



5 | 2021  
special issue

# Beyond *Guanxi*: Chinese Historical Networks

**HENRIKE RUDOLPH | SONG CHEN**



## **Imprint**

### **Université du Luxembourg 2021**

#### **Luxembourg Centre for Contemporary and Digital History (C<sup>2</sup>DH)**

Université du Luxembourg  
Belval Campus  
Maison des Sciences Humaines  
11, Porte des Sciences  
L-4366 Esch-sur-Alzette

The publication of this special issue was in part supported by the Max Weber Foundation and the Fritz Thyssen Foundation.

#### **Editors**

Asst. Prof. Dr. Marten Düring (Luxembourg Centre for Contemporary and Digital History | C<sup>2</sup>DH)  
apl. Prof. Dr. Robert Gramsch-Stehfest (Friedrich-Schiller-Universität Jena)  
Dr. Christian Rollinger (Universität Trier)  
Dr. Martin Stark (ILS – Institut für Landes- und Stadtentwicklungsforschung, Dortmund)  
Clemens Beck, M. A. (Friedrich-Schiller-Universität Jena)

#### **Guest Editors**

Dr. Henrike Rudolph, University of Göttingen  
Dr. Song Chen, Bucknell University

ISSN 2535-8863

#### **Contact**

##### **Principal Contact**

JHNR-editors@historicalnetworkresearch.org

##### **Support Contact**

Dr. Marten Düring (Université du Luxembourg)  
JHNR-support@historicalnetworkresearch.org

#### **Typesetting**

text plus form, Dresden, Germany

#### **Cover image**

Chinese star chart, British Library, Or.8210/S.3326 recto,  
<https://www.bl.uk/collection-items/chinese-star-chart>

#### **Copyediting**

Andy Redwood, Barcelona, Spain

#### **Published online at**

<https://doi.org/10.25517/jhnr.v5i1>

This work is licensed under a Creative Commons License:  
Attribution-NoDerivatives 4.0 (CC BY-ND 4.0)

This does not apply to quoted content from other authors.

To view a copy of this license, please visit

<https://creativecommons.org/licenses/by-nd/4.0/deed.en>

## TABLE OF CONTENTS

### Introduction

- Song Chen/Henrike Rudolph*  
Beyond Relationships and *Guanxi*:  
An Introduction to the Research of Chinese Historical Networks iii

### Articles

- Shang Wenyi/Sang Zizhou*  
Solidity in a Turbulent Flow: The Social Network  
of Aristocratic Families in the Eastern Jin Dynasty (317–420 C.E.) 1
- Xiong Huei-Lan*  
Path toward the Top Leadership: A Network Analysis  
of the Civil Service System in the Early Southern Song (1131–1164) 33
- Peter K. Bol*  
From Kinship to Collegiality: Changing Literati Networks, 1100–1400 87
- Cécile Armand/Christian Henriot*  
From Textual to Historical Networks: Social Relations  
in the Biographical Dictionary of Republican China 114
- Anne S. Chao/Jie Yang/Zhandong Liu/Qiwei Li*  
Network of Words: A Co-Occurrence Analysis of Nation-Building  
Terms in the Writings of Liang Qichao and Chen Duxiu 154
- Marilyn Levine*  
Post WWI Chinese Revolutionary Leaders in Europe 187

### Databases and Tools

- Marcus Bingenheimer*  
The Historical Social Network of Chinese Buddhism 233
- Michael Fuller/Hongsu Wang*  
Structuring, Recording, and Analyzing Historical Networks  
in the China Biographical Database 248

<i>Marilyn Levine</i> Biography for Historical Analysis: A Chinese Biographical Database	271
<i>Matthias Arnold/Henrike Rudolph</i> Network Data in the Early Chinese Periodicals Online Database (ECPO)	288
<i>Baptiste Blouin/Nora van den Bosch/Pierre Magistry</i> Creating Biographical Networks from Chinese and English Wikipedia	303





The Journal of  
**HISTORICAL  
NETWORK  
RESEARCH**

SONG CHEN/HENRIKE RUDOLPH

# **Beyond Relationships and *Guanxi*: An Introduction to the Research of Chinese Historical Networks**

*Journal of Historical Network Research 5 (2021) iii–xxxii*

Network research promises to bridge the divide between humanities, arts, and sciences, as well as to further our understanding of the past. This special issue sets out to bridge yet another gap, one that results from disciplinary divisions and language barriers. The rapid growth of Digital Humanities in general, and network science in particular, within the field of East Asian Studies has given rise to a thriving community of scholars with their own journals and conferences, which has gone mostly unnoticed in Anglophone circles. We thus seek to cross this divide by introducing the emerging field of network research in Chinese history to a broader audience. We hope to engage both experts on Chinese history who are not yet familiar with the theories and methods of network research, and network scholars specializing in other world regions who may draw inspiration from the way historians have applied network analysis to the study of Chinese history. This introduction thus serves to touch base with both groups and stimulate constructive dialogues across disciplinary boundaries. It sets out to explain network analysis as a modern *Kulturtechnik* and questions the ahistorical, culturalist assumptions, including the concept of *guanxi*, that cast a long shadow on previous studies of interpersonal relationships in Chinese society. It emphasizes this with a brief discussion of how interpersonal relationships evolved over the two millennia of imperial and modern Chinese history, followed by an overview of the state of Chinese historical network research as it moves beyond networks as metaphors for social histories to the reconstruction and structural analysis of social relationships. As the articles collected in this special issue demonstrate, research on Chinese networks is no longer confined to interpersonal ties but includes explorations of texts, bureaucratic practices, and material objects.<sup>1</sup> Today, Chinese historical network research has benefited from the development of databases that collect and synthesize biographical data from discrete historical sources as well as tools that facilitate text markup and data visualization. Therefore, a special section of this issue is designated to database projects that have elevated – and continue to elevate – the quantitative study of China’s past to a new level. We conclude with a discussion of the future and potential of scholarship on Chinese historical networks.

---

1 In comparison to the European community of network scholars, in Chinese studies exchanges between historians and archaeologists employing the methodology of network analysis are still rare and leave much room for improvement. Despite an explicit reference to archaeological network research in the Call for Papers, we unfortunately received no submissions from this field.

## 1. Network Analysis as a Modern *Kulturtechnik*

In this special issue, we ground our work in the assumption that relational thinking and an awareness of the self, others, objects, and nature as being interrelated is an anthropological constant. Past societies and individuals were acutely aware of how kinship ties, social affinity, geographical proximities and trade routes opened or limited their possibilities of action. However, even though human thinking has always been relational, the analytical exploration of networks in a narrower sense is a modern *Kulturtechnik*, a cultural technique that operates with network conceptualizations.<sup>2</sup> Relationships only become a network once they are aggregated and abstracted as a totality that is presumed to possess identifiable and describable structural characteristics. Networks are therefore products of a collective imagining, and can “gain social and cultural influence through their performative implementation as models of action and description.”<sup>3</sup>

The German sociologist Georg Simmel (1858–1918) never used the term “social network,” but his relational and formalist theory of society has foreshadowed many core ideas in the methodology of formal network analysis that we practice today. Simmel rejects the Durkheimian notion of society as a real, material entity or a thing-like substance. For Simmel, society exists only in and through “stable and patterned forms of reciprocal interaction between individuals,” which he calls “sociation” (*Vergesellschaftung*).<sup>4</sup> “Society,” he argues, “merely is the name for a number of individuals, connected by interaction,” and “[t]he large systems and the super-individual organizations that customarily come to mind when we think of society, are nothing but immediate interactions that occur among men constantly...but that have become crystallized as permanent fields, as autonomous phenomena.”<sup>5</sup> Simmel’s conceptualization of society is relational, and his approach to analyzing society and sociation is formalist.<sup>6</sup> Drawing inspiration from geometry that abstracts spatial forms from concrete objects, Simmel argues that sociology must study the *forms* of sociation – e.g., domination and subordination, competition, inclusion and exclusion – that underlie the widely varying contents of actual social interactions. Simmel’s ideas, although considered vague by many of his contemporaries, provide the conceptual apparatus for the sociograms and quantitative analysis of social networks that developed in the twentieth century.

---

2 Sebastian Gießmann, *Netze und Netzwerke: Archäologie einer Kulturtechnik, 1740–1840*, Kultur- und Medientheorie (Bielefeld: transcript, 2006), 13.

3 Gießmann, *Netze und Netzwerke*, 18.

4 Alex Law, *Key Concepts in Classical Social Theory* (London: SAGE Publications Ltd, 2011), 181.

5 Kurt H. Wolff ed., *The Sociology of Georg Simmel* (New York: The Free Press, 1950), 10.

6 For an in-depth discussion of formalism and relationalism in social network research, see Emily Erikson, “Formalist and Relationalist Theory in Social Network Analysis,” *Sociological Theory* 31, no. 3 (2013): 219–42.

As an analytical approach, network analysis has a strongly structuralist orientation. It asserts that relational ties between actors create a structural environment, i.e. “the network,” which both empowers and constrains individual action. Accordingly, actors in a network are presumed to be interdependent, and the relative importance of each actor hinges on his or her position in the network. Therefore, relationships are not by themselves networks. To view them as networks is an analytical decision grounded in a structuralist worldview. It directs attention from each individual actor to a totality of interacting actors and their linkages. It is the structural properties and topological features of this totality, or network, that become the focus of analysis.

We should not assume that actors in a network possess the same God’s eye view, as do network analysts, concerning the topology of that network.<sup>7</sup> However, it would be just as problematic to surmise that the network and its structural forces are purely the invention of modern theorists, of which historical actors were totally oblivious. We argue that the truth was somewhere in between. Historical actors often viewed a collection of crisscrossing relationships as forming a totality that both provided opportunities for and imposed constraints upon individual actors. Yet their view of this networked totality was usually simultaneously holistic and agnostic, characterized by a fatalistic acceptance of its complexities and ambiguities. Political writings from imperial China, for example, were replete with accusations against men whose personal relations – through marriage, political patronage, or else – allegedly enabled them to create alternative centers of power that solidified their own position in government at the expense of dynastic and public interests. These accusations, however, often vaguely described interpersonal relationships as forming a disorderly totality, comparing it to the root system of plants that was “entwined” and “entangled” (*pangen cuojie* 盤根錯節) beyond deciphering.<sup>8</sup> By contrast, modern network analysts presume that a network has structural features that can be teased out. By carefully defining the scope of analysis (e.g., selecting some nodes and ties while ignoring others) and employing graph-theoretical methods of analysis (e.g., partitioning a network into components and clusters, counting links and distances), they reduce networks to theoretical models in order to tame them and make them “manageable.” In this sense, network analysis is a modern scientification of historical relationships and social structures.

---

7 Usually they do not: people do not even know all the friends of their friends. That actors lack full knowledge of the topology of the network that they participate in is, for instance, a major issue in Stanley Milgram’s famous small-world experiment. See Duncan J. Watts, *Six Degrees: The Science of a Connected Age* (New York and London: W. W. Norton and Company, 2004), 132–156.

8 For an example, see Chen Dong 陳東, “Dengwenjianyuan san shang Qinzong huangdi shu” 登聞檢院三上欽宗皇帝書 [Third Memorial Submitted to Emperor Qinzong through the Public Petition Review Bureau], in *Quan Song wen* 全宋文, ed. Zeng Zaozhuang 曾棗莊 and Liu Lin 劉琳 (Shanghai: Shanghai cishu chubanshe, 2006), 175: 3832.194.

Historians usually point to two factors to explain this conceptual shift: the changing understanding of the human body, and the growing transregional integration through transportation and communication in the nineteenth century.<sup>9</sup> “The network-term is one of the concepts that produce what they recognize and also recognize where nothing was produced but emerged evolutionarily.”<sup>10</sup> Indeed, the use of the term “network” in English-language publications increased after 1920 and then gained in frequency again after 1980. Ahnert et al. note that “the word for network in certain other [European] languages carries the same lineage from the word for the material act of weaving nets.... The etymology of ‘network’ in both Germanic and Romance languages, therefore, contains a set of assumptions about structure, pattern, order, and distribution. In many cases, a maker or designer is implied.”<sup>11</sup>

In Chinese history, we can observe similar etymologies. The term “network” gained popularity as a technical term for electrical networks (*dian wangluo* 電網絡) in the first half of the twentieth century.<sup>12</sup> As in the European context, the characters *wang* 網 and *luo* 絡 are both associated with weaving, such as nets for fishing or catching birds. *Wangluo* is today generally used for the internet and serves as a standard translation for “network analysis” (*wangluo fenxi* 網絡分析). The character *luo* is also part of the vocabulary of traditional Chinese medicine and describes the meridians through which fluids are said to circulate in the body.<sup>13</sup> And as the Dunhuang star charts demonstrate, we can already find visualizations of imagined networks in Tang-dynasty (618–907 CE) manuscripts, where stars are recorded as nodes in different colors, connected by lines forming constellations.<sup>14</sup> Yet such pre-modern notions of circulation and connectivity, or conversely of “blocked” channels of exchange, differ substantially from the structuralist approaches of present-day network analysis. Even though charts of star signs or the meridians in a human body can be seen as precursors to network graphs, they function as maps, similar to maps of a metro network. They help vi-

- 
- 9 Jürgen Osterhammel, *The Transformation of the World: A Global History of the Nineteenth Century*, America in the World (Princeton: Princeton University Press, 2014), 711.
  - 10 Hartmut Böhme, “Netzwerke. Zur Theorie und Geschichte einer Konstruktion,” in *Netzwerke: eine Kulturtechnik der Moderne*, ed. Jürgen Barkhoff (Köln: Böhlau, 2004), 27.
  - 11 Ruth Ahnert et al., *The Network Turn: Changing Perspectives in the Humanities*, 1st ed. (Cambridge: Cambridge University Press, 2020), 14–16.
  - 12 For example, see Zishan 紫珊, “Dianwangluo (NETWORK) de jiefa” 電網絡 (NETWORK) 的解法 [Solutions to Electric Networks], *Kangzhan kexue* 抗戰科學, no. 3 (1939): 36–37.
  - 13 Chen Qiao 陳峭 et al., “Guanyu goujian Zhongyi ‘jingluo tizhi’ xueshuo de shexiang” 關於構建中醫“經絡體制”學說的設想 [The Construction of a Theory of “Meridian System” in Chinese Medicine], *Zhonghua Zhongyiyao zazhi* 中華中醫藥雜誌 33, no. 6 (2018): 2448–51.
  - 14 Jean-Marc Bonnet-Bidaud, Francoise Praderie, and Susan Whitfield, “The Dunhuang Chinese Sky: A Comprehensive Study of the Oldest Known Star Atlas,” *Journal of Astronomical History and Heritage*, no. 12 (2009): 39–59.

sualize spatial or conceptual connectedness, but they do not subscribe to the holistic, structuralist, and statistical aspirations of present-day network research, which features new analytical perspectives, such as clustering, reachability, and centrality.

## 2. Questioning the Otherness of Chinese Networks

This special issue does not set out to depict the “otherness” of Chinese networks, nor argue in an Orientalist vein that Chinese networks (whether past or present) differ substantially from those of other cultures or world regions. In the preceding section, we argued that Chinese historical etymologies, conceptualizations, and visualizations of connectedness do not show any qualitative difference from, for example, European counterparts. Despite all the regional differences, one central theme in the project of modernization, shared across cultures, is the “awareness of a great variety of roles existing beyond narrow, fixed, local, and familial ones.”<sup>15</sup> Some scholars maintain, however, that there is something peculiar about networks in Chinese culture and our motives to study them. Fröhlich and Heinrich argue that “the network approach in China studies was born as a counter-narrative to explain in what respect Chinese society *differed* from that of the western world.”<sup>16</sup> In the eyes of some twentieth-century sociologists, like Fei Xiaotong, the difference between China and the West lies in their pattern of organization. That is, as Fröhlich and Heinrich have neatly summarized, “[w]here in the West there were formal hierarchies, China was built on informal networks.”<sup>17</sup> This belief gave rise to *guanxi* as a catchword in social sciences and business management, which is used by some observers not merely as a Chinese wording for personal connections that are found in all societies, but to signify “an essential and defining element of Chinese culture, handed down relatively unchanged through time and space.”<sup>18</sup> *Guanxi* carrying these connotations is not only adopted as an Orientalist label, but also as a Chinese self-fashioning. For instance, some Chinese scholars maintain that a certain type of networking has shaped Chinese social interactions since the time of Confucius.

I believe that self-organization, the network-like structure and *guanxi* management are the main keys to understanding the nature of Chinese management. And

---

15 Shmuel N. Eisenstadt, “Multiple Modernities,” *Daedalus* 129, no. 1 (2000): 4.

16 Hajo Fröhlich and Mathias Heinrich, “China – A Network Society? The Network Concept in Research on Chinese History and Society,” in *China Networks*, ed. Jens Damm, Berliner China-Hefte 35 (Berlin: LIT-Verl, 2009), 3.

17 Fröhlich and Heinrich, “China – A Network Society?,” 4.

18 Thomas Gold, Doug Guthrie, and David Wank, “An Introduction to the Study of *Guanxi*,” in *Social Connections in China: Institutions, Culture, and the Changing Nature of Guanxi*, ed. Thomas Gold, Doug Guthrie, and David Wank (Cambridge: Cambridge University Press, 2002), 3.



they can all be traced to Zhongyong, or the doctrine of dynamic balance, developed by Ziji [sic], the grandson of Confucius, more than two thousand years ago.<sup>19</sup>

This agentless, ahistorical view of *guanxi* as a millennia-old “Chinese” cultural trait is established as a counter-narrative to Euro-American claims to cultural hegemony and the related assumption that all countries will eventually “reach” a Western standard of civil society and governance. With that said, this infatuation with *guanxi* is – at least to some extent – also the result of a research environment where the study of culture and habits is seen as politically less conspicuous than formal studies of institutionalized structures and systemic inequalities. In other words, the generalization of Chinese society made up of informal ties was, in part, a result of the difficulty to study formal hierarchies in contemporary China. Andrew Kipnis, for example, admitted that he “found it convenient to hide my own interest in the sensitive politics of gender, age, and state regulation behind the innocuous label of ‘customs and habits’ (*fengsu xiguan* 風俗習慣).”<sup>20</sup> Yet, is there really a perennial “Chinese” mode of networking? Does the concept of *guanxi* hold any explanatory potential? Or, in other words, is *guanxi* in the Straussian sense, “good to think with”?

This use of *guanxi* as an analytical concept to capture a presumably unique feature of Chinese society reflects an uneasy tension that scholars of networks more generally (whether past or present, European or non-European) find themselves caught in: Network analysis promises a level of objectivity and comparability that extricates measurable structures from the thicket of cultural ambiguities, but explanations of the emergence and utilization of the observed structures often revert to culturalist assumptions, for which the use of *guanxi* is only one example. Network research along these lines takes social interactions and relationships in a society out of their historical context. It falls back on a static and essentialist view of the so-called “Chinese” or some other culture, while downplaying the agency of historical actors and paying no heed to how social relationships are shaped by, and in turn shape, the evolving institutional environment and social structures.

This leaves a theoretical void in network research, which Emirbayer and Goodwin already observed in the 1990s and that remains unresolved:

Network analysis all too often denies in practice the crucial notion that social structure, culture, and human agency presuppose one another; it either neglects or inadequately conceptualizes the crucial dimension of subjective meaning and

---

19 Jiade Luo and Yong Zhou, *Social Networking and Chinese Indigenous Management* (Reading, UK: Paths International Ltd, 2014), 6.

20 Andrew Kipnis, “Practices of *Guanxi* Production and Practices of *Ganqing* Avoidance,” in *Social Connections in China: Institutions, Culture, and the Changing Nature of Guanxi*, 22.

motivation – including the normative commitments of actors – and thereby fails to show exactly how it is that intentional, creative human action serves in part to constitute those very social networks that so powerfully constrain actors in turn.<sup>21</sup>

One possible way out of this trap, and to bring human agency back in, is not to study social networks as fixed structures but as practices of establishing or maintaining relationships in specific institutional, social, and cultural contexts, as demonstrated by Andrew Kipnis in his study of a North China village from the 1940s to the 1990s.<sup>22</sup> By studying not only the cultural logic of *guanxi* practices, but also how modern institutions and politics shaped the rules and meaning of these practices and how individuals actively manipulated and interpreted the rules to which they were subject, Kipnis challenges the view that some sort of essential Chineseness could provide an adequate explanation for *guanxi* practices. He thus avoids Orientalist fallacies of defining *guanxi* in contrast to non-Chinese types of social relations and manages to include critical theories of culture like Bourdieu's practice theory.<sup>23</sup>

A study of social relationships, therefore, must pay due attention to the social meaning of a relationship in its specific historical context. Historians, with their sensitivity to historical change, their training in the critical reading of sources, and the necessary awareness of their own potential misreading of historical "facts," are well equipped to engage with questions of biased network data, the dialectics between social practice and social structure, the role of individual agency, or revolutionary tipping points. Hence, historical network research enables us to trace the transformation of social networks as a result of evolving practices of developing and sustaining social relationships, which were both responses to and constitutive of dynamic institutional, social, and cultural environments. It therefore gives new impetus to the critical engagement with the conceptualization of social relationships in Chinese society and the concept of *guanxi* in particular, which is still dominated by social science research on post-1980s mainland China.

In brief, we argue that relational thinking is powerful in coming to terms with past and present phenomena, but we reject the idea that there are relationships "with Chinese characteristics" that take up functions or forms unmatched in any other time or culture. The following section offers an overview of the changing nature of elite networks from imperial to modern China, thus providing an entry point to the research articles.

---

21 Mustafa Emirbayer and Jeff Goodwin, "Network Analysis, Culture and the Problem of Agency," *American Journal of Sociology* 99, no. 6 (1994): 1413.

22 Andrew B. Kipnis, *Producing Guanxi: Sentiment, Self, and Subculture in a North China Village* (Durham, NC: Duke University Press, 1997).

23 Kipnis, "Practices of *Guanxi* Production and Practices of *Ganqing* Avoidance," 23–25.

### 3. Interpersonal Relationships

Contributions in this issue underscore that throughout Chinese history, both large historical processes (e.g., the rise and fall of centralized states and technological revolutions) and short-term episodes (e.g., foreign invasions, migration, and dynastic change) had constantly shaped the nature and scope of interpersonal relationships that were integral to the evolving structures of power in Chinese society. These studies show that throughout Chinese history, informal networks of the elite shaped institutionalized systems of government and, conversely, that the continuous changing of institutions also left its mark on China's social fabric. They also demonstrate that the mutually constitutive relationship between formal institutions and informal social ties was always mediated by ever-changing cultural notions of the historical actors. These studies dismantle the misconception of an immobile social order in imperial China (221 BCE–1912 CE) and a static, essentialist view of “Chinese” culture. They demonstrate that cultural notions were always in flux and that the decision to build what relationships, with whom, and by what means were also apt to change in response to the times. These works call into question any attempt that takes informal interpersonal ties, or *guanxi*, as a perennial and unique cultural trait of Chinese society while ignoring the specific form of social relationship that was valued and developed under specific institutional, social, and cultural conditions.

In the past decades, social historians have challenged narratives of a Chinese society awoken by foreign intrusion in the mid-nineteenth century from its millennia of slumber, and paid more attention to social relationships than bureaucratic institutions. They find that the imperial bureaucracy not only operated according to the meticulous rules laid out in the administrative codifications, but was also permeated by a myriad of private ties of loyalty and obligations. The Later Han (25–220 CE) officials were enmeshed in a web of patron-client relations that tied disciples to their teachers and subordinates in office to their former superiors. These hierarchical relations carried demanding social and political obligations on the clients, who were expected to support their patrons in political struggles and mourn for them if they died. Although ties of locality underlay most patron-client relations, high-ranking officials sometimes attracted clients from all over the country. As Patricia Ebrey has observed, “[p]atron-client ties could be extended beyond the two individuals involved to form networks of men linked to common patrons, and to the patrons of their patrons.”<sup>24</sup> These ties, therefore, brought together elite men from widely separated places and fostered among them a common identity that transcended their geographic origins.<sup>25</sup> In

---

24 Patricia Ebrey, “Patron-Client Relations in the Later Han,” *Journal of the American Oriental Society* 103, no. 3 (1983): 533.

25 Ebrey, “Patron-Client Relations in the Later Han,” 541.

Later Han, the desire of elite males to enter this web of connections, which was inseparable from holding office and participating in the metropolitan culture, were counterbalanced by a persistent effort to consolidate their social station in the provinces, where they owned landed estates and headed large descent groups with a retinue of dependents and retainers. This effort became extremely rewarding at the turn of the second century when domestic rebellions, foreign invasions, and bloody political strife ripped apart the Han imperial order. The elite men's solid local position provided the necessary social and economic resources to tide them over the crisis.

In the four centuries that followed, foreign conquerors and military dynasts seeking to establish effective control eagerly sought support from these elites by appointing them to offices. The most influential elite families (the so-called "great clans") intermarried with each other and with the imperial clan and enjoyed, in effect, hereditary and exclusive access to high office. They constituted a distinctive and privileged superelite that modern historians have conveniently referred to as the "aristocracy." Unlike nobility in many parts of medieval Europe, aristocracy in China was not a legally defined status but a *de facto* social category growing out of an elite culture that claimed talent and character were prerequisites for holding office, but were also fundamentally inborn and thus hereditary. The superelite arrogated to itself the power of evaluating any candidate's talent and character, which they asserted were subtle and mysterious and could only be discerned by men of extraordinary perception.<sup>26</sup> In the new "Nine-Rank" system of bureaucratic recruitment, men from the superelite families monopolized the office of the Impartial Judge (*zhongzheng* 中正), who ranked each candidate based on his deportment and pedigree; the candidate's rank, in turn, determined the level of office for which he was eligible. These superelite families thus formed a status group, closed to itself, that was both the arbiter of its own culture and the master of its own political fate. The social prestige of the superelite and its political privileges hinged upon its purity, which the constituent families guarded jealously by associating only with one another and marrying exclusively among themselves. In this altered environment, patron-client ties lost their social meaning and function, while horizontal ties of marriage and affinity within the status group played a key role in defining the aristocracy's social distinctiveness and solidifying its high status (see SHANG in this issue).

The Chinese aristocracy was thus, from the very beginning, entwined with imperial power. Pedigree and officeholding reinforced each other: proof of an illustrious pedigree was a *sine qua non* for holding high office, but only a pedigree replete with ancestors in high office were considered illustrious. Consequently,

---

26 Mark Edward Lewis, *China between Empires: The Northern and Southern Dynasties* (Cambridge, MA: Belknap Press of Harvard University Press, 2009), 39–40.

during the centuries at the peak of their power, one after another the aristocratic families left their provincial estates to live in the capital, where they kept company with social equals, stayed close to the court, and molded their temperament in highbrow culture. By the early seventh century, when the Tang founders successfully rebuilt a strong centralized authority, aristocratic families constituted an endogamous, semi-hereditary officeholding elite that resided mostly in the dynasty's capital region.<sup>27</sup>

This superelite maintained political dominance well into the ninth century, but the foundations of its dominance were seriously undermined as early as the sixth century, when dynastic rulers abolished the Nine-Rank system and instituted, in its place, the civil service examinations as a new way of recruiting state bureaucrats. The aristocratic families successfully adapted themselves to this challenge, and until the late ninth century the majority of those who passed the examinations and held office were men of aristocratic descent. In spite of this success, the aristocracy's acceptance of the new rules sowed the seed of its ultimate demise. Having acknowledged the need to compete among themselves and prove their worth in the examinations, the aristocrats were shattering the cultural premises of their social position: that talent was hereditary, that high office was a prerogative of those of good ancestry, and that the superelite was the arbiter of its own social worth and its eligibility for government service.<sup>28</sup> Under the new rules, the privileged status of the aristocracy perched precariously on its ability to monopolize the cultural and social resources that were necessary to sustain its examination success.

By the early years of the Song (960–1279 CE), a series of changes brought this monopoly to an end. The spread of printing technology made books widely available at lower prices. State-sponsored schools were established across the country, equipped with student dorms and libraries, run by salaried instructors, and financed by ringfenced resources. Reforms of the civil service examinations, such as the introduction of regional quotas and the enforcement of anonymity, ensured a degree of fairness for candidates of diverse geographical and family origins. By the mid-eleventh century, it was clear that capital residence and status-group endogamy were no longer requisites for political eminence. The cultural premises and social practices underpinning the status of the aristocracy thus became ancient history. Officials, as well as those whose classical education prepared them for the examinations and government service, now constituted a distinctive social stratum that encompassed a much broader segment of society than the aris-

---

27 Nicolas Tackett, *The Destruction of the Medieval Chinese Aristocracy* (Cambridge, MA: Harvard University Asia Center, 2014), chapters 2–3.

28 David G. Johnson, *The Medieval Chinese Oligarchy* (Boulder, Colorado: Westview Press, 1977), 149–52.

tocracy of earlier times. These men, the self-styled literati, were scattered all over the country and defined themselves by learning, not pedigree. But who decided upon learning? The court and the literati offered different arguments. To those who embarked on a bureaucratic career, the court's argument meant a great deal. It was generally accepted that a good government was a government run by good men, but the imperial administration of Song times operated on the premise that talent and character were neither inborn nor hereditary and therefore ancestry was irrelevant. Instead, men had to demonstrate their talent in the examinations and nourish it by serving in the academic institutes at the imperial court (see XIONG in this issue).

Learning, however, was also flourishing without state sponsorship and outside its control. Scholars shared their interpretations of classical texts by writing and publishing commentaries. They spread their views by building academies and taking on students. Teachers were invited to lecture in different places, and students traveled to study under different masters. Intellectuals of different persuasions debated in letters and at face-to-face meetings.<sup>29</sup> Activities like these created a vibrant community of learned men, or literati, in which membership was a matter of mutual recognition and outside the purview of state authority. Underpinning this community were teacher-disciple relations and crisscrossing horizontal ties of literary and intellectual exchanges that were geographically extensive. Thus, when kinship networks were more localized, as in the thirteenth and fourteenth centuries, these ties of learning played an instrumental role in sustaining a supralocal literati identity (see BOL in this issue).

The late nineteenth and early twentieth century witnessed an acceleration of social change. The rule of the Qing dynasty (1644–1912 CE) was gradually hollowed out and remained a mere symbolic reference point for continuity in a spiral of reform and revolution, until it finally collapsed in 1912. Industrialization, tied to increasing urbanization, regional connectedness through new means of communication and travel, the abolition of the examination system for recruiting state bureaucrats, and the accumulation of political power by local elites, all contributed to a reconfiguration of existing social structures. Causal as well as symptomatic of change, “the emergence of a private sphere and of interest groups to represent it both reflected and contributed to the breakdown of traditional authority.”<sup>30</sup>

---

29 For an incisive discussion of the rise of the *shi*-oriented culture that competed with the court-oriented culture, see Robert Hymes, “Sung Society and Social Change,” in *Cambridge History of China*, Volume 5, Part Two, ed. John W. Chaffee and Denis Twitchett (Cambridge: Cambridge University Press, 2015), 631–32.

30 Joseph Fewsmith, “From Guild to Interest Group: The Transformation of Public and Private in Late Qing China,” *Comparative Studies in Society and History* 25, no. 4 (1983): 618.



These challengers to traditional authority included previously marginalized groups, such as women or workers, who organized themselves to become political forces, as well as an increasingly self-confident urban elite. In part, the new intelligentsia recruited itself from the traditional literati families now struggling to find their place in a modern nation-state, as well as from a young, less affluent generation that embraced the possibilities of educational expansion and state-sponsored foreign studies programs. “New learning” (*xinxue* 新學), which included, for example, the mastery of one or more foreign languages, knowledge of natural and social sciences, engineering, and law, as well as practical industrial skills, emerged as a new source of cultural capital.<sup>31</sup> Guilds, professional organizations, and native-place associations mobilized and structured civil society and offered anchorage in turbulent times by integrating premodern and modern social fabrics. In Beijing, for example, “native-place ties served to facilitate a multi-directional flow of information and influence between center and region.”<sup>32</sup>

The mobility of China’s young elites increased even on a global scale. They travelled the world in search of new knowledge and returned not only with the latest scientific news, but also with a head full of utopian visions for a new China. Especially in urban centers like Shanghai, schools and study societies formed new nodes in the highly politicized networks.<sup>33</sup> The personal ties that had been forged abroad now became a framework for political activism, from anarchist circles to the first Communist party cells (see LEVINE in this issue). In many of the newly emerging social organizations, however, traditional hierarchical patron-client networks continued to exist. Personal loyalty was a glue that tied troops to their leaders, political activists to their mentors, and even within the Leninist-style party organizations of the Nationalists and Communists, family ties, seniority in party membership, and shared biographical experiences continued to shape social networks. Therefore, amidst the growing importance of modern institutions, like an expanding state bureaucracy, courts and parliaments, public schools and universities, a deeper understanding of interpersonal relationships remains crucial in uncovering the social history of twentieth-century China. These institutions not only framed social interaction, they also opened the pathways of social mobility, opportunities for exchange, and the forging of new alliances.

---

31 Barbara Schulte, “Webs of Borrowing and Lending: Social Networks in Vocational Education in Republican China,” in *World Yearbook of Education 2012*, ed. Gita Steiner-Khamisi and Florian Waldow (London: Routledge, 2012), 115–38.

32 Richard Belsky, “Placing the Hundred Days: Native-Place Ties and Urban Space,” in *Rethinking the 1898 Reform Period: Political and Cultural Change in Late Qing China*, ed. Rebecca E. Karl and Peter Gue Zarrow (Cambridge, MA: Harvard University Press, 2002), 132.

33 Stephen Averill, “The Cultural Politics of Local Education in Early Twentieth-Century China,” *Twentieth-Century China* 32, no. 2 (2007): 22–23.

In sum, social relationships must be studied as practices that are grounded in historically specific institutional, social, and cultural contexts. Social relationships are developed and maintained by individual actors who constantly interpret the structures and circumstances of the institutional environment they are in and, on the basis of these interpretations, appreciate the meaning and value of different types of social relationships. In other words, social networks are always embedded in culture, but culture is dynamic and evolves in tandem with social conditions and political institutions. The shifting views of talent, pedigree, and learning in Chinese history, which profoundly shaped the scope and nature of elite networks from one period to another, were but one example. Therefore, any study of informal ties, or *guanxi*, in Chinese society must dispel the specter of cultural essentialism, which not only overlooks individual agency but, in the final analysis, also denies the causal power of culture itself. By studying social networks as evolving practices and by integrating cultural and institutional perspectives, historical network research gives due attention to human agency and holds the promise of reaching more meaningful interpretations of social network structures. Moreover, to understand how exactly individual agency – as an expression of cultural, religious, or political convictions or as a strategic choice – was possible within these sets of social and institutional boundaries is a challenge for historians of any period or world region. Therefore, the diversification of the field of historical network research hopefully further challenges our assumptions about “standard” (i.e. Euro-American) development paths and increases our awareness of multifaceted modes of establishing and exploiting social ties.

#### 4. Network Analysis and Prosopography

Social historians have studied interpersonal relationships in Chinese history for a long time, but only recently have they begun to adopt the concepts and tools of network analysis. This is nevertheless an unsurprising outcome of their enduring efforts to understand how the character of China’s dominant elite changed over the two millennia of imperial history. Many scholars working on this topic in the twentieth century embraced the traditional approach of prosopography.<sup>34</sup> They collected data on a well-defined population (e.g., officials who were sufficiently prominent to warrant a biography in dynastic histories) and investigated their common characteristics (e.g., ancestry and regional origin).<sup>35</sup> Research along these lines led them to also look at the different ways in which elite men

---

34 Lawrence Stone, “Prosopography,” *Daedalus* 100, no. 1 (1971), 46.

35 Sun Guodong 孫國棟, “Tang Song zhi ji shehui mendi zhi xiaorong: Tang Song zhi ji shehui zhuanbian yanjiu zhiyi” 唐宋之際社會門第之消融——唐宋之際社會轉變研究之一 [The Dissipation of Prominent Families in the Society of Tang and Song Times: A Study of Social Transformations in Tang and Song Times], *Xinya xuebao* 新亞學報 4, no. 1 (1959): 211–304.

and women interacted and made connections with one another. Some scholars collected data on elite marriages, teacher-disciple ties, and literary exchanges; others published case studies on individual clans. Until the recent adoption of formal network analysis, however, historians lacked the wherewithal to analyze the structural pattern in elite networks. Several prosopographical studies published in the last quarter of the twentieth century, for instance, took on elite marriages in Tang (618–907 CE) and Song (960–1279 CE) times as a central subject of investigation.<sup>36</sup> These works typically study marriage practices from a spatial perspective and ask, for example, whether elite families married within or across prefectural borders. They fail to consider the possibility that through marriage, elite families at different moments of history may have formed networks that were markedly different *in structure*. For instance, these studies show that official families in the Tang and early Song preferred to live in the capital region and arranged marriages mainly among themselves, but they have not explored whether these marriages constituted a network that centered on any particular family or fragmented into several clique-like subgroups. To ask questions like this requires a conceptual leap that views interpersonal relationships, such as marriage, as constitutive of a network with structural properties that can be discovered and described.

As in the field of ancient European history, the adoption of formal network analysis among scholars of imperial China is a “logical extension of traditional prosopographical research.”<sup>37</sup> Network analysis offers enticing conceptual and operational tools for analyzing a historical phenomenon that has long enamored prosopographers. A recent study by Nicolas Tackett, for example, reveals the structural properties of the ninth-century capital elite’s marriage network, which partitioned into two distinct clusters. One cluster was organized around the Tang imperial family and included many eminent clans of northwestern origin. The other cluster, in contrast, was composed almost entirely of prestigious families that resided in the Tang Eastern Capital, which had a more egalitarian structure and was not dominated by any single descent group.<sup>38</sup> A similar study by Song Chen compares prefectural governors’ marriage networks in the mid-eleventh and early thirteenth centuries. It shows that mid-eleventh-century prefects were intricately connected, by consanguinity and marriage, in a dense and expan-

- 
- 36 Robert M. Hartwell, “Demographic, Political, and Social Transformations of China, 750–1550,” *Harvard Journal of Asiatic Studies* 42, no. 2 (1982): 365–442. Robert P. Hymes, *Statesmen and Gentlemen: The Elite of Fu-chou, Chiang-hsi, in Northern and Southern Sung* (Cambridge: Cambridge University Press, 1986). Beverly Bossler, *Powerful Relations: Kinship, Status, and the State in Sung China (960–1279)* (Cambridge, MA: Harvard University Asia Center, 1998).
- 37 Christian Rollinger, “Prolegomena. Problems and Perspectives of Historical Network Research and Ancient History,” *Journal of Historical Network Research* 4 (2020): 7.
- 38 Tackett, *The Destruction of the Medieval Chinese Aristocracy*, 125–26.

sive network that centered on the capital elite, but those in the early thirteenth century only formed small regional clusters, each spanning a few adjacent prefectures.<sup>39</sup>

The union between network analysis and prosopography is cemented by their shared methodological outlook. Both approaches look beyond the individual historical actor and explore patterns at the group level, which makes network analysis an easy sell to social historians who have already embraced prosopography in their research. Nevertheless, the difference between these two approaches is equally pronounced. Whereas traditional prosopography focuses on the shared characteristics of the study population (e.g., family and educational background, economic status, careers, and religious affiliations), network analysis concentrates on the interactions and relationships between and among historical actors.<sup>40</sup> By bringing the practices of interaction and relationships into focus, network analysis provides an alternative means of conceptualizing social stratification. Twentieth-century studies of the imperial Chinese elite, having flourished under Marxist and Weberian influences, often defined social status on the basis of personal and familial attributes, such as landownership, ancestry, and educational achievements. By contrast, the new-style prosopography inspired by network analysis operates from the premise that a person's social standing was not merely an outcome of their individual characteristics, but also of their interactions with others. Seeking and gaining recognition from social peers was essential for establishing a person's social position, and the boundaries of social classes were demarcated by the decisions of historical actors to interact with some but not others.

Two articles in this issue have pursued this line of enquiry. They explore the character and identity of elite families in Chinese history by studying the pattern of their interactions. SHANG Wenyi and SANG Zizhou focus on aristocratic life in the fourth century, when war in northern China forced the Jin court (266–420) and some northern aristocratic families to flee southward across the Yangzi River. Throughout the fourth century, the northern émigrés fought bitterly among themselves for domination at court, while leaving only the power in the provinces to the prominent southern families. Using a fifth-century collection that provides snippets of aristocratic life in this era, SHANG and SANG show that the patterns

---

39 Song Chen, "Governing a Multicentered Empire: Prefects and Their Networks in the 1040s and 1210s," in *State Power in China, 900–1325*, ed. Patricia Buckley Ebrey and Paul Jakov Smith (Seattle: University of Washington Press, 2016), 101–52.

40 Dion Smythe makes a distinction between "old" prosopography, which explores each person's external characteristics, and "new-style" prosopography that is equally concerned with the relationships between individuals that enmesh them in overlapping social networks. Dion Smythe, "Prosopography," in *The Oxford Handbook of Byzantine Studies*, ed. Robin Cormack, John F. Haldon, and Elizabeth Jeffreys (New York: Oxford University Press, 2008), 176–81.

of their daily interactions reflect the social distance between different aristocratic families and the gap in their political status. The five most powerful families, all of northern origin, which once dominated court politics also rank highest by all centrality measures. Men from these families were also central figures in seven of the nine major clusters detected by the Louvain method. Moreover, the persistent social chasm that separated the northern émigrés from the southerners is also evident in the much higher intensity of interaction within either group than between the two groups. Nonetheless, the authors argue that close interactions among the aristocrats – despite differences in their political leaning and regional origin – indicate their willingness to recognize each other as social equals, which fostered a degree of cohesion that sustained the aristocracy’s privileged social status and its century-long political dominance.

Similarly, who was and was not a literatus in Yuan times (1279–1368 CE) was also a matter of mutual recognition that was evidenced by literary exchanges. In his case study of Wuzhou (a prefecture in southeast China), Peter K. BOL shows that in the Southern Song (1127–1279 CE), participation in civil service examinations provided a state-sanctioned way for local men to claim themselves as “literati” (*shi*), while marriages between literati families in different counties held them together across the prefecture. In the Yuan, however, the examination system was first abolished and then restored at a far diminished scale and marriages, driven by a growing desire to build alliances with surrounding descent groups, became strictly confined to within county borders. Under these new circumstances, argues BOL, local men in the Yuan relied heavily on learning to build connections with each other that provided the necessary recognition to bolster their claims to literati status.<sup>41</sup> By combining spatial and centrality analyses, BOL further reveals that the literati learning networks in the Southern Song and Yuan differed substantially in size, intensity, and leadership. The Yuan network was larger, more active, and also centered on local men in Wuzhou, instead of influential national figures outside the prefecture. Unlike traditional prosopographical studies of the elite that focus mostly on what they were (e.g., descendants of which clan, natives of which place, whether degree- or office-holders), these articles turn the spotlight on what they did. In other words, these articles look not merely at the elite’s social attributes but also their networking practices – i.e., how they interacted with each other and cultivated close relationships among themselves. These studies share the view, explicitly or not, that the pattern of these interactions and relationships was an outcome of the choices made by the elites, and thus an expression of how they viewed themselves and each other.

---

41 Hilde De Weerdts makes a similar argument about how literati identity was constructed interactively through networks of information exchange during the Southern Song. Hilde De Weerdts, *Information, Territory, and Networks: The Crisis and Maintenance of Empire in Song China* (Cambridge, MA: Harvard University Press, 2015), 392.

This is certainly not to say that the social attributes of historical actors were irrelevant to the pattern of their interactions and relationships. Both articles discussed above have fruitfully explored the interrelation between social attributes and network structures. SHANG and SANG show that aristocratic families in the Eastern Jin clustered by regional origin in their daily interactions, whereas BOL demonstrates that the central figures in the Wuzhou literati learning network had, by Yuan times, become local scholars instead of national celebrities from outside the prefecture. The article by Cécile ARMAND and Christian HENRIOT provides another excellent example, invoking the background characteristics of historical actors to explain observed structural properties in the networks. With the education reforms in early twentieth-century China, the social category of “literati” gradually vanished. China’s most educated men and women now drew on a more diverse set of self-denominations. On closer inspection, however, professionals, bureaucrats, cadres, educators, industrialists, and many other groups did form a diverse, though not segmented, elite. As part of a larger endeavor to reconsider the networks of Republican elites, ARMAND and HENRIOT examine the depiction of eminent men and women in Boorman’s *Biographical Dictionary of Republican China*. They employ Natural Language Processing to trace the co-occurrence of names as textual links between historical actors. Critically engaging with the dictionary’s known biases, they asked what level of elite connectivity can be gleaned from such co-occurrences and whether they reflect historical reality in its full complexity. The results show that in some instances, the editors chose individuals as representatives of social groups, isolating them from other circles. At the same time, besides such smaller ego-networks, a highly connected main component emerged. Centrality measures and clustering methods revealed subgroups within this polycentric network. Even though network research focuses on relationships, this study demonstrates that the exploration of shared attributes should not be neglected in the analysis. To test whether these subgroups shared common traits, the data was further enriched with individual traits, such as provincial origin, military background, foreign education, and party affiliation. The findings show that clusters represent combinations of attributes, supporting the above-mentioned assumption of multiplex elite structures in Republican China.

## 5. Beyond Interpersonal Relationships

We have argued that relationships are not by themselves networks, and that to construe them as such is a modern *Kulturtechnik*, which operates on the premise that relationships constitute a totality with identifiable structural properties and aims to reveal, interpret, and explain these properties. As an analytical approach, therefore, network analysis holds great promise for a wide range of research questions, beyond the reconstruction of social interactions and interpersonal relationships, and may be used to explore the structure of any totality of connections. The first graph-theoretical analysis of networks concerned movement in space. This



was Leonhard Euler's rumination on the Seven Bridges of Königsberg, where each node was a landmass and each edge was a bridge. In the past decades, sociologists and archeologists have also routinely adopted network models to study the relationships between countries, organizations, excavation sites, and even cultural notions and relief practices.<sup>42</sup> Several articles in this issue provide a snapshot of how China historians have used network analysis to investigate a diverse range of topics in institutional and intellectual history. XIONG Huei-Lan studies the career trajectory of Southern Song officials who had an appointment in the Imperial Library. Instead of looking at interpersonal relationships, she constructed a directed network where each node is a government position and each edge represents a transfer between positions.<sup>43</sup> Using modularity analysis and current-flow betweenness centrality, among other algorithms, and drawing on her erudition in the Song institutional history to interpret the results, XIONG argues that appointment to the Imperial Library served as an important stepping stone to high office in the twelfth century, although this practice encountered a brief setback between 1140 and 1155 when Qin Gui 秦檜 (1091–1155), a powerful grand councilor, monopolized power at the Song court. Network analysis on a rich data set that is systematically harvested from diverse historical sources provides solid quantitative evidence for XIONG to evaluate the dynasty's declared policy of using scholarly institutes as the grooming ground for top administrators.

Marilyn LEVINE combines network analysis and prosopographical studies in a different way, exploring the network of persons but constructing connections based on node attributes. Her study focuses on a group that we already know to be highly exclusive and interconnected, namely the leaders of the Chinese Communist Party who received training in the Soviet Union during the early twentieth century ("Soviet Returned Leader," SRL). From archives in France to interviews in Beijing, LEVINE collected an impressive abundance of biographical data on each leader, such as their educational background, careers, political affiliations, and the major events in which they participated. She used these data to construct a network of Chinese revolutionary leaders, where edges represent not interpersonal relationships but the degree of similarity between each pair of persons in their background characteristics. To put this in more technical terms, her one-mode network data are similarity matrices computed from node

---

42 John W. Mohr and Vincent Duquenne, "The Duality of Culture and Practice: Poverty Relief in New York City, 1888–1917," *Theory and Society* 26, no. 2/3 (1997): 305–56. Søren Michael Sindbæk, "The Small World of the Vikings: Networks in Early Medieval Communication and Exchange," *Norwegian Archaeological Review* 40, no. 1 (2007): 59–74.

43 For a study of bureaucracy that employs a similar methodology, see the brief discussion in Franziska Barbara Keller, "Analyses of Elite Networks," *The Palgrave Handbook of Political Elites*, ed. Heinrich Best and John Higley (London: Palgrave Macmillan, 2018), 142–143.

attributes. Using the Louvain algorithm, LEVINE finds that this network partitions neatly into clusters that, with some important exceptions, separate those leaders who traveled only to the Soviet Union (the Soviet group) from others who had also spent some years in Western Europe (the Euro-Soviet group). She shows that leaders in the Euro-Soviet group were on average a few years older than those in the Soviet group. Many of those in the Euro-Soviet group died at the very beginning of China's communist revolution, but of those who survived, many played important roles in the subsequent decades and therefore rank high on nearly all centrality measures. After the founding of the People's Republic of China in 1949, they reemerged as the central figures of a new society and cut ties with those that fled the mainland for Taiwan, creating a schism that has persisted until this day.

Anne CHAO et al. analyze texts, not persons. They also employ network analysis not as a tool of verification but as a device of exploration. This guides CHAO et al. throughout the extensive literature of the twentieth century and alerts them to passages that deserve close reading. To compare the political views of two leading intellectuals in the early twentieth century (Liang Qichao 梁啟超 [1873–1929] and Chen Duxiu 陳獨秀 [1879–1942]) and trace the evolution of their ideas, CHAO et al. constructed several networks where nodes represent words and edges indicate their co-occurrences in Liang's and Chen's writings. Using the Louvain method, they partitioned each network into several clusters, where each cluster signifies a specific topic. Next, they used centrality measures to identify the key terms in each cluster, which they then used as guides for interpreting Liang's and Chen's writings. A comparison of these co-occurrence networks of terms shows that Liang employed a much "richer and more varied vocabulary" in his writings than Chen. This led CHAO et al. to the discovery that although both thinkers were occupied with national salvage, Liang and Chen had different views on nation-building and also adopted different rhetorical strategies. For Liang, nation-building was a great enterprise that encompassed a diversity of interconnected issues, ranging from citizens' responsibilities to institutional checks on governmental power. These issues were tied together in Liang's writings by notions of constitutional rule and popular sovereignty. By contrast, Chen was an avowed adherent of social Darwinism and Marxism and placed emphasis on evolution and class revolution. Unlike Liang, he charted a more specific course of action and drove home his point by repeatedly invoking the same key terms.

## 6. Databases and Tools

Whether to investigate the pattern of interpersonal relationships or explore the structure of any other type of connections, historical network analyses have to be grounded in solid empirical data. Contributions in this issue demonstrate two markedly different approaches that scholars take in building their datasets. Some collect data from a single source. SHANG and SANG rely exclusively on *A New Account of the Tales of the World* (*Shishuo xinyu* 世說新語), a fifth-century collection of historical anecdotes and character sketches, to build their dataset on the interaction among fourth-century aristocrats. Likewise, ARMAND and HENRIOT use only Howard L. Boorman's *Biographical Dictionary of Republican China* when reconstructing the elite networks of twentieth-century China. Other contributors to this special issue, in contrast, glean and synthesize data from a multitude of historical sources. BOL investigates kinship relations and literary exchanges that are evidenced in a variety of literary writings and aggregated into the gigantic China Biographical Database (CBDB). XIONG starts out with service records in the twelfth century and supplements them with an extensive survey of biographical information that is preserved in a plethora of historical texts and databases. The work of LEVINE draws on archival materials in China and Europe, as well as a series of transcribed interviews that she conducted in 1985 and 1990.

Either approach to data collection has its benefits and drawbacks. To combat the danger of cherry-picking sources, Giovanni R. Ruffini argues that a historical network analyst must use data from all sources or data from only one.<sup>44</sup> Any constructed network inevitably reproduces whatever selection bias exists in the source itself, and the historian who relies exclusively on a single source has to face an almost insurmountable challenge to give a convincing argument about whether any structural pattern observed in the constructed network pertains to actual historical relationships or merely the representation of history in the chosen source. Moreover, a single source often fails to provide adequate data for addressing many historical questions. For instance, to investigate the structural and spatial features of Wuzhou men's kinship and intellectual networks, BOL needs data on ancestry, marriage, literary exchanges, and so forth. These data are scattered throughout large numbers of biographies, letters, and other literati writings, and to assemble data from these diverse sources is not a choice but a necessity.

Nonetheless, to collect and synthesize data from a wide range of sources is not without its own methodological challenges. Historians who adopt this approach may find it difficult to assess how the constructed network is distorted by the ensemble of biases embedded in the diverse body of source materials. Do these

---

44 Giovanni R. Ruffini, "An Epilogue. Social Network Analysis and Greco-Roman Politics," *Journal of Historical Network Research* 4 (2020): 335.

biases cancel each other out? Or do they amplify each other? There is no easy answer to these questions. Yet one has to recognize that these are perennially nagging issues that confront all historians, and that they are not unique to those who practice formal network analysis. In traditional historical research, these thorny issues have not led historians to confine themselves to a single source. Rather, a historian's work has typically benefited from piecing together anecdotal information from different sources and from studying a person's life from multiple angles, each of which is often accentuated in a particular genre of historical texts.

Therefore, instead of limiting the range of sources one uses to reconstruct the network, it is perhaps more sensible for historians to do what we have always been good at doing: to stay intellectually vigilant at every step of the analysis and make our decisions transparent. Historical network analysis is an assiduous practice. The analyst starts with defining nodes and edges and converting historical records into datasets. They move on to graph the network, measure its structural properties, and then develop a historical interpretation based on the graphs and metrics. Every step in this process involves judgment. Constant intellectual vigilance requires that the analysts make choices at every step that are informed by their knowledge of possible source biases. They must be fully transparent with these choices and make their dataset available to readers who may want to reevaluate its quality, or use it to test alternative hypotheses and assess the robustness of the findings (e.g., by weighting the ties differently and using different parameter values).

Ruffini dreams of historical network research that uses "data from *all* sources" but implores that this is often "a practical impossibility." He envisions that one day someone will create a "platform...capable of reading any literary, documentary or epigraphic data and generating standardized data-sets of connections customized in response to user-controlled criteria."<sup>45</sup> Many scholars in the China field share his visions and have set out to turn them into reality. The Historical Social Network of Chinese Buddhism project, which Marcus BINGENHEIMER introduces in this special issue, is a large dataset consisting of approximately 17,500 actors and 25,000 links, spanning from the late third century to the early twentieth century. The data was based on thoroughly marked-up biographical literature of Chinese Buddhism and the Buddhist Studies Authority Database.<sup>46</sup> The undisputedly largest and most comprehensive database for Chinese history is the China Biographical Database (CBDB) project discussed by Michael FULLER and WANG Hongsu in this issue. The long-term goal of the CBDB is to collect all significant biographical information from all sources of Chinese history before the twentieth century. As of May 2021, it already contains biographical data

---

45 Ruffini, "An Epilogue. Social Network Analysis and Greco-Roman Politics," 335.

46 On the Buddhist Studies Authority Database, see Buddhist Studies Authority Database Project, "Introduction," <https://authority.dila.edu.tw/> (accessed August 21, 2021).

for approximately 491,000 individuals, spanning from the seventh through nineteenth centuries. By systematically collecting diverse categories of biographical data from different sources and integrating them into a single relational database, the CBDB offers a treasure trove where everyone can set their own querying parameters according to their individual research needs and quickly “check out” a subset of structured data for academic use. More importantly, as it attracts more users and contributors, the CBDB coding practices (e.g., how to code kinship and non-kin social relations) will likely gain greater influence and help build consensus among China historians. In this sense, the CBDB is on its way towards setting a gold standard in the global community of China studies for transforming historical narratives into structured data. Today, the CBDB has been an integral part of a burgeoning digital humanities ecosystem for Chinese studies. Its API bridges the CBDB and more specialized databases (e.g., Ming-Qing Women’s Writings Database<sup>47</sup>); its code tables for offices, places, persons, and so forth are used in online text markup platforms (MARKUS<sup>48</sup> and DocuSky<sup>49</sup>); it exports query results in several formats that can be directly imported into different spatial and network analysis programs; its data are integrated, along with data from several other systems, into the knowledge graph of the Chinese Text Project Data Wiki<sup>50</sup>; and the CBDB has enabled many innovative studies of Chinese history<sup>51</sup> and become a source of inspiration for a growing list of linked open data projects<sup>52</sup>, online data visualization projects, and pedagogical platforms<sup>53</sup> that train the next generation of digital humanists.

For twentieth-century China, however, no equivalent of the CBDB exists. Even though historians can draw on an abundance of published sources from the Republican period (1912–1949 CE) and the People’s Republic of China (since 1949), no comprehensive database has yet emerged. Funding logics favoring the digiti-

- 
- 47 “Mingqing funü zhuzuo” 明清婦女著作 [Ming-Qing Women’s Writings Database], <https://digital.library.mcgill.ca/mingqing/chinese/index.php> (accessed August 21, 2021).
  - 48 Hou leong Brent Ho and Hilde De Weerd, “MARKUS. Text Analysis and Reading Platform,” <https://dh.chinese-empires.eu/markus/beta/> (accessed August 21, 2021).
  - 49 “Shuwei renwen xueshu yanjiu pingtai” 數位人文學術研究平台 [DocuSky Collaboration Platform], <https://docusky.org.tw/> (accessed August 21, 2021).
  - 50 Chinese Text Project, “Linked Open Data and the Semantic Web” <https://ctext.org/tools/linked-open-data> (accessed August 21, 2021).
  - 51 For a list of publications that use CBDB data, see China Biographical Database Project, “Publications that Use CBDB Data,” <https://projects.iq.harvard.edu/cbdb/publications-use-cbdb-data> (accessed August 21, 2021).
  - 52 For CBDB Linked Open Data developed by Shanghai Library, see China Biographical Database Project, “SPARQL Editor,” <https://cbdb.library.sh.cn/sparql> (accessed August 21, 2021).
  - 53 For example, see Tsinghua University’s Digital Humanities Teaching and Research Platform “Tsinghua daxue shuzi renwen jiaoxue yu yanjiu pingtai” 清華大學數字人文教學與研究平臺 [Tsinghua Digital Humanities Teaching and Research Platform], <http://qh.nqcx.net/> (accessed August 21, 2021).

zation of clearly demarcated source bodies and project timeframes have resulted in a number of smaller yet noteworthy projects. One of the pathbreakers in creating biographical databases for the political elites of twentieth-century China is Marilyn LEVINE, who combined data collected from archival and published sources with insights gained from interviewing cadres to form the “Chinese Biographical Database” (CBD). Two projects have produced datasets focusing on progressive women of the first half of the twentieth century. One is Academia Sinica’s (Taiwan) “Authorship of Chinese Women’s Periodicals” (ACWP) database, which has already allowed users to download the data of ego networks.<sup>54</sup> The second project on “Chinese Women’s Magazines in the Late Qing and Early Republican Period” is based at Heidelberg University and has now expanded to include relational data linking not only persons and texts, but also signifying interpersonal social relationships (see Matthias ARNOLD and Henrike RUDOLPH in this issue). Christian Henriot and his team at Aix-Marseille University have started a renewed attempt to create a more comprehensive “Modern China Biographical Database” (MCBD) with an integrated “Modern China Geospatial Database.”<sup>55</sup> These databases will greatly facilitate geospatial analysis of modern Chinese history, which has been impeded by the repeated restructuring of administrative units, changing place names, and urbanization.

Baptiste BLOUIN, Nora van den BOSCH, and Pierre MAGISTRY show in their contribution how they addressed the challenges of processing Chinese-language sources and applied named-entity recognition techniques, which are particularly challenging for Chinese as a logographic system. The resulting lists of named entities, whether of persons, places, or institutions, will facilitate text encoding in future projects that make use of digitized Chinese sources.

## 7. Challenges and Potentials

As a nascent field, the study of Chinese historical networks still faces many challenges. Instead of seeing them as an indication of methodological or theoretical flaws of the network approach, we believe that these challenges demonstrate the potential in this burgeoning field of research and that they call for further intellectual innovation and collaboration.

---

54 Institute of Modern History, Academia Sinica, “Funü qikan zuozhe yanjiu pingtai” 婦女期刊作者研究平臺 [Authorship of Chinese Women’s Periodicals], <http://mhdb.mh.sinica.edu.tw/ACWP/index.php> (accessed September 1, 2021).

55 ENP China, “Modern China Biographical Database,” <https://heurist.huma-num.fr/h6-alpha/?db=ModernChinaBiographicalDatabase&website&id=109237&pageid=109242> (accessed September 1, 2021).



We are yet to realize the full explanatory potential of network analysis in historical research. First, the majority of the present scholarship on Chinese historical networks remains descriptive and static in nature. A great deal of work is devoted to describing the structural features of historical networks at a given moment of history and comparing network structures between different historical periods. These structural features are typically treated as *outcomes* that need to be explained by other historical factors, such as war and migration, institutional arrangements, technological innovations, and cultural shifts. We are yet to fully explore how historical networks may have constrained or empowered individual actors and thereby could provide *explanations for* historical change. Second, we have not adequately explored the dynamic and evolution of historical networks. Few studies discuss how historical actors strategically shaped the network in which they participated or how they mobilized resources in the network to achieve their ends. In brief, the temporal dynamics of historical networks and the explanatory potential of these networks for historical change<sup>56</sup> remain largely uncharted territory in the field of Chinese studies.

Also, we have only started to explore how to best operationalize network concepts in historical studies. When we apply network analysis to historical relationships that are not between natural persons, the question arises as to how we define a “node.” In the network of bureaucratic offices, should we code the magistracy of each county as a node, or should we code the position of magistracy as a node regardless of its geographic jurisdiction? When we take a network approach to intellectual writings, should we code each word as a node, or should we code each concept – which historical writers may have expressed using different terms – as a node? We need to mull over what social theory, empirical studies, and historical research exist to justify the way we classify and weight different kinds of social interactions and relationships (e.g., any theoretical and empirical grounds for weighting kinship relations, teacher-disciple ties, and letter exchanges similarly or differently). Some authors in this issue have begun to address these methodological concerns. They have been forthright about their choices, and many have tested the robustness of their conclusions under alternative definitions and weightings of nodes and edges.

Theory provides more than justification. As XIONG shows with the “four-continent theory,” abstract models are a source of inspiration for historical work. Like Weberian ideal types, they provide useful analytical constructs for conceptualizing historical networks. Yet using theoretical models as a guide for exploring historical networks often poses challenges, not only because historical sources are

---

56 For an example of a historical study that nicely handles some of these issues, see Padgett and Ansell’s work on the Medici family. John F. Padgett and Christopher K. Ansell, “Robust Action and the Rise of the Medici, 1400–1434,” *American Journal of Sociology* 98, no. 6 (1993): 1259–319.

often fragmentary but also because many models were originally constructed as mathematical representations of materially different network phenomena, such as connectivity in the World Wide Web. Moreover, some algorithms for measuring network properties (e.g., Google's PageRank) were first developed to accomplish specific tasks in a very specific type of networks (e.g., to rank web pages discovered by the search engine). Although they are included in some network analysis programs for general use, it is still up for discussion whether these algorithms are meaningful for historical networks and how we should interpret their outputs. These challenges suggest that it may not be adequate to simply "borrow" existing theories and tools for historical analysis. These inadequacies underscore the need for a more constructive dialogue between historians, sociologists, mathematicians, and computer scientists who work on past and present networks, graph theory, and network analysis algorithms.

Interdisciplinary collaboration has been particularly fruitful in increasing the digital availability of Chinese sources. In the past two decades, computer scientists and humanities scholars have made concerted efforts to resolve issues of character recognition, word and sentence segmentation, the detection of reading directions, text markup, and data extraction (for example, see FULLER and WANG in this issue). Although solutions to these issues still leave much room for improvement, historians of China can now delve into vast digital resources and employ digital tools that would surely have made previous generations of scholars envious. For archival sources from the People's Republic of China, however, the picture is somewhat mixed. Here, digitization efforts had adverse effects. The systematic cataloging and scanning of archival sources facilitated state censorship, especially regarding documents that might challenge official narratives of twentieth-century history. Yet even here, network analysis offers new perspectives as it allows scholars to make full use of the available published sources while combining them with the archival sources that remain accessible because they were deemed not worthy of censorship, such as economic statistics or inventories.

## 8. References

- Ahnert, Ruth, Sebastian E. Ahnert, Catherine Nicole Coleman, and Scott B. Weingart. *The Network Turn: Changing Perspectives in the Humanities*. 1st ed. Cambridge: Cambridge University Press, 2020.
- Averill, Stephen. "The Cultural Politics of Local Education in Early Twentieth-Century China." *Twentieth-Century China* 32, no. 2 (2007): 4–32.
- Belsky, Richard. "Placing the Hundred Days: Native-Place Ties and Urban Space." In *Rethinking the 1898 Reform Period: Political and Cultural Change in Late Qing China*, edited by Rebecca E. Karl and Peter Gue Zarrow. Cambridge, MA: Harvard University Press, 2002.

- Böhme, Hartmut. "Netzwerke. Zur Theorie und Geschichte einer Konstruktion." In *Netzwerke: eine Kulturtechnik der Moderne*, edited by Jürgen Barkhoff, 17–36. Köln: Böhlau, 2004.
- Bonnet-Bidaud, Jean-Marc, Françoise Praderie, and Susan Whitfield. "The Dunhuang Chinese Sky: A Comprehensive Study of the Oldest Known Star Atlas." *Journal of Astronomical History and Heritage*, no. 12 (2009): 39–59.
- Bossler, Beverly. *Powerful Relations: Kinship, Status, and the State in Sung China (960–1279)*. Cambridge, MA: Harvard University Asia Center, 1998.
- Buddhist Studies Authority Database Project, "Introduction." <https://authority.dila.edu.tw/> (accessed August 21, 2021).
- Chen, Dong 陳東. "Dengwenjianyuan san shang Qinzong huangdi shu" 登聞檢院三上欽宗皇帝書 [Third Memorial Submitted to Emperor Qinzong through the Public Petition Review Bureau]. In *Quan Song wen* 全宋文, edited by Zeng Zaozhuang 曾棗莊 and Liu Lin 劉琳, 175: 3832.194. Shanghai: Shanghai cishu chubanshe, 2006.
- Chen, Song. "Governing a Multicentered Empire: Prefects and Their Networks in the 1040s and 1210s." In *State Power in China, 900–1325*, edited by Patricia Buckley Ebrey and Paul Jakov Smith, 101–52. Seattle: University of Washington Press, 2016.
- Chen, Qiao 陳峭, Zhou Xiaoling 周曉玲, Zhang Danxuan 張丹璇, and Zhou Yani 周婭妮. "Guanyu goujian Zhongyi 'jingluo tizhi' xueshuo de shexiang" 關於構建中醫“經絡體制”學說的設想 [The Construction of a Theory of "Meridian System" in Chinese Medicine]. *Zhonghua Zhongyiyao zazhi* 中華中醫藥雜誌 33, no. 6 (2018): 2448–51.
- China Biographical Database Project. "Publications that Use CBDB Data." <https://projects.iq.harvard.edu/cbdb/publications-use-cbdb-data> (accessed August 21, 2021).
- China Biographical Database Project. "SPARQL Editor." <https://cbdb.library.sh.cn/sparql> (accessed August 21, 2021).
- De Weerd, Hilde. *Information, Territory, and Networks: The Crisis and Maintenance of Empire in Song China*. Cambridge, MA: Harvard University Press, 2015.
- Ebrey, Patricia. "Patron-Client Relations in the Later Han." *Journal of the American Oriental Society* 103, no. 3 (1983): 533–42.
- Eisenstadt, Shmuel N. "Multiple Modernities." *Daedalus* 129, no. 1 (2000): 1–29.
- Emirbayer, Mustafa, and Jeff Goodwin. "Network Analysis, Culture and the Problem of Agency." *American Journal of Sociology* 99, no. 6 (1994): 1411–54.
- ENP China. "Modern China Biographical Database." <https://heurst.huma-num.fr/h6-alpha/?db=ModernChinaBiographicalDatabase&website&id=109237&pageid=109242> (accessed September 1, 2021).
- Erikson, Emily. "Formalist and Relationalist Theory in Social Network Analysis." *Sociological Theory* 31, no. 3 (2013): 219–42.

- Fewsmith, Joseph. "From Guild to Interest Group: The Transformation of Public and Private in Late Qing China." *Comparative Studies in Society and History* 25, no. 4 (1983): 617–40.
- Fröhlich, Hajo, and Mathias Heinrich. "China – A Network Society? The Network Concept in Research on Chinese History and Society." In *China Networks*, edited by Jens Damm, 3–19. Berliner China-Hefte 35. Berlin: LIT-Verl, 2009.
- Gießmann, Sebastian. *Netze und Netzwerke: Archäologie einer Kulturtechnik, 1740–1840*. Kultur- und Medientheorie. Bielefeld: transcript, 2006.
- Gold, Thomas, Doug Guthrie, and David Wank. "An Introduction to the Study of Guanxi." In *Social Connections in China: Institutions, Culture, and the Changing Nature of Guanxi*, edited by Thomas Gold, Doug Guthrie, and David Wank, 1–20. Cambridge, UK: Cambridge University Press, 2002.
- Hartwell, Robert M. "Demographic, Political, and Social Transformations of China, 750–1550." *Harvard Journal of Asiatic Studies* 42, no. 2 (1982): 365–442.
- Hymes, Robert P. *Statesmen and Gentlemen: The Elite of Fu-chou, Chiang-hsi, in Northern and Southern Sung*. Cambridge: Cambridge University Press, 1986.
- Ho, Hou leong Brent, and Hilde De Weerdt. "MARKUS. Text Analysis and Reading Platform." <https://dh.chinese-empires.eu/markus/beta/> (accessed August 21, 2021).
- Hymes, Robert. "Sung Society and Social Change." In *Cambridge History of China*, Volume 5, Part Two, edited by John W. Chaffee and Denis Twitchett, 526–664. Cambridge: Cambridge University Press, 2015.
- Institute of Modern History, Academia Sinica, "Funü qikan zuozhe yanjiu pingtai" 婦女期刊作者研究平臺 [Authorship of Chinese Women's Periodicals]. <http://mhdb.mh.sinica.edu.tw/ACWP/index.php> (accessed September 1, 2021).
- Johnson, David G. *The Medieval Chinese Oligarchy*. Boulder, Colorado: Westview Press, 1977.
- Keller, Franziska Barbara. "Analyses of Elite Networks." In *The Palgrave Handbook of Political Elites*, edited by Heinrich Best and John Higley, 135–52. London: Palgrave Macmillan, 2018.
- Kipnis, Andrew. "Practices of Guanxi Production and Practices of Ganqing Avoidance." In *Social Connections in China: Institutions, Culture, and the Changing Nature of Guanxi*, edited by Thomas Gold, Doug Guthrie, and David Wank, 21–36. Cambridge: Cambridge University Press, 2002.
- Kipnis, Andrew B. *Producing Guanxi: Sentiment, Self, and Subculture in a North China Village*. Durham, NC: Duke University Press, 1997.
- Lewis, Mark Edward. *China between Empires: The Northern and Southern Dynasties*. Cambridge, MA: Belknap Press of Harvard University Press, 2009.

- Luo, Jiade, and Yong Zhou. *Social Networking and Chinese Indigenous Management*. Reading, UK: Paths International Ltd, 2014.
- “Mingqing funü zhuzuo” 明清婦女著作數字 [Ming-Qing Women’s Writings]. <https://digital.library.mcgill.ca/mingqing/chinese/index.php> (accessed August 21, 2021).
- Mohr, John W., and Vincent Duquenne. “The Duality of Culture and Practice: Poverty Relief in New York City, 1888–1917,” *Theory and Society* 26, no. 2/3 (1997): 305–56.
- Osterhammel, Jürgen. *The Transformation of the World: A Global History of the Nineteenth Century*. America in the World. Princeton: Princeton University Press, 2014.
- Padgett, John F., and Christopher K. Ansell. “Robust Action and the Rise of the Medici, 1400–1434,” *American Journal of Sociology* 98, no. 6 (1993): 1259–319.
- Rollinger, Christian. “Prolegomena. Problems and Perspectives of Historical Network Research and Ancient History,” *Journal of Historical Network Research* 4 (2020): 1–35.
- Ruffini, Giovanni R. “An Epilogue. Social Network Analysis and Greco-Roman Politics,” *Journal of Historical Network Research* 4 (2020): 325–339.
- Schulte, Barbara. “Webs of Borrowing and Lending: Social Networks in Vocational Education in Republican China.” In *World Yearbook of Education 2012*, edited by Gita Steiner-Khamsi and Florian Waldow, 115–38. London: Routledge, 2012.
- “Shuwei renwen xueshu yanjiu pingtai” 數位人文學術研究平台 [DocuSky Collaboration Platform]. <https://docusky.org.tw/> (accessed August 21, 2021).
- Sindbæk, Søren Michael. “The Small World of the Vikings: Networks in Early Medieval Communication and Exchange,” *Norwegian Archaeological Review* 40, no. 1 (2007): 59–74.
- Smythe, Dion. “Prosopography.” In *The Oxford Handbook of Byzantine Studies*, edited by Robin Cormack, John F. Haldon, and Elizabeth Jeffreys, 176–81. New York: Oxford University Press, 2008.
- Stone, Lawrence. “Prosopography,” *Daedalus* 100, no. 1 (1971): 46–71.
- Sun, Guodong 孫國棟. “Tang Song zhi ji shehui mendi zhi xiaorong” 唐宋之際社會門第之消融 [The Dissipation of Prominent Families in the Society of Tang and Song Times: A Study of Social Transformations in Tang and Song Times]. *Xinya xuebao* 新亞學報 4, no. 1 (1959): 211–304.
- Tackett, Nicolas. *The Destruction of the Medieval Chinese Aristocracy*. Cambridge, MA: Harvard University Asia Center, 2014.
- “Tsinghua daxue shuzi renwen jiaoxue yu yanjiu pingtai” 清華大學數字人文教學與研究平臺 [Tsinghua Digital Humanities Teaching and Research Platform]. <http://qh.nqcx.net/> (accessed August 21, 2021).
- Watts, Duncan J. *Six Degrees: The Science of a Connected Age*. New York and London: W. W. Norton and Company, 2004.

- Zeng, Zaozhuang 曾棗莊, and Liu Lin 劉琳, eds. *Quan Song wen* 全宋文 [Complete Collection of the Song-Dynasty Prose]. Shanghai: Shanghai cishu chubanshe, 2006.
- Zishan 紫珊. “Dianwangluo (NETWORK) de jiefa” 電網絡 (NETWORK) 的解法 [Solutions to Electric Networks]. *Kangzhan kexue* 抗戰科學, no. 3 (1939): 36–37.



The Journal of  
**HISTORICAL  
NETWORK  
RESEARCH**

SHANG WENYI/SANG ZIZHOU

# Solidity in a Turbulent Flow: The Social Network of Aristocratic Families in the Eastern Jin Dynasty (317–420 C.E.)

*Journal of Historical Network Research* 5 (2021) 1–32

**Keywords** Aristocratic society, Eastern Jin dynasty, *A New Account of the Tales of the World*, Social network analysis

**Abstract** Historians have characterized the Eastern Jin (317–420 C.E.) as an era dominated by aristocratic families. Network analysis provides a unique perspective to understand these families by expanding the scope of research from imperial politics to the aristocrats' social world. This article reconstructs the aristocratic social network from *A New Account of the Tales of the World* (*Shishuo xinyu* 世說新語), a collection of anecdotes that provides rich information on the aristocrats' social interactions in this period, and provides a comprehensive analysis of the network. This article uses network simulation and Exponential Random Graph Models (ERGM) to describe the structure of this network. It calculates cumulative edge weights to measure the homophily by family affiliation in the network and computes several centrality measures to evaluate the importance of various historical figures. These results demonstrate that homophily by family affiliation exists in this network, but is weak. Although members of the same family tend to have close ties, social relationships across aristocratic families were also common. In contrast to many existing studies that emphasize the conflicts between aristocratic families, this article argues that there was a notable degree of social cohesion that united the aristocratic families as a status group during the Eastern Jin. This cohesion provides an element of solidity in the turbulent flow of their political fortunes, which was one of the main reasons why, as a privileged status group, these aristocratic families successfully dominated politics and social life throughout the century of the Eastern Jin.

## 1. Introduction\*

In the early fourth century, the Jin court fled south after losing control over northern China. There it reestablished itself and ruled for another century. The period, known as the Eastern Jin (317–420 C.E.), was characterized by historians as a “golden age of the great families”, since these families monopolized political power under the formal rule of an actually “powerless imperial line.”<sup>1</sup> Tian Yuqing called this “the politics of powerful families” (*menfa zhengzhi* 門閥政治). According to Tian, “the politics of powerful families in the strict sense existed only during the Eastern Jin in south China. Neither its predecessor, the Wu regime (222–280 C.E.), nor its successor, the Southern dynasties (420–589 C.E.), satisfied the strict definition, whereas in north China, the politics of powerful families never existed.”<sup>2</sup>

Naitō Konan claimed that these powerful families constituted an “aristocratic class,” characterized by an enduring lineage, monopoly over office, and marriage within its own group.<sup>3</sup> However, Dennis Grafflin has argued that an aristocracy such as that described by Naitō did not exist, since only the Wang family

---

\* **Acknowledgements:** The data used in this article is a slightly modified version of the data manually collected for a previous article of the first author. That previous article investigates a similar topic, the social network of *A New Account of the Tales of the World*, but focuses on aspects substantially different from those in the present work. See Shang Wenyi 尚聞一 and Yu Zixuan 于子軒, “Dongjin guizuzhi shehui de wending jizhi – Shishuoxinyu gongci fenxi” 東晉貴族制社會的穩定機制 – 《世說新語》共詞分析 [The Stability Mechanism of the Aristocratic Society in the Eastern Jin Dynasty: A Co-Word Analysis of *A New Account of the Tales of the World*], *Tushuguan luntan* 圖書館論壇, no. 1 (2019): 46–57. We thank Yu Zixuan 于子軒 for his significant contribution to that previous work, which also inspired this article. We would also like to express our gratitude to Professor Ted Underwood, Professor Jana Diesner, and Ly Dinh for their insightful advice on this paper.

**Corresponding author:** Wenyi Shang, School of Information Sciences, University of Illinois at Urbana-Champaign, 501 E. Daniel St. Champaign, IL, 61820, USA. Email: wenyis3@illinois.edu

- 1 Mark Edward Lewis, *China Between Empires: The Northern and Southern Dynasties*, vol. 2, History of Imperial China (Cambridge, MA and London, UK: The Belknap Press of Harvard University Press, 2009), 51–52.
- 2 Tian Yuqing 田余慶, *Dongjin menfa zhengzhi* 東晉門閥政治 [The Politics of the Powerful Families during the Eastern Jin] (Beijing: Beijingdaxue chubanshe, 2012), 1–2.
- 3 Naitō Konan 內藤湖南, “Gaikuoxing de Tang Song shidai guan” 概括性的唐宋時代觀 [General Views on the Tang-Song Period], in *Dongyang wenhuashi yanjiu* 東洋文化史研究 [A Study of Oriental Cultural History], trans. Lin Xiaoguang 林曉光 (Shanghai: Fudan daxue chubanshe, 2016), 103–112. On Naitō’s theory on this period of Chinese society, see Naitō Konan 內藤湖南, *Zhongguoshi tonglun – Neiteng Hunan Boshi Zhongguo shixue zhuzuo xuanyi* 中國史通論 – 內藤湖南博士中國史學著作選譯 [General Comments of Chinese History – Selected Translations of Dr. Naitō Konan’s Works on Chinese History], trans. Xia Yingyuan 夏應元, et al. (Beijing: Shehui kexue wenxian chubanshe, 2004), 227–312. A sophisticated examination of Naitō’s works and Japanese sinology in the late



of Langya (Langya Wangshi 琅琊王氏) “can be described as perennially important.” To Grafflin, “a social paradigm that can only with the aid of self-contradictory qualifications be applied to even a single case is worse than no paradigm at all.”<sup>4</sup> Grafflin’s argument was based on a remarkable reconstruction of the genealogies of the great families, and therefore empirically solid. Nevertheless, the swift rise and fall of individual families does not necessarily disprove the existence of an aristocracy as a far more enduring status group with a high degree of internal cohesion.

Tian argued that neither the imperial clan nor the aristocratic families could establish a stable government on their own. “Only when imperial power ruled along with the powerful families could balance and order be maintained.” None of the aristocratic families in power managed to break this pattern.<sup>5</sup> This suggests that in the turbulent flow of the politics in the Eastern Jin, a certain degree of solidity ensured the aristocratic families’ grip on politics. Kawakatsu Yoshio provided an additional perspective by expanding the scope of analysis from imperial politics to social conditions in general. He described the Eastern Jin as an “aristocratic society,” characterized by “a social system where a class of aristocrats or powerful families was widely present and held a pivotal position in all realms – political, social, economic, and cultural.”<sup>6</sup> Therefore, the aristocracy in the Eastern Jin can be investigated not only from the perspective of political history, but also from that of social history.

Building on Tian and Kawakatsu, this research investigates the social history of aristocratic society<sup>7</sup> in the Eastern Jin. Our source material is *A New Account of the Tales of the World* (*Shishuo Xinyu* 世說新語, henceforth *The Tales*

---

19th and early 20th century can be found in Joshua A. Fogel, *Politics and Sinology: the Case of Naitō Konan (1866–1934)* (Cambridge, MA: Harvard University Press, 1984).

4 Dennis Grafflin, “The Great Family in Medieval South China,” *Harvard Journal of Asiatic Studies* 41, no. 1 (1981): 70–71.

5 Tian, *Dongjin menfa zhengzhi*, 329.

6 Kawakatsu Yoshio 川勝義雄, *Liuchao guizuzhi shehui yanjiu* 六朝貴族制社會研究 [A Study on Aristocratic Society in the Six Dynasties], trans. Xu Gupeng 徐谷芃 and Li Jicang 李濟滄 (Shanghai: Shanghai guji chubanshe, 2007), 53.

7 The ‘aristocracy’ in medieval China was substantially different from the ‘aristocracy’ in medieval Europe, although they share many common characteristics. In this article, we follow the convention of using “aristocratic families” to call the “great families” that enjoyed political, economic, and cultural privileges during the Eastern Jin. Membership of a great family is determined on this basis of a person’s patriline. For more detailed discussions of the concepts of ‘aristocracy’ and ‘aristocratic families’ in imperial China, see Nicolas Tackett, *The Destruction of the Medieval Chinese Aristocracy* (Cambridge, MA: Harvard University Asia Center, 2014), and Patricia Ebrey, *The Aristocratic Families of Imperial China: A Case Study of the Po-Ling Ts’ui Family* (Cambridge, UK and New York, NY: Cambridge University Press, 1978), among others.

of the World)<sup>8</sup>, a collection of 1,130 historical anecdotes compiled and edited in the 5th century, immediately following the fall of the Eastern Jin. Although the reliability of this text is not without debate, many historians agree that it has value as a historical record and was a major source for the official dynastic history, the *Book of Jin* (*Jinshu* 晉書), compiled in the seventh century. Furthermore, with its focus on social relationships among the aristocrats, this text has unique value as a supplement to the *Book of Jin*, which focuses on political and military events.

*The Tales of the World* is also particularly suited for a study of the aristocratic society in the Eastern Jin from a network perspective. Each anecdote in the book narrates an event (usually several sentences in length) involving historical figures and their relationships, which can be easily represented by nodes and edges. This social network provides a unique treasure trove for a comprehensive investigation of aristocratic society.

In this research, we first reconstruct undirected social networks (both weighted and unweighted) from the social relationships between historical figures described in the anecdotes in *The Tales of the World*. Some relationships are positive (i.e., two people engage in favorable interactions or show affirmative attitudes), and others negative (i.e., two people come into conflict or show hostile attitudes). In this study, only positive relationships are used for constructing the networks. In the next step, we analyze these networks at three different levels to address the following questions:

- 1) At the network level, we use a network evolution model to simulate the process that generated the structure of the network and use a node attribute – each person's 'family affiliation' – to predict the formation of the network. We then visualize the network and detect communities in it, which leads to analyses at the group level.
- 2) At the group level, we calculate the cumulative weights of edges that link historical figures from different aristocratic families. We pay particular attention to the best-connected families, and examine whether connections within each family were more intensive than connections between families.
- 3) At the node level, we ask who the most important historical figures were in terms of degree centrality, eigenvector centrality, betweenness centrality, and closeness centrality, and which families they were from.

Based on the above investigations, this article furthers our understanding of the Eastern Jin aristocracy by quantifying the correlation between the aristocrats' so-

---

8 The text has been translated in full into English, see Liu I-ch'ing 劉義慶, *Shih-shuo Hsin-yü: A New Account of Tales of the World*, trans. Richard B. Mather, Second Edition (Ann Arbor, MI: Center for Chinese Studies, The University of Michigan, 2002).

cial relationships and their family affiliations, thereby exploring the solidity in the relationships of the Eastern Jin aristocratic families. Here, ‘solidity’ refers to the social cohesion that bound different aristocratic families together as a status group. If our study shows a robust correlation between the aristocrats’ social relationships and their family affiliations, it suggests that such solidity did not exist, because strong homophily by family affiliation indicates clear boundaries between families. On the contrary, if our results demonstrate a stable, more heterogeneous network (i.e., a weak homophily by family affiliation), it suggests that different families in the aristocratic class bonded together as a self-contained social circle.

## 2. Background

### 2.1 Aristocratic Society in the Eastern Jin Dynasty

The aristocracy in the Eastern Jin has always been an important field of research in Chinese historical studies. Naitō Konan characterized medieval Chinese politics as being dominated by “powerful families,” which he called the “aristocracy.”<sup>9</sup> Kawakatsu Yoshio further investigated the social character of this aristocracy. He argued that this aristocratic society was a result of the Eastern Jin regime’s military origin<sup>10</sup> and was maintained through patron-client relations.<sup>11</sup> Tanigawa Michio adopted a similar perspective in his discussion of medieval Chinese society. He described the aristocratic society of this period as an “autonomous world” whose cohesiveness was ensured by a “communitarian” bond.<sup>12</sup> Tian Yuqing described the Eastern Jin politics as “the politics of powerful families.” Five powerful families took turns to dominate the government: the Wang family of Langya, the Yu family of Yingchuan 潁川庾氏, the Huan family of Qiaoguo 譙國桓氏, the Xie family of Chenjun 陳郡謝氏, and the Wang family of Taiyuan 太原王氏. Additionally, the Xi family of Gaoping 高平郗氏 also exerted a significant influence on politics.<sup>13</sup>

The solid strength of this aristocracy was deeply rooted in the dynasty’s procedures of selecting officials. A practice known as the “Nine-Rank System” (*jiu-pin guanren fa* 九品官人法), meticulously studied by Miyazaki Ichisada, assigned each family into one of nine grades and gave families of higher grades exclu-

---

9 Naitō, *Zhongguoshi tonglun – Neiteng Hunan Boshi Zhongguo shixue zhuzuo xuanyi*, 227–312.

10 Kawakatsu, *Liuchao guizuzhi shehui yanjiu*, 154–186.

11 Kawakatsu, *Liuchao guizuzhi shehui yanjiu*, 187–220.

12 Tanigawa Michio 谷川道雄, *Medieval Chinese Society and the Local “Community”*, trans. Joshua A. Fogel (Berkeley, CA: University of California Press, 1985), 100–19.

13 Tian, *Dongjin menfa zhengzhi*, 316.

sive access to the upper echelons of government.<sup>14</sup> This conferred an advantage on aristocratic families. In his statistical analyses of 4,137 officials of the Jin (265–420 C.E.) and the Southern and Northern Dynasties (420–589 C.E.), Mao Hanguang found that about 70% of them came from aristocratic families, about 20% were from minor clans, while only about 10% were from families of humble origins.<sup>15</sup> In addition to the institutional underpinnings, Tang Zhangru argued that the aristocratic families also enjoyed economic privileges as large landowners with large numbers of retainers and dependents.<sup>16</sup> In terms of cultural prestige, Yu Yingshi maintained that “Wei (220–265 C.E.), Jin (265–420 C.E.), and the Southern and Northern Dynasties (420–589 C.E.) can be considered as a prosperous era for Confucianism which placed the ‘family’ at the center. This was a result of the actual need of the aristocratic society.”<sup>17</sup> Mark Edward Lewis argued that the aristocratic families wrapped themselves “in a refined style of philosophical quietude, nominal eremitism, and literary attainments” as badges of “true nobility.”<sup>18</sup>

Despite the consensus surrounding the aristocracy’s central role in the Eastern Jin, the question of when the aristocratic families first appeared in Chinese history, and whether they continuously held on to political power for centuries, remains controversial. Mao Hanguang argued that it was in the Han (206 B.C.–220 C.E.) that the great families of medieval China first rose to prominence.<sup>19</sup> Drawing on his analysis of several lists of great families that have survived, David Johnson further suggested that these families formed an “oligarchy” in medieval China. He maintained that “the social stratum...had been in existence for at least

- 
- 14 Miyazaki Ichisada 宮崎市定, *Jiupin guanren fa yanjiu: Keju qianshi* 九品官人法研究: 科舉前史 [Research on the Method of Selecting Officials According to the Nine Ranks: The Prehistory of the Examination System], trans. Han Sheng 韓昇 and Liu Jianying 劉建英 (Beijing: Zhonghua shuju, 2008).
  - 15 Mao Hanguang 毛漢光, “Liangjin Nanbeichao zhuyao wenguan shizu chengfen de tongjifenxi yu bijiao” 兩晉南北朝主要文官士族成分的統計分析與比較 [A Statistical Analysis and Comparison of the Family Affiliations of Major Officials in the Two Halves of Jin and the Southern and Northern Dynasties], in *Zhongguo zhonggu shehuishi lun* 中國中古社會史論 [Essays on the Social History of Medieval China] (Shanghai: Shanghai shudian chubanshe, 2002), 141–86.
  - 16 Tang Zhangru 唐長孺, *Wei Jin Nanbeichao shilun shiyi* 魏晉南北朝史論拾遺 [Miscellaneous Treatises on the History of Wei, Jin, and the Southern and Northern Dynasties] (Beijing: Zhonghua shuju, 1983), 5–11.
  - 17 Yu Yingshi 余英時, *Zhongguo zhishi jieceng shilun (gudai pian)* 中國知識階層史論 (古代篇) [A Historical Treatise on Chinese Intelligentsia (Ancient Times)] (Taipei: Lianjing chuban shiye gongsi, 1980), 326.
  - 18 Lewis, *China between Empires: The Northern and Southern Dynasties*, 52–53.
  - 19 Mao Hanguang 毛漢光, “Zhonggu da shizu zhi ge'an yanjiu – Langya Wangshi” [A Case Study of the Great Aristocratic Families in Medieval China: The Wang Family of Langya] 中古大士族之個案研究 – 琅琊王氏, in *Zhongguo zhonggu shehuishi lun* 中國中古社會史論 [Essays on the Social History of Medieval China] (Shanghai: Shanghai shudian chubanshe, 2002), 365–404.

four or five centuries, and that during that period its membership had remained remarkably stable.”<sup>20</sup> However, Dennis Grafflin opposed this view and denied the existence of an aristocracy composed of perennially important great families.<sup>21</sup> Tian Yuqing showed that of the five most powerful families that once dominated the top echelons of government, none could trace their origins to the great families in the Eastern Han (25–220 C.E.).<sup>22</sup> In a case study of Huan Xuan 桓玄, a member of the aristocratic Huan family of Qiaoguo who rebelled against the imperial government of Jin and received support from other aristocratic families, Zhu Zongbin pointed out that “the high-ranking aristocratic families were generally in favor of Huan Xuan’s attempt to replace Jin and resorted to him for protection of their privileges and solution of the crisis they faced in their times.”<sup>23</sup> From his point of view, Huan Xuan’s rebellion was actually the last attempt of the aristocratic families to stay in power, which implies a certain degree of social cohesion among the aristocratic families as a status group.

Last but not least, since the Eastern Jin dynasty was a regime of northern elites that ruled southern China, the complex relationship between the northern émigré families (*qiaoxing shizu* 僑姓士族) and the native southern ones (*wuxing shizu* 吳姓士族<sup>24</sup>) has also attracted much attention. Chen Yinke noted that the rise of powerful native southern families with a strong military force exerted a significant influence on the Eastern Jin politics.<sup>25</sup> He went as far as to argue that the reconciliation between the northern émigré families and the native southern ones was the most significant accomplishment of Wang Dao 王導, the crucial political figure of the early Eastern Jin.<sup>26</sup> Kawakatsu Yoshio likewise emphasized the importance of the native southern families and maintained that in the early East-

- 
- 20 David G. Johnson, *The Medieval Chinese Oligarchy* (Boulder [CO]: Westview Press, 1977), 126.
- 21 Grafflin, “The Great Family in Medieval South China”.
- 22 Tian, *Dongjin menfa zhengzhi*, 318.
- 23 Zhu Zongbin 祝總斌, “Shilun Dongjin houqi gaoji shizu zhi moluo ji Huan Xuan dai Jin zhi xingzhi” 試論東晉後期高級士族之沒落及桓玄代晉之性質 [On the Decline of the High Aristocratic Families and the Nature of Huan Xuan’s Replacement of Jin in the Late Eastern Jin], *Beijingdaxue xuebao (zhexue shehui kexue ban)* 北京大學學報 (哲學社會科學版), no. 3 (1985): 85.
- 24 Literally, the aristocratic families from the Wu 吳 area (rough the lower Yangzi region), the locus of the Eastern Jin court.
- 25 Chen Yinke 陳寅恪, “Weishu Sima Rui zhuan Jiangdong minzu tiao shizheng ji tuilun” 魏書司馬睿傳江東民族條釋證及推論 [An Explanation and Inference Concerning the Section on Ethnicities in the Jiangdong Area in Biography of Sima Rui in the *Book of Wei*], in *Jinmingguan congkao chubian* 金明館叢稿初編 [Writings from Jinmingguan, Vol. 1] (Shanghai: Shanghai guji chubanshe, 1980), 78–119.
- 26 Chen Yinke 陳寅恪, “Shu Dong Jin Wang Dao zhi gongye” 述東晉王導之功業 [On the Accomplishments of Wang Dao in the Eastern Jin], in *Jinmingguan congkao chubian* 金明館叢稿初編 [Writings from Jinmingguan, Vol. 1] (Shanghai: Shanghai guji chubanshe, 1980), 55–77.

ern Jin, the dynasty had to rely on the armies of these southern families to build its military foundation.<sup>27</sup> But the relationship between the northern émigrés and the southern natives was far more complicated than mere cooperation. Kawakatsu Yoshio pointed out that Wang Dun 王敦, a general from a northern émigré family (the Wang family of Langya), suppressed the forces of the native southern families that had once cooperated with him “to prevent them from gathering too much strength.”<sup>28</sup> On the other hand, when Wang Dun himself revolted against the court, some other native southern aristocrats nevertheless joined him “to air their grievances.”<sup>29</sup> In his interpretation of this event, Tang Zhangru further claimed that the native southern families participated in the rebellion of Wang Dun because Wang’s opposition to imperial policy helped protect the privileges of these southern families.<sup>30</sup>

To conclude, previous scholars have discussed the importance of the aristocratic families in the Eastern Jin dynasty, debated their origins, explored the continuity of their political dominance, and analyzed the divide between the northern émigré families and the native southern ones. These arguments have significantly contributed to our understanding of the aristocratic society in the Eastern Jin. However, we still lack a deeper understanding of complex social relationships, for which the methods of historical network analysis provided an enticing solution.

## 2.2 Historical Network Analysis

As noted in the introductory section, this article analyzes the aristocratic social network, constructed from *The Tales of the World*, at three levels (network, group, and node). In what follows, we will briefly discuss the methods used at each level of analysis and review the pertinent scholarship that has adopted similar approaches. At the network level, we employ statistical methods of SNA. Examples of such methods include Quadratic Assignment Procedure (QAP) and Exponential Random Graph Models (ERGM), the basic ideas of which are hypothesis testing and logistic regression of the network models. These methods provide convincing quantitative evidence and are attracting growing attention in historical network research. For example, Andrew Schauf and Miguel Escobar Varela analyzed the co-occurrence network of characters in Javanese *wayang kulit* (a shadow puppet theatre tradition) using betweenness centrality and closeness centrality. Adopting the method of QAP, they generated artificial networks by re-

---

27 Kawakatsu, *Liuchao guizuzhi shehui yanjiu*, 164. For Kawakatsu’s detailed discussions on this issue, see Kawakatsu, *Liuchao guizuzhi shehui yanjiu*, 154–86.

28 Kawakatsu, *Liuchao guizuzhi shehui yanjiu*, 164.

29 Kawakatsu, *Liuchao guizuzhi shehui yanjiu*, 171.

30 Tang, *Wei Jin Nanbeichao shilun shiyi*, 151–67.

wiring the connections in the empirical network, which allowed them to identify those “nodes whose centrality measures are higher or lower than expected given the overall strengths of their connections.”<sup>31</sup> A typical use of ERGM can be found in a study that reconstructed science networks of the malacological authors in the mid-19<sup>th</sup> century. The article interpreted the eponyms as “immaterial goods that resemble the properties of regular social contact” and utilized ERGM to explore these connections. With ERGM-based network analysis, the article revealed the network’s endogenous structure and its homophilic tendency.<sup>32</sup>

For analysis at the group level, we calculate the cumulative weights for edges between different families. We also create visualizations of the network and analyze its structure using the technique of community detection. Partitioning the network into different communities helps reveal its structural properties, such as the degree of homophily by family affiliation. In her study of Petronius’ *Satyricon*, Elena Köstner analyzed the interpersonal interactions related to the last will of the character Trimalchio.<sup>33</sup> In another example, Tom Brughmans and Matt Peeples conducted a bibliometric analysis of a large corpus of publications on the topic between 1965 and 2016 and used community detection to uncover the trends in archaeological network research. They used the Louvain clustering algorithm to identify groups of authors in the co-authorship network.<sup>34</sup> Similarly, our previous work on the social network of *The Tales of the World* also used the Louvain clustering algorithm to detect communities in the aristocratic society of the Eastern Jin.<sup>35</sup>

Finally, at the node level, we use centrality measures to evaluate the importance of different historical figures. In a study of Carib attacks on colonial forces in the sixteenth and seventeenth century, the authors calculated eigenvector centrality from a binary incidence matrix that has data on the attackers, the attacked colonial forces, as well as the locations and years of the attacks.<sup>36</sup> Focusing on the events of a similar era, Aline Deicke investigated the intra-Protestant controversies using polemical pamphlets from the late sixteenth century. She

---

31 Andrew Schauf, and Miguel Escobar Varela, “Searching for Hidden Bridges in Co-Occurrence Networks from Javanese *Wayang Kulit*,” *Journal of Historical Network Research* 2 (2018): 26–52.

32 Abraham S. H. Breure, and Raphael H. Heiberger, “Reconstructing Science Networks from the Past. Eponyms Between Malacological Authors in the Mid-19th Century,” *Journal of Historical Network Research* 3 (2019): 92–117.

33 Elena Köstner, “Trimalchio’s Last Will: Shifting Interactions Between Seeming and Being,” *Journal of Historical Network Research* 3 (2019): 1–29.

34 Tom Brughmans, and Matt Peeples, “Trends in Archaeological Network Research: A Bibliometric Analysis,” *Journal of Historical Network Research* 1 (2017): 1–24.

35 Shang, and Yu, “Dongjin guizuzhi shehui de wending jizhi – Shishuoxinyu gongci fenxi”.

36 Termeh Shafie, *et al.*, “Hypergraph Representations: A Study of Carib Attacks on Colonial Forces (AD 1509–1700),” *Journal of Historical Network Research* 1 (2017): 52–70.

constructed a directed network between the authors of the pamphlets and their opponents and calculated various centrality measures.<sup>37</sup>

In summary, scholars have various methods to analyze historical networks at different levels. Inspired by this body of scholarship, this article combines a variety of methods to analyze the aristocratic social network reconstructed from *The Tales of the World*. We aim to answer whether the pattern of the aristocrats' social relationships is correlated with their family affiliations.

### 3. Data and Method

#### 3.1 Data Collection and Preprocessing

Our data on social relationships between the aristocrats are extracted from the historical anecdotes in *The Tales of the World*. The book contains 1,130 anecdotes in total, and each anecdote recounts a distinct event. Therefore, we treat each anecdote as an independent unit of analysis and extract data from it based on the co-occurrences of named persons (i.e., two or more persons from the Eastern Jin mentioned in the same anecdote) as the first step of data collection. In those cases where more than two persons are mentioned in a single anecdote, we create a tie between each pair of these persons. In the next step, we examine the nature of the relationship between each pair of historical figures and filter out all cases of negative relationships. What remains are a total of 484 anecdotes that give us 736 instances of positive relationships between 250 persons.

Since our source text frequently refers to the same person by different names (e.g., courtesy names, Dharma names, etc.), in the third step, we disambiguate and standardize the names of these historical figures. In the final step, we manually assign each instance of relationship into one of two categories: (a) strong positive (i.e., when two people engage in direct interaction, such as bureaucratic appointment, and gift-giving); (b) weak positive (i.e., when two people show respectful and affirmative attitudes towards each other or when they are said to have had positive interactions in conversations, or as colleagues). We assign a weight of 3 to strong positive relationships and 1 to weak positive relationships.<sup>38</sup> We then sum up the weights of all reported relationships between each pair of historical figures and use this as the final strength of their relationship. In summary, our data collection starts with finding name co-occurrences, but the above

---

37 Aline Deicke, "Networks of Conflict: Analyzing the 'Culture of Controversy' of Polemical Pamphlets of Intra-Protestant Disputes (AD 1548–1580)," *Journal of Historical Network Research* 1 (2017): 71–105.

38 Here we give strong positive relationships a weight of 3 (instead of 2) in order to magnify the difference between 'strong' and 'weak' relationships.



steps of data processing ensure that the final data set we obtain is not merely a representation of textual co-occurrences but reflects different types of ‘actual’ social relationships. In other words, the network we have constructed is not semantic, but rather one that is sociologically and historically meaningful.

The following excerpt provides an example of a historical anecdote from *The Tales of the World*. In our interpretation, this anecdote reports a strong positive relationship between Wang Meng 王濛 and Sima Yu (Ssu-ma Yu) 司馬昱, because the latter ‘appointed’ the former:

(In about 345) when Wang Meng had requested the post of grand warden of Tung-yang (Chekiang), the General Governing the Army (Ssu-ma Yu) had not used him. Later, after he became critically ill and was about to die (347), the general, sighing with grief, said, “I’m obligated to Wang Meng in this matter,” and issued an order to employ him. Wang Meng said, “People say the Prince of K’uai-chi (Ssu-ma Yu) is an idiot. He really is an idiot!”<sup>39</sup>

In brief, treating historical figures as nodes and their relationships as edges, we have constructed a network that contains 250 nodes connected by 505 edges. We work with both a weighted and an unweighted version of this network for different levels of analysis. In the weighted version, an edge between two nodes is weighted by the sum of the strength of all reported instances of their relationships. In the unweighted version, edge data is binary (either 1 when a tie exists or 0 when it is absent).

We have also added ‘family affiliation’ as an attribute for each node (person) in the network. Using historical records, we have assigned 195 of the total 250 people into 62 aristocratic families. Each family is labeled by a combination of their home prefecture (choronym) and surname.<sup>40</sup>

---

39 Liu, *Shih-shuo Hsin-yü: A New Account of Tales of the World*, 183.

40 A family must be widely accepted by contemporaries as an ‘prominent family’ (*gaomen* 高門) in the Eastern Jin to be categorized by us as an aristocratic family. We use the home prefecture (*junguo* 郡國), instead of the home county (*xian* 縣), as a family’s hometown (e.g., the Huan family of Qiaoguo, instead of the Huan family of Longkang 龍亢桓氏). These prefectures and counties are based on the administrative divisions in the Eastern Jin dynasty.

### 3.2 Data Analysis

After we completed the data collection, we conducted three rounds of analyses to address the three aforementioned research questions.

#### *Network evolution model and exponential random graph models*

At the network level of analysis, we used the unweighted network because network evolution models generally do not take edge weights into account. First, to understand the dynamics that generated the structure of the network, we used an evolving network model to simulate the evolution of the social network that we constructed from *The Tales of the World*. Inspired by the results of community detection in our previous study of a similar network that we constructed from the same source text, where the number of people in each community follows a long-tail distribution (17% of the communities contained 82% of all the people in the network),<sup>41</sup> we hypothesize that the preferential attachment model<sup>42</sup> may be suitable for simulating the evolution of this network. This is also a reasonable hypothesis given that people with more connections are generally more likely to develop more connections with others.

Specifically, we simulated the preferential attachment process, where the probability of any given node acquiring a new edge is proportional to the number of edges it already possesses.<sup>43</sup> We set the total number of nodes to 250 (that is, the total number of nodes we have in the observed network), adjusted the parameters to control the generation rate of new edges, and replicated the simulation until the total number of edges in the simulation reached 505 (i.e., the total number of edges in the observed network). Finally, we plotted on one graph the distribution of the degree centrality for all the nodes in the unweighted observed network and in the simulated network, so as to evaluate whether the preferential attachment model provides a good simulation of the observed network.

Moreover, we used exponential random graph models (ERGM)<sup>44</sup> to assess whether the presence or absence of an edge can be explained by the node attribute ‘family affiliation.’ First, we set the baseline model in which only an ‘edge’ term (i.e., number of edges) is specified. We then added a new variable, ‘node-

---

41 Shang, and Yu, “Dongjin guizuzhi shehui de wending jizhi – Shishuoxinyu gongci fenxi”, 50.

42 A. L. Barabási, and R. Albert, “Emergence of scaling in random networks,” *Science* 286, no. 5439 (1999): 509–12.

43 For the algorithm, see Blake Stacey, “Preferential Attachment in Python,” June 19, 2007, <https://www.sunclipse.org/?p=163>.

44 Paul W. Holland, and Samuel Leinhardt, “An Exponential Family of Probability Distributions for Directed Graphs,” *Journal of the American Statistical Association* 76, no. 373 (1981): 33–50.

match (family affiliation),’ to evaluate whether nodes from the same family have a higher probability of being connected.

### *Community detection and group similarity measures*

For group-level analysis, we used the Louvain algorithm<sup>45</sup> to partition the weighted network into communities and created visualizations in Gephi to provide an overview of its structure.<sup>46</sup> We laid out the network with the Fruchterman-Reingold algorithm,<sup>47</sup> and used different colors for different communities (Fig. 2).

To assess the relationship between family affiliation and community formation (i.e., whether people from the same family tended to cluster in the network), we used the Kulczynski distance to measure the similarity between each detected community and aristocratic family. The Kulczynski distance is a common null-invariant (i.e., the number of null records does not influence the results) measure of the similarity between two sets of data. Here it is defined as:

$$\text{Similarity}(A,B) = \frac{1}{2} \left( \frac{|A \cap B|}{|A|} + \frac{|A \cap B|}{|B|} \right)^{48}$$

where  $A$  is a community detected in the network,  $B$  is an aristocratic family,  $|A|$  is the number of people in community  $A$ ,  $|B|$  is the number of people in family  $B$ , and  $|A \cap B|$  is the number of people in community  $A$  who also belong to family  $B$ .

---

45 Vincent D. Blondel, *et al.*, “Fast Unfolding of Communities in Large Networks,” *Journal of Statistical Mechanics: Theory and Experiment* 10 (2008): 155–68.

46 M. Bastian, S. Hermann, and M. Jacomy, “Gephi: An Open Source Software for Exploring and Manipulating Network,” in *Proceedings of the Third International AAAI Conference on Weblogs and Social Media (ICWSM 2009)*, ed. Eytan Adar, *et al.* (San Jose, CA: 2009), 361–62.

47 The Fruchterman-Reingold Algorithm is a force-directed layout algorithm designed for drawing undirected graphs with straight edges. It is suitable only for small graphs due to its high time complexity ( $O(N^2 + E)$ ). We chose this algorithm based on the following considerations: (1) the social network investigated in this research is naturally suitable for this algorithm, as it is undirected and relatively small in size (containing 250 nodes); and (2) the algorithm uses a roughly circular layout, with higher degree nodes at the center, which places important nodes visually together and does not create clear boundaries. This simulates the aristocratic social network in *The Tales of the World* very well, because the interactions between the aristocrats were complex and not confined within small clusters of people separated by clear boundaries. For details on the algorithm, see Thomas M. J. Fruchterman, and Edward M. Reingold, “Graph drawing by force-directed placement,” *Software: Practice and Experience* 21, no. 11 (1991): 1129–64.

48 Guang R. Shi, “Multivariate Data Analysis in Palaeoecology and Palaeobiogeography – A Review,” *Palaeogeography, Palaeoclimatology, Palaeoecology* 105, no. 3–4 (1993): 206.

### *Cumulative weights between families*

At the group level, we also calculated the cumulative weights between families in the weighted network. Here, the cumulative weight between two families is defined as the sum of all edge weights between all node-pairs (pairs of people) where each of the two nodes (people) is from a different one of these two families. The cumulative weight between a family and itself is defined as the sum of all edge weights between all node-pairs where the two nodes are both from this family.

A family with a large number of nodes in the network is more likely to have a higher cumulative weight with other families and with itself. Therefore, we also calculated the average weight between each pair of families by dividing the cumulative weight between these families by the total number of possible node pairs between them. Specifically, the average weight between two different families is the cumulative weight between them divided by the product of the number of people in each of these two families, and the average weight between a family and itself is its cumulative weight divided by  $\binom{n}{2}$ , where  $n$  is the number of people in this family.

### *Centrality measures*

At the node level, both weighted and unweighted networks were used for computing degree centrality and eigenvector centrality. Betweenness centrality and closeness centrality, however, were calculated only for the unweighted network, since these two metrics focus on the position of each node in the network, and for this reason edge weights do not matter.<sup>49</sup> We employed the Python package NetworkX<sup>50</sup> for these calculations.<sup>51</sup> We thus obtained six centrality mea-

---

49 Additionally, calculations of weighted betweenness centrality and weighted closeness centrality measures assume that the “distances” between nodes can be represented by the multiplicative inverse (or reciprocal) of edge weights (i.e., dividing 1 by the edge weight). However, this assumption is not suitable for the network investigated in this research, where the edge weight is the strength of social relationships, and its reciprocal cannot be simply interpreted as the “distance”. For calculating weighted betweenness centrality, see Ulrik Brandes, “A Faster Algorithm for Betweenness Centrality,” *Journal of Mathematical Sociology* 25, no. 2 (2001): 163–77; for calculating weighted closeness centrality, see M. E. J. Newman, “Scientific Collaboration Networks. II. Shortest Paths, Weighted Networks, and Centrality,” *Physics Review E* 64 (2001): 016132.

50 Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart, “Exploring Network Structure, Dynamics, and Function Using NetworkX,” in *Proceedings of the 7th Python in Science Conference (SciPy2008)*, ed. Gäel Varoquaux, Travis Vaught, and Jarrod Millman (Pasadena, CA: 2008), 11–15.

51 NetworkX does not support the calculation of weighted degree centrality. Therefore, we calculated it by summing up the weights of all adjacent edges of each node. This is consistent with the conventional definition of the weighted degree centrality. For de-

asures for each person: weighted and unweighted degree centrality, weighted and unweighted eigenvector centrality, unweighted betweenness centrality, and unweighted closeness centrality.

After calculating these centrality metrics for each person, we examined the family affiliations of those who are ranked highest on each of these metrics and summed up the metrics of all the nodes in each aristocratic family.

## 4. Results

### 4.1 Network Simulation and Prediction

Figure 1 shows that degree distribution follows a similar pattern in the observed and the simulated networks, although the latter has a longer tail than the former. The significant overlap between the two curves of degree distribution strongly demonstrates that the formation of the aristocratic social network in *The Tales of the World* can be considered a preferential attachment process. This suggests that cumulative advantage exists in this network: prestigious individuals attracted more attention from the author of *The Tales of the World*, and this significantly increases their visibility in the text.

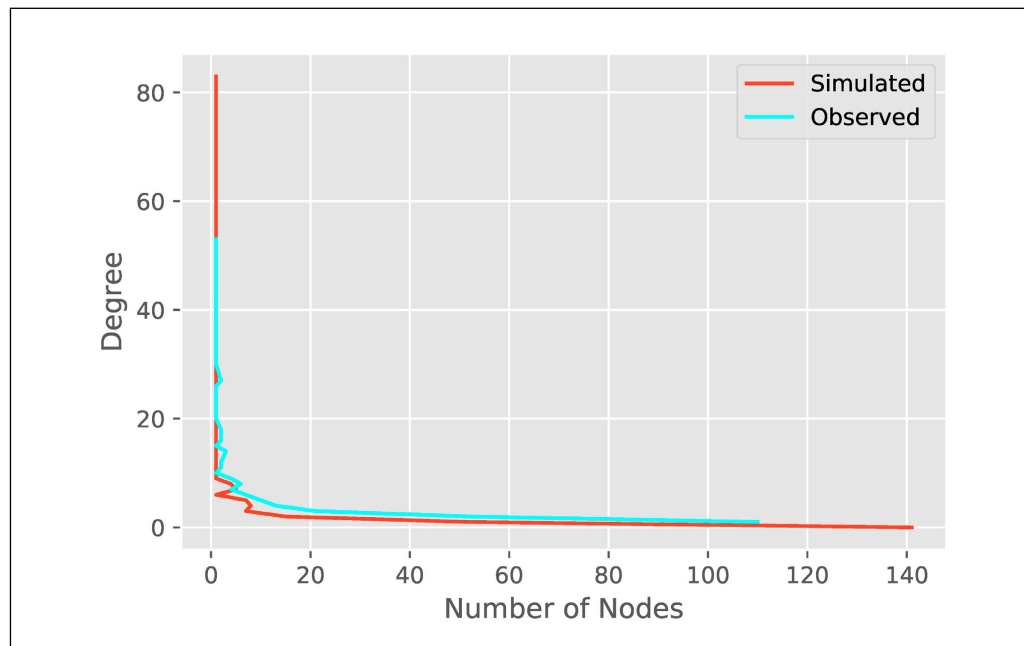
Moreover, the prediction based on ERGM further suggests that a specific node attribute – namely “family affiliation” – has a notable influence on the network structure.

The ERGM analysis (Table 1) reveals that two nodes (people) that share the same family affiliation have a higher probability of having a relationship with each other. In other words, people from the same aristocratic family are more likely to have interactions recorded in *The Tales of the World*. Table 1 provides quantitative evidence for this observation: it shows that the variable ‘nodematch (family affiliation)’ has a positive estimated logistic coefficient and is statistically significant ( $p < 0.001$ ). The values of the estimated logistic coefficients mean that the probability of forming an edge (relationship) is  $\frac{e^{-4.13}}{1 + e^{-4.13}} \approx 0.016$  for two random nodes (people), but slightly more than twice as high ( $\frac{e^{-4.13+0.78}}{1 + e^{-4.13+0.78}} \approx 0.034$ ) for two nodes that share the same family affiliation.

However, these statistics only reveal that people from the same family are more likely than those from different families to have positive social interactions in *The Tales of the World*. They do not tell us how common it was for people from dif-

---

tails of the definition, see T. Opsahl, F. Agneessens, and J. Skvoretz, “Node Centrality in Weighted Networks: Generalizing Degree and Shortest Paths,” *Social Networks* 32, no. 3 (2010): 241–51.



**Fig. 1** Degree distribution in the observed network and the simulated network based on the preferential attachment model.

	Estimated Logistic Coefficient	Std. Error	MCMC %	z value	Pr (>  z )
edges	-4.13	0.05	0	-89.81	< 1e-04***
nodematch (family affiliation)	0.78	0.21	0	3.75	0.0002***

**Tab. 1** Results of ERGM. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

ferent families to have such interactions and what proportion of the relationships in the text were between people from the same family. To answer these questions, we will evaluate the strength of homophily by family affiliation in the network in the following sections.

## 4.2 Visualization and Community Detection

In Figure 2, we visualized the aristocratic social network in *The Tales of the World* to provide a panoramic view of it.

Using the Louvain algorithm with a resolution of 1.0, we partitioned the network into 18 communities. Nine of these communities each have 15 or more members, and the others each have three or fewer members. Each of the nine major communities centers on a man from a powerful family: Wang Dao (center of the community in purple) and Wang Dun (grey) from the Wang family of Langya, Yu Liang 庾亮 (pink) from the Yu family of Yingchuan, Huan Wen 桓温

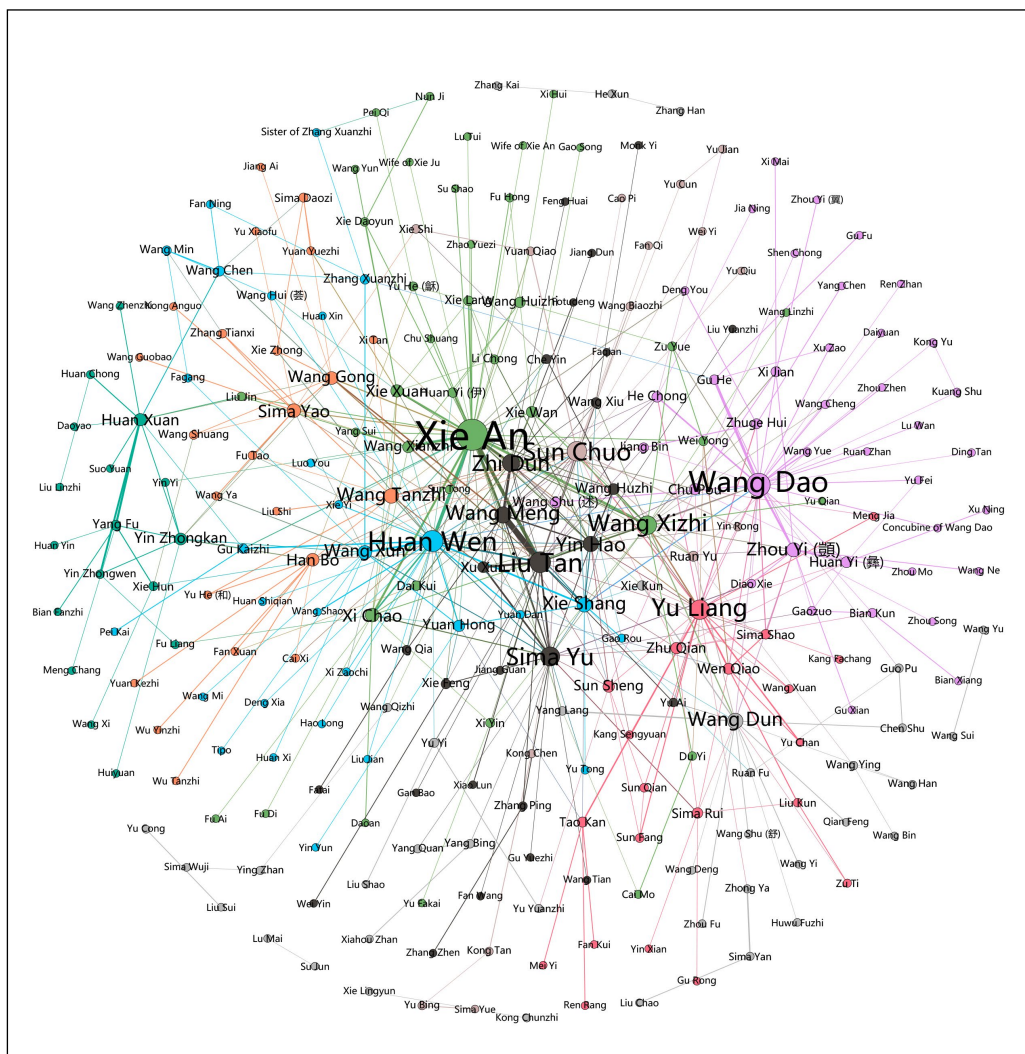


Fig. 2 Aristocratic Social network of the Eastern Jin in *The Tales of the World*.

(blue) and Huan Xuan (cyan) from the Huan family of Qiaoguo, Xie An 謝安 (green) from the Xie family of Chenjun, Wang Tanzhi 王坦之 (orange) from the Wang family of Taiyuan, Sun Chuo 孫綽 (brown) from the Sun family of Taiyuan 太原孫氏, and Liu Tan (black) from the Liu family of Peiguo 沛國劉氏. Therefore, the central figures in seven of the nine major communities (i.e., all except for the communities in green and brown) are from the five most powerful families that once dominated the positions of power in the Eastern Jin, and all of these five families each have a member at the center of at least one of the major communities.

Hence, the social interactions of the five most powerful families attracted exceptional attention in *The Tales of the World*, and each of these families had a remarkable influence within their respective social circles. However, the boundaries of these social circles are porous, and the detected social communities in the network should not be considered as a mirror image of different aristocratic families.

Table 2 reports the Kulczynski similarity measures between the nine major communities detected by the Louvain algorithm and the five most powerful aristocratic families. These statistics show that the Wang family of Langya and the community in grey have the highest similarity score of 0.363, but this is nonetheless only a moderate value. The similarity scores for other families and detected communities are even lower: none of the five most powerful families has a similarity score above 0.3 with any of the major communities. This indicates that heterogeneity in family affiliation is strong and homophily by family affiliation is weak in all the communities.<sup>52</sup>

These results provide a preliminary answer to the question raised in section 4.1 concerning the strength of homophily by family affiliation: although the aristocrats in the Eastern Jin were more likely to have positive interactions with people from their own family, it was still common for them to associate with people from other families. Instead of consolidating around their own kin and keeping a distance from other families, aristocrats in the Eastern Jin dynasty formed a congen-

---

52 The Kulczynski similarity of two sets is defined as the average of a) the size of the intersection of the two sets divided by the size of set A, and b) the size of the intersection of the two sets divided by the size of set B (its formula can be found in section 3.2). Therefore, if one-third of the members of community A are from family B, and one-third of the members of family B are in community A, the Kulczynski similarity between community A and family B would be  $(1/3 + 1/3)/2 = 1/3$ . However, in this case, the heterogeneity is still strong, because two-thirds of the members of community A are from other families and two-thirds of the members of family B are in other communities. Therefore, 0.363, which is only slightly greater than 1/3, should not be considered a high value that indicates strong similarity, and values less than 0.3 are very low.



Aristocratic Family	Detected community (denoted by color)	Central Figure in the Community	Similarity
Wang family of Langya	Grey	Wang Dun	0.363
Huan family of Qiaoguo	Blue	Huan Wen	0.291
Xie family of Chenjun	Green	Xie An	0.272
Huan family of Qiaoguo	Cyan	Huan Xuan	0.255
Wang family of Taiyuan	Orange	Wang Tanzhi	0.254
Wang family of Langya	Blue	Huan Wen	0.179
Wang family of Taiyuan	Purple	Wang Dao	0.163
Wang family of Langya	Black	Liu Tan	0.146
Yu family of Yingchuan	Pink	Yu Liang	0.139
...	...	...	...

**Tab. 2** Kulczynski similarity between the five most powerful families and the nine major communities (in descending order, based on similarity).

ial social circle in *The Tales of the World*, a sign of their shared consciousness of being part of a privileged status group.

### 4.3 Cumulative Weights Between Families

The preceding discussion has shown that homophily by family affiliation is weak within the network. In this section, we take a closer look at homophily in the network and examine whether the degree of homophily varies for different aristocratic families by calculating the cumulative weights between these families in the network.

We focus particularly on the five most powerful families that once dominated the positions of power in the Eastern Jin. Tables 3 and 4 show that the cumulative and average weights between two different aristocratic families are not generally smaller than those between an aristocratic family and itself. For example, of the highest four cumulative weights (the four over 30) in Table 3, three are between different families and only one is between a family and itself. Table 4 also shows that three out of the five families (the Wang family of Langya, the Huan family of Qiaoguo, and the Wang family of Taiyuan) have the highest average weights with other families, rather than themselves. This suggests that it was common for these families to have interactions with one another in *The Tale of the World*. In other words, homophily by family affiliation is relatively weak in this network, which implies a high degree of social cohesion uniting these aristocratic families as a status group.

	Wang family of Langya	Yu family of Yingchuan	Huan family of Qiaoguo	Xie family of Chenjun	Wang family of Taiyuan
Wang family of Langya	37	–	–	–	–
Yu family of Yingchuan	17	10	–	–	–
Huan family of Qiaoguo	25	2	7	–	–
Xie family of Chenjun	51	9	32	28	–
Wang family of Taiyuan	21	3	10	33	15

**Tab. 3** Cumulative weights between the five most powerful families.

	Wang family of Langya	Yu family of Yingchuan	Huan family of Qiaoguo	Xie family of Chenjun	Wang family of Taiyuan
Wang family of Langya	0.105	–	–	–	–
Yu family of Yingchuan	0.057	0.182	–	–	–
Huan family of Qiaoguo	0.103	0.020	0.194	–	–
Xie family of Chenjun	0.157	0.068	0.296	0.424	–
Wang family of Taiyuan	0.065	0.023	0.093	0.229	0.227

**Tab. 4** Average weights between the five most powerful families.

The Xie family of Chenjun is particularly notable for their intensive interactions with other families. This family traced their descent only to the mid-third century and therefore could not look back on a long aristocratic tradition as an “old family.”<sup>53</sup> More or less a parvenu, the Xie family of Chenjun may have been particularly eager to build social connections with the other powerful families as a means of securing its status in the aristocratic circles. In Table 3, three out of the four highest average weights in the network are between the Xie and other

53 Grafflin, “The Great Family in Medieval South China”, 69.

powerful families, while in Table 4, all three families that have the highest average weights with other families rather than themselves have the highest average weights with the Xie family of Chenjun.

Next, we expand our scope of investigation beyond the five most powerful families and look at homophily between different categories of aristocratic families in the Eastern Jin. We place aristocratic families into two broad categories: the native southern families, and the northern émigré families. Within each category, we make a further distinction and categorize the native southern families into top-tier families and other families, and the northern émigré families into top-tier families, second-tier families, and other families.<sup>54</sup> Table 5 compares the intensity of cumulative weights (ICW)<sup>55</sup> within each group of families (indicative of homophily within the group) and also between different groups (indicative of intimacy between groups).

Historians argue that the top-tier native southern families had a deep distrust of the Eastern Jin regime and the powerful northern émigré families that dominated the court. The statistics in Table 5 are broadly consistent with this view. It shows that the top-tier native southern families tended to focus more on building relationships among themselves (hence the exceptionally high ICW of 2.585

54 The top-tier native southern families are the Gu 顧, Lu 陸, Zhu 朱, Zhang 張 families of Wujun 吳郡 and the Yu 虞, Xie 謝, Kong 孔, Wei 魏 families of Kuaiji 會稽. This definition is based on the contemporary recognition of these families' prominent status in the Eastern Jin. For example, *The Tales of the World* lists 'Gu, Lu, Zhu, Zhang' as the 'four principal surnames' of Wujun, while *Jiangxi kaogu lu* lists 'Gu, Lu, Zhu, Zhang' as the top-tier families of Wujun and 'Yu, Xie, Kong, Wei' as the top-tier families of Kuaiji. See Liu, *Shih-shuo Hsin-yü: A New Account of Tales of the World*, 261, and Wang Mo 王謨, *Jiangxi kaogu lu* 江西考古錄 [Investigations on Ancient Matters of Jiangxi] (n.p.: Fumei shuwu, 1891), 371. In this research, only seven of the eight families are included, since no member of the Zhu family of Wujun appears in the network. All other native southern families are second-tier native southern families. The five second-tier northern émigré families are those that each have 5 to 7 members included in the network: the Xi family of Gaoping, the Sun family of Taiyuan, the Yin family of Chenjun 陳郡殷氏, the Yuan family of Chenjun 陳郡袁氏, and the Yang family of Taishan 泰山羊氏. By contrast, the five most powerful families and the imperial clan (the Sima family of Henei 河內司馬氏) are the top-tier northern émigré families, which each have 8 to 27 members in the network. All other northern émigré families are the third-tier northern émigré families.

55 In this research, A's intensity of cumulative weights with B is defined as:  $I_{AB} = \frac{CW_{AB}/CW_A}{\frac{n_A n_B / (n_A(n - n_A) + \binom{n_A}{2})}{\binom{n}{2}}}$ , where  $CW_{AB}$  is the cumulative weights between all nodes in A and all nodes in B,  $CW_A$  is the cumulative weight between all nodes in A and all the nodes in the entire network,  $n_A$ ,  $n_B$ ,  $n$  are the number of nodes in A, B and the whole network respectively. For example, if there are two nodes in A, five nodes in B, and one hundred nodes in the entire network, and if the cumulative weight between the two nodes in A and the five nodes in B is 10 and that between the two nodes in A and all the nodes of network is 50, then A's 'intensity of cumulative weights' with B is  $(10 \div 50) \div ((2 \times 5) \div (2 \times (100 - 2) + 1)) = 3.94$ .

	All northern émigré families	Top-tier northern émigré families	Second-tier northern émigré families	Third-tier northern émigré families	All native southern families	Top-tier native southern families	Second- tier native southern families
All northern émigré families	1.284	1.48	1.146	0.906	0.306	0.315	0.278
Top-tier northern émigré families	1.29	1.633	1.176	0.882	0.36	0.351	0.389
Second-tier northern émigré families	1.179	1.506	1.567	0.613	0.2	0.264	0
Third-tier northern émigré families	1.001	1.356	0.736	0.823	0.196	0.21	0.152
All native southern families	0.917	1.344	0.583	0.475	1.837	1.832	0.709
Top-tier native southern families	0.906	1.257	0.739	0.489	1.758	2.585	0.517
Second-tier native southern families	1.023	1.782	0	0.452	0.872	0.663	2.366

**Tab. 5** Intensity of cumulative weights for different groups of families.

within this group, which more than doubles that of a randomly selected group of families) than with the top-tier families of northern origins (hence the low ICW of 1.257, which is slightly higher than that for a randomly selected group of families). This seems to suggest that the top-tier native southern families cautiously avoided intimacy with the top-tier northern émigré families in order to distance themselves from political conflicts.<sup>56</sup>

The second-tier native southern families showed a similar tendency: they interacted more intensively among themselves than with the northern émigrés. However, there is also a crucial difference: the second-tier southern families could not solely rely on each other to shore up their social position, but also had to build connections with the powerful northern émigré families. Compared to the top-tier southern families, the second-tier southern families had fewer connections among themselves (ICW = 2.366) and a much closer relationship with the top-tier northern émigré families (ICW = 1.782).

Unlike native southern families, which have closer relationships among themselves than with the top-tier northern émigré families, the second-tier northern émigré families were almost as closely connected with the top-tier northern émigré families (ICW = 1.506) as they were with each other (ICW = 1.567), and the third-tier northern families had closer relationships with the five most powerful ones (ICW = 1.356) than among themselves (ICW = 0.823). These results suggest that these less prominent northern émigré families acted as ‘adhesives’ in the aristocratic society of the Eastern Jin, mediating the relationship between the top-tier northern émigré families and the others. Tian Yuqing’s study of Xi Jian 郗鑒 from the Xi family of Gaoping corroborates this finding. He argued that “the efforts of Xi Jian maintained the balance of power between different powerful families, perpetuated ‘the politics of powerful families,’ and allowed the Eastern Jin regime to exist for a long time.”<sup>57</sup>

---

56 An alternative explanation for why the native southern families had a low intensity of cumulative weights with the top-tier northern émigré families is that the northern émigré families looked down upon the southerners and edged them out. However, it should be noted that the intensity of cumulative weights between the top-tier native southern families and the top-tier northern émigré families is even lower than that between second-tier southern families and the top-tier northern émigré families. Since it is unlikely that the top-tier native southern families were more marginalized by the northern émigrés than were the second-tier native southern families, their low intensity of interaction with the top-tier northern émigré families seems more likely a result of their deliberate choice.

57 Tian, *Dongjin menfa zhengzhi*, 96.

#### 4.4 Node Centrality Measures

Statistics on node-level centrality measures (Table 6) show that members of the five most powerful families occupied the most important structural positions in the aristocratic social network of *The Tales of the World*. For example, the three people with the highest ranking in degree centrality were all from the five most powerful families: Xie An was from the Xie family of Chenjun, Wang Dao from the Wang family of Langya, and Huan Wen from the Huan family of Qiaoguo.

Table 6 shows that a person who ranks high in one centrality measure tends to have a high ranking in other centrality measures as well. However, there are a few outliers, which deserve attention. For example, Wang Dao and Yu Liang rank considerably lower in eigenvector centralities (weighted and unweighted) but high in all other centrality measures, while Liu Tan ranks high in eigenvector centrality but low on betweenness centrality.

To fully understand these variances requires that we look into the different roles played by different aristocratic families in the network. We calculated the centrality scores for each family by adding up the relevant centrality scores for all members of that family. In Table 7, we focus again on the five most powerful families.

Table 7 shows that the Wang family of Langya and the Xie family of Chenjun dominated the network. They were the most securely established, ranking highest in all centrality measures. The high connectivity of these two families might explain their dominance in the aristocratic network and why they were the only

Person	Rank in degree centrality (un-weighted)	Rank in degree centrality (weighted)	Rank in eigenvector centrality (un-weighted)	Rank in eigenvector centrality (weighted)	Rank in betweenness centrality	Rank in closeness centrality
Xie An	1	1	1	2	1	1
Wang Dao	2	3	11	19	2	8
Huan Wen	3	4	7	4	3	2
Liu Tan	4	2	2	1	7	3
Yu Liang	5	7	10	16	5	11
...	...	...	...	...	...	...

**Tab. 6** Rank of people in different centrality measures (showing only those that rank in the top five in unweighted degree centrality).

Family	Rank in degree centrality (un-weighted)	Rank in degree centrality (weighted)	Rank in eigenvector centrality (unweighted)	Rank in eigenvector centrality (weighted)	Rank in betweenness centrality	Rank in closeness centrality
Wang family of Langya	1	1	1	3	1	1
Xie family of Chenjun	2	2	2	1	2	2
Wang family of Taiyuan	3	3	3	2	6	3
Huan family of Qiaoguo	4	4	4	5	3	5
Yu family of Yingchuan	6	7	7	9	5	4

**Tab. 7** Rank of the five most powerful families in different centrality measures.

two that continued to produce a good number of eminent figures in the fifth and sixth centuries following the fall of the Eastern Jin.<sup>58</sup>

Not all families rank similarly in all centrality measures. The Huan family of Qiaoguo ranks high in betweenness centrality (which measures how often a node is traversed along the shortest paths between two other nodes) but relatively low in closeness centrality (which measures the multiplicative inverse of the average length of the shortest path between a node and all other nodes in the network). The Huan were at the peak of their power in the mid-fourth century when Huan Wen dominated the court. The Huan's high rank in betweenness centrality reflects the historical importance of Huan Wen and Huan Xuan, who act as 'gatekeepers' or 'bridges' in this network, and through whom many shortest links between people pass.

The Wang family of Taiyuan is just the opposite: it ranks low in betweenness centrality but relatively high in closeness centrality. The Wang reached the peak of their power in the late fourth century. The Wang's high rank in closeness centrality reflects their role as 'monitors' in the network, who are capable of reaching other people quickly. This appears to partly explain why the Wang continued to be influential after they fell from the peak of their power, unlike the Huan, whose influence declined dramatically after their golden days were over.

58 Grafflin, "The Great Family in Medieval South China", 70.

Family	Rank in degree centrality (un-weighted)	Rank in degree centrality (weighted)	Rank in eigenvector centrality (un-weighted)	Rank in eigenvector centrality (weighted)	Rank in betweenness centrality	Rank in closeness centrality
Gu family of Wujun	13	12	16	13	14	9
Zhang family of Wujun	15	18	25	15	16	15
Yu family of Kuaiji	20	28	23	32	35	12
Kong family of Kuaiji	23	25	28	35	13	16
Xie family of Kuaiji	28	29	22	25	23	32
Wei family of Kuaiji	40	44	39	42	44	25
Lu family of Wujun	40	51	47	45	44	44

**Tab. 8** Rank of the top-tier native southern families in different centrality measures (in ascending order, based on the rank of degree centrality).



The Yu family of Yingchuan's influence also declined rapidly after losing their power, which may be partly explained by the fact that many of those who had relations with the Yu family were themselves poorly connected. This is reflected in the Yu family's low ranking in weighted eigenvector centrality, which measures the influence of a node on its connections by assigning higher weights to those connections that are themselves well connected, as well as taking edge weights (the strength of relationships) into account.

The top-tier native southern families, in general, tend to rank higher in closeness centrality than in weighted eigenvector centrality. This suggests that although they had contacts with a variety of people in the network and were generally capable of reaching different parts of the network efficiently, their connections were themselves usually not well-connected. These findings corroborate the argument we made earlier in this article on how these southern families successfully maintained their status in the Eastern Jin: despite their close relationships among themselves, different native southern families often sided with different northern émigrés in the political struggle, which in effect reduced the overall risk facing the southern families as a whole. Tian Yuqing also pointed out that “the native southern families were always divided into two camps, each allying itself with a different side in the conflict within the northern émigré families.”<sup>59</sup>

## 5. Conclusions, Limitations, and Future Work

The foregoing analyses lead to several conclusions. First, at the network level, the formation of the aristocratic social network of *The Tales of the World* follows the logic of preferential attachment. Furthermore, an identical family affiliation of two nodes increases the probability of a positive social connection between them. Community detection reveals that the five most powerful families had a remarkable influence in the network, with each at the center of a sizable social circle (i.e., a detected community), although each circle had participants from many different families.

Second, at the group level, each group of families had internal connections of varying intensity. The native southern families – the top-tier ones in particular – were more strongly connected to each other than the northern émigré families, while all families had intensive connections with the five most powerful families.

Third, at the node level, although people with a high score in one centrality measure generally have a high score in all the other centrality measures, a close look at the outliers leads to historically meaningful discoveries. Different northern émigré families rank relatively higher in different centrality measures,

---

59 Tian, *Dongjin menfa zhengzhi*, 72.

suggesting that they occupied different structural positions and played different roles in the network. Native southern families generally rank high in closeness centrality and low in weighted eigenvector centrality, reflecting the unique challenges they faced in Eastern Jin politics and how they tackled them.

These findings lead us to a more general conclusion. *The Tales of the World* shows that homophily by family affiliation existed in the aristocratic society of the Eastern Jin, but that such homophily was weak. While many previous studies have emphasized the conflicts between the aristocratic families in the Eastern Jin, some going so far as to argue that some of these families were “as incompatible as water and fire,”<sup>60</sup> our findings reveal a different dimension to their relationship: despite the political struggles that divided them, aristocratic families in the Eastern Jin developed intensive and favorable relationships with each other, which helped them to maintain a high degree of social cohesion among themselves and ensured their century-long dominance in the Eastern Jin society as a privileged status group.

This argument corroborates some earlier studies. Building on his case study of Huan Xuan’s rebellion against the Eastern Jin court, Zhu Zongbin has suggested that aristocratic families may have formed alliances for their common interests.<sup>61</sup> However, our study goes beyond microscopic case studies to provide a comprehensive analysis of the social relationships among the aristocratic families. In doing so, we have expanded the social historical research started by Kawakatsu Yoshio with the help of social network analysis.

Admittedly, the data source for this study has its limitations. Like all historical texts, *The Tales of the World* is not free from the author’s bias. It is arguably more selective and less thorough than official dynastic histories in its representation of history.<sup>62</sup> Therefore, the aristocratic social network we have constructed from *The Tales of the World* is far from an exhaustive or unbiased representation of the social interactions between aristocrats. For example, members of the same family were obviously connected by consanguinity, which creates arguably strong intra-family connections but is not explicitly mentioned in *The Tales of the World*. Therefore, weak homophily by family affiliation means only that interactions between members of the same family in a social setting were limited or

---

60 Tian, *Dongjin menfa zhengzhi*, 123.

61 Zhu, “Shilun Dongjin houqi gaoji shizu zhi moluo ji Huan Xuan dai Jin zhi xingzhi”.

62 Kawakatsu Yoshio provided insightful discussions on the historical context for the composition of *The Tales of the World* and the author’s inclinations. See Kawakatsu, *Liuchao guizuzhi shehui yanjiu*, 238–252. The bias has also been noted in the introduction to the English translation of *The Tales of the World*, where Richard B. Mather observed that the author of the book had a preference for the “adherents of naturalness” over the “adherents of conformity.” See Richard B. Mather, “Introduction: The World of the Shih-Shuo Hsin-Yü,” in *Shih-Shuo Hsin-Yü: A New Account of Tales of the World*, xviii.

went unnoticed in *The Tales of the World*. This caveat notwithstanding, the network we have constructed provides sufficient evidence to show that although different aristocratic families frequently came into conflict, there was a high degree of social cohesion that united them as a status group, which vigorously defended their common interests when they were threatened by the imperial power or foreign enemies (for example, in the Battle of Fei River against invaders from the north in 383 C.E.). This was the element of solidity in the turbulent flow of aristocratic society, which explains why the entire century of the Eastern Jin was characterized by “the politics of powerful families.”<sup>63</sup>

Many other aspects of the aristocratic social network of *The Tales of the World* invite future research. Particularly promising topics include, for example, the structural equivalence of people from different aristocratic families in the network, the balance of triads, and its relationship with the family affiliation of individual nodes. It may also be worthwhile to collect network data from the official dynastic history, the *Book of Jin*, and combine it with what we have already reconstructed from *The Tales of the World*.

This research is an example of how network analysis methods can be adopted for the study of pre-modern history. Network analysis allows us to abstract relationships from complex historical sources and create manageable graph models based on a set of well-defined rules. This methodological conciseness, inherent in the network analysis approach, not only forces scholars to clarify their research questions, but also allows them to synthesize a large corpus of source materials and draw more convincing conclusions than is possible with case studies. Combining theories from mathematics, computer sciences, and social sciences, network analysis significantly improves the methodological power of humanities research.

## References

- Barabási, A. L., and R. Albert. “Emergence of Scaling in Random Networks.” *Science* 286, no. 5439 (1999): 509–12.
- Bastian, M., S. Hermann, and M. Jacomy. “Gephi: An Open Source Software for Exploring and Manipulating Network.” In *Proceedings of the Third International Conference on Weblogs and Social Media, ICWSM 2009*, edited by Eytan Adar, Matthew Hurst, Tim Finin, Natalie Glance, Nicolas Nicolov, and Belle Tseng. San Jose, CA, 2009.
- Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. “Fast Unfolding of Communities in Large Networks.” *Journal of Statistical Mechanics: Theory and Experiment*, no. 10 (2008): 155–68.

---

63 Tian, *Dongjin menfa zhengzhi*, 1–2.

- Brandes, Ulrik. "A Faster Algorithm for Betweenness Centrality." *Journal of Mathematical Sociology* 25, no. 2 (2001): 163–77.
- Breure, Abraham S. H., and Raphael H. Heiberger. "Reconstructing Science Networks from the Past. Eponyms Between Malacological Authors in the Mid-19th Century." *Journal of Historical Network Research* 3 (2019): 92–117.
- Brughmans, Tom, and Matt Peeples. "Trends in Archaeological Network Research: A Bibliometric Analysis." *Journal of Historical Network Research* 1 (2017): 1–24.
- Chen, Yinke 陳寅恪. "Shu Dong Jin Wang Dao zhi gongye" 述東晉王導之功業 [On the Accomplishments of Wang Dao in the Eastern Jin]. In *Jinmingguan congkao chubian* 金明館叢稿初編 [Writings from Jinmingguan, Vol. 1], 55–77. Shanghai: Shanghai guji chubanshe, 1980.
- "Weishu Sima Rui zhuan Jiangdong minzu tiao shizheng ji tuilun" 魏書司馬睿傳江東民族條釋證及推論 [An Explanation and Inference Concerning the Section on Ethnicities in the Jiangdong Area in Biography of Sima Rui in the *Book of Wei*]. In *Jinmingguan congkao chubian* 金明館叢稿初編 [Writings from Jinmingguan, Vol. 1], 78–119. Shanghai: Shanghai guji chubanshe, 1980.
- Deicke, Aline. "Networks of Conflict: Analyzing the 'Culture of Controversy' of Polemical Pamphlets of Intra-Protestant Disputes (AD 1548–1580)." *Journal of Historical Network Research* 1 (2017): 71–105.
- Ebrey, Patricia. *The Aristocratic Families of Imperial China: A Case Study of the Po-Ling Ts'ui Family*. Cambridge, UK: Cambridge University Press, 1978.
- Fogel, Joshua A. *Politics and Sinology: The Case of Naitō Konan (1866–1934)*. Cambridge, MA: Harvard University Press, 1984.
- Fruchterman, Thomas M. J., and Edward M. Reingold. "Graph Drawing by Force-Directed Placement." *Software: Practice and Experience* 21, no. 11 (1991): 1129–64.
- Grafflin, Dennis. "The Great Family in Medieval South China." *Harvard Journal of Asiatic Studies* 41, no. 1 (1981): 65–74.
- Hagberg, Aric A., Daniel A. Schult, and Pieter J. Swart. "Exploring Network Structure, Dynamics, and Function Using NetworkX." In *Proceedings of the 7th Python in Science Conference (SciPy2008)*, edited by Gäel Varoquaux, Travis Vaught, and Jarrod Millman, 11–15. Pasadena, CA, 2008.
- Holland, Paul W., and Samuel Leinhardt. "An Exponential Family of Probability Distributions for Directed Graphs." *Journal of the American Statistical Association* 76, no. 373 (1981): 33–50.
- Johnson, David G. *The Medieval Chinese Oligarchy*. Boulder, CO: Westview Press, 1977.
- Kawakatsu, Yoshio 川勝義雄. *Liuchao guizuzhi shehui yanjiu* 六朝貴族制社會研究 [A Study on Aristocratic Society in the Six Dynasties]. Translated by Xu Gupeng 徐谷芑 and Li Jicang 李濟滄. Shanghai: Shanghai guji chubanshe, 2007.

- Köstner, Elena. "Trimalchio's Last Will: Shifting Interactions Between Seeming and Being." *Journal of Historical Network Research* 3 (2019): 1–29.
- Lewis, Mark Edward. *China between Empires: The Northern and Southern Dynasties*. Vol. 2. History of Imperial China. Cambridge, MA and London, UK: The Belknap Press of Harvard University Press, 2009.
- Liu, I-ch'ing 劉義慶. *Shih-shuo Hsin-yü: A New Account of Tales of the World*. Translated by Richard B. Mather. Second Edition. Ann Arbor, MI: Center for Chinese Studies, The University of Michigan, 2002.
- Mao, Hanguang 毛漢光. "Liangjin Nanbeichao zhuyao wenguan shizu cheng-fen de tongjifenxi yu bijiao" 兩晉南北朝主要文官士族成分的統計分析與比較 [A Statistical Analysis and Comparison of the Family Affiliations of Major Officials in the Two Halves of Jin and the Southern and Northern Dynasties]. In *Zhongguo zhonggu shehuishi lun* 中國中古社會史論 [Essays on the Social History of Medieval China], 141–86. Shanghai: Shanghai shudian chubanshe, 2002.
- "Zhongguo da shizu zhi ge'an yanjiu – Langya Wangshi" 中古大士族之個案研究 – 琅琊王氏 [A Case Study of the Great Aristocratic Families in Medieval China: The Wang Family of Langya]. In *Zhongguo zhonggu shehuishi lun* 中國中古社會史論 [Essays on the Social History of Medieval China], 365–404. Shanghai: Shanghai shudian chubanshe, 2002.
- Mather, Richard B. "Introduction: The World of the Shih-Shuo Hsin-Yü." In *Shih-Shuo Hsin-Yü: A New Account of Tales of the World*, Second Edition., xiii–xxxv. Ann Arbor, MI: Center for Chinese Studies, The University of Michigan, 2002.
- Miyazaki, Ichisada 宮崎市定. *Jiupinguanrenfa yanjiu: Keju qianshi* 九品官人法研究：科舉前史 [Research on the Method of Selecting Officials According to the Nine Ranks: The Prehistory of the Examination System]. Translated by Han Sheng 韓昇 and Liu Jianying 劉建英. Beijing: Zhonghua shuju, 2008.
- Naitō, Konan 內藤湖南. *Zhongguoshi tonglun – Neiteng Hunan Boshi Zhongguo shixue zhuzuo xuanyi* 中國史通論 – 內藤湖南博士中國史學著作選譯 [General Comments of Chinese History – Selected Translations of Dr. Naitō Konan's Works on Chinese History]. Translated by Xia Yingyuan 夏應元, Liu Wenzhu 劉文柱, Xu Shihong 徐世虹, Zheng Xianwen 鄭顯文, and Xu Jianxin 徐建新. Beijing: Shehui kexue wenxian chubanshe, 2004.
- "Gaikuoxing de Tang Song shidai guan" 概括性的唐宋時代觀 [General Views on the Tang and-Song Period]. In *Dongyang wenhuashi yanjiu* 東洋文化史研究 [A Study of Oriental Cultural History], translated by Lin Xiaoguang 林曉光, 103–112. Shanghai: Fudandaxue chubanshe, 2016.
- Newman, M. E. J. "Scientific Collaboration Networks. II. Shortest Paths, Weighted Networks, and Centrality." *Physics Review E* 64 (2001): 016132.
- Opsahl, T., F. Agneessens, and J. Skvoretz. "Node Centrality in Weighted Networks: Generalizing Degree and Shortest Paths." *Social Networks* 32, no. 3 (2010): 241–51.

- Schauf, Andrew, and Miguel Escobar Varela. "Searching for Hidden Bridges in Co-Occurrence Networks from Javanese Wayang Kulit." *Journal of Historical Network Research* 2 (2018): 26–52.
- Shafie, Termeh, David Schoch, Jimmy Mans, Corinne Hofman, and Ulrik Brandes. "Hypergraph Representations: A Study of Carib Attacks on Colonial Forces (AD 1509–1700)." *Journal of Historical Network Research* 1 (2017): 52–70.
- Shang, Wenyi 尚聞一, and Yu Zixuan 于子軒. "Dongjin guizuzhi shehui de wending jizhi – Shishuoxinyu gongci fenxi" 東晉貴族制社會的穩定機制 – 《世說新語》共詞分析 [The Stability Mechanism of the Aristocratic Society in the Eastern Jin Dynasty: A Co-Word Analysis of A New Account of the Tales of the World]. *Tushuguan luntan* 圖書館論壇, no. 1 (2019): 46–57.
- Shi, Guang R. "Multivariate Data Analysis in Palaeoecology and Palaeobiogeography – A Review." *Palaeogeography, Palaeoclimatology, Palaeoecology* 105, no. 3–4 (1993): 199–234.
- Stacey, Blake. "Preferential Attachment in Python," June 19, 2007. <https://www.sunclipse.org/?p=163>.
- Tackett, Nicolas. *The Destruction of the Medieval Chinese Aristocracy*. Cambridge, MA: Harvard University Asia Center, 2014.
- Tang, Zhangru 唐長孺. *Wei Jin Nanbeichao shilun shiyi* 魏晉南北朝史論拾遺 [Miscellaneous Treatises on the Historical Theories of Wei, Jin, and the Southern and Northern Dynasties]. Beijing: Zhonghua shuju, 1983.
- Tanigawa, Michio 谷川道雄. *Medieval Chinese Society and the Local "Community"*. Translated by Joshua A. Fogel. Berkeley, CA: University of California Press, 1985.
- Tian, Yuqing 田余慶. *Dongjin menfa zhengzhi* 東晉門閥政治 [The Politics of the Powerful Families during the Eastern Jin]. Beijing: Beijingdaxue chubanshe, 2012.
- Wang, Mo 王謨. *Jiangxi kao gu lu* 江西考古錄 [Investigations on Ancient Matters of Jiangxi]. n.p.: Fu mei shu wu, 1891.
- Yu, Yingshi 余英時. *Zhongguo zhishi jieceng shilun (gudai pian)* 中國知識階層史論（古代篇） [A Historical Treatise on Chinese Intelligentsia (Ancient Times)]. Taipei: Lianjing chuban shiye gongsi, 1980.
- Zhu, Zongbin 祝總斌. "Shilun Dongjin houqi gaoji shizu zhi moluo ji Huan Xuan dai Jin zhi xingzhi" 試論東晉後期高級士族之沒落及桓玄代晉之性質 [On the Decline of the High Aristocratic Families and the Nature of Huan Xuan's Replacement of Jin in the Late Eastern Jin]. *Beijingdaxue xuebao (zhexue shehuikexue ban)* 北京大學學報（哲學社會科學版）, no. 3 (1985): 77–90.



XIONG HUEI-LAN

# Path toward the Top Leadership: A Network Analysis of the Civil Service System in the Early Southern Song (1131–1164)

*Journal of Historical Network Research* 5 (2021) 33–86

**Keywords** Career trajectory, directed acyclic graphs, talent-nurturing, the Imperial Library, Song Dynasty

**Abstract** This study takes a network approach to explore the career trajectories of Song officials who served in the Imperial Library from 1131 to 1164. It exemplifies how officials' career records, abundant in Chinese historical sources, can provide a trove of information that we may utilize to decipher the operations of imperial bureaucracy and the career paths of officials. "Nurturing talents in the academic institutes" was a policy that the Song dynasty developed to prepare promising junior officials for top civilian leadership. Under this "talent-nurturing" policy, "talented officials" were appointed to the Imperial Library. Such appointments were considered a launchpad for a successful career, as the appointees often soon advanced to other influential offices across diverse branches of government. As an official was transferred from one position to another, the connections between offices reveal early Southern Song bureaucratic practices. Using the career records of officials appointed to the Imperial Library in the early Southern Song, this paper constructs a directed network of bureaucratic transfers. It demonstrates that many offices held by beneficiaries of the "talent-nurturing" policy were stepping-stones towards the highest echelon of government. Their duties involved personnel administration, especially the evaluation of the character and performance of other officials. By focusing on the early Southern Song, a period in which the dynasty was embroiled in war and struggling to survive, this paper also demonstrates how political crises impacted the implementation of the "talent-nurturing" policy.

## Bureaucratic Careers as a Network\*

A bureaucratic system is fundamental to a functioning state. In Chinese history, the Song dynasty (960–1279) is especially famous for the complexity of its bureaucracy.<sup>1</sup> Though historical records on the Song bureaucracy may appear impenetrable to some modern readers, several eminent scholars have argued that they are useful for deciphering government operations and the career trajectories of its officials.<sup>2</sup> These records offer invaluable information on how government agencies were functionally connected and how officials moved from one office to another in a sequential, if puzzling, manner. Based on the connectedness between government agencies and the directedness of bureaucratic transfers, I propose that network analysis provides an innovative approach for exploiting the rich data on the officials' careers preserved in Chinese historical sources and will deepen our understanding of the political culture of Song times.

This article demonstrates the promise of this approach with a case study that investigates the revival and implementation of a Song-dynasty personnel policy known as “nurturing talents” (*yucai* 育才/材) in academic institutes, with a particular focus on the period from 1131 to 1164. The policy dates back to the founding years of the dynasty and was maintained until its end.<sup>3</sup> Twentieth-century re-

---

\* **Acknowledgements:** I thank our editors, Henrike Rudolph and Chen Song, and two anonymous reviewers for this special issue in *JHNR*. They provided insightful and constructive comments which have sharpened my argument. In particular, I owe our editors a debt of gratitude for their technical support of network analysis and language help. I am grateful to Professor Charles Hartman and my supervisor Professor Hilde De Weerd for giving me instructive advice on how to revise this article. I am also grateful to Professor Deng Xiaonan and my friend Fang Chengfeng for deepening my understanding of Song institutional history during my visit to Beijing in 2019. All infelicities remain my own.

**Corresponding author:** Xiong Huei-Lan, hueilanxiong@gmail.com

1 Umehara Kaoru 梅原郁, “Civil and Military Officials in the Sung: The *Chi-lu-kuan* System,” *Acta Asiatica* 50 (1986): 2.

2 For example, Deng Guangming in the 1950s called the study of bureaucratic institutions one of “the four keys to understanding history.” See Liu Pujiang 劉浦江, “Deng Guangming lüezhuan” 鄧廣銘略傳, in *Xiangnian Deng Guangming* 想念鄧廣銘, ed. Zhang Shilin 張世林 (Beijing: Xinshijie chubanshe, 2012), 25. Denis C. Twitchett, Lai Ruihe, and Charles Hartman also expressed a similar view. See Denis C. Twitchett, *The Writing of Official History under the Tang* (Cambridge: Cambridge University Press, 1992), 2; Lai Ruihe 賴瑞和, *Tangdai zhongceng wenguan* 唐代中層文官 (Beijing: Zhonghua shuju, 2011), 1–4; Charles Hartman, “Introduction: Bureaucratic Institutes and Information Networks,” 2–5.

3 Jin Zhongshu 金中樞, “Songdai guan’ge de jianzhi yu zhishi kao” 宋代館閣的建置與職始考, in *Songdai de xueshu he zhidu yanjiu jinian xuanji* 宋代的學術和制度研究紀念選集 (Xinbei, Taiwan: Daoxiang chubanshe, 2016), 339–87. For the development of the “talent-nurturing in academic institutes” throughout the Song dynasty, see Li Geng 李更, *Songdai guan’ge jiaokan yanjiu* 宋代館閣校勘研究 (Nanjing: Fenghuang chubanshe, 2006), 58–61, 74–77, and 86–90.



searchers have acknowledged the political significance of this policy, noting that a majority of top government leaders in the Song had served in these scholarly institutes early in their careers.<sup>4</sup> However, the general trajectory of officials' careers under this "talent-nurturing" arrangement remains vague. Thus, by examining the bureaucratic transfers of officials who held offices in these scholarly institutes, this study aims to reconstruct the career path(s) designed to cultivate top civilian leadership in the Song government. It focuses on the period from 1131 to 1164, when the Song ruling house, struggling to survive under the threat of the Jurchens, made every attempt to legitimate its authority. One such effort was to claim that it followed the policies of the dynastic founders, including the talent-nurturing policy. These political circumstances provide a window on how the Song court in 1131–1164 upheld a tradition of personnel policy, how dynastic crises impacted its implementation, and what impacts this had on the careers of officials.

The remainder of this article is arranged as follows: the ensuing section explains the methodology and data used in this study. Section 3 gives a brief introduction to the historical context and reconstructs the career trajectories of officials recruited under the talent-nurturing policy between 1131 and 1164. It illustrates a general pattern in the career paths of these officials and identifies the major posts they held that define this pattern. Section 4 elucidates the impacts that the political circumstances of 1131–1164 had on the implementation of the talent-nurturing policy and the officials' careers. To conclude, this article will reflect on the prospects for applying network analysis to more general research on the Song bureaucratic system.

## 1. Methodology and Data

The imperial bureaucracy and its incumbents have long been popular subjects in Chinese studies. A traditional approach to Chinese institutional history explores governmental regulations on the selection and promotion of officials. These studies aim to elaborate the design and the operating logic of the bureaucratic system, but hardly explain the interconnectedness between different personnel decisions.<sup>5</sup> Another approach focuses on the career patterns of officials who staffed the bureaucracy. For instance, Sun Guodong charts the major career

---

4 Umehara Kaoru, "Sōdai no kanshoku" 宋代の館職, in *Sōdai kanryō seidō kenkyū* 宋代官僚制度研究 (Kyoto: Dōhōsha, 1985), 329–422; Li Changxian 李昌憲, "Songdai wenguan tiezhi zhidu" 宋代文官貼職制度, *Wenshi* 文史 30 (1982): 109–35.

5 Edward A. Kracke, Jr., *Civil Service in Early Sung China, 960–1067, with Particular Emphasis on the Development of Controlled Sponsorship to Foster Administrative Responsibility* (Cambridge, Mass: Harvard University Press, 1957); Winston W. Lo, *An Introduction to the Civil Service of Sung China: With Emphasis on Its Personnel Admin-*

paths of officials in different periods of the Tang dynasty (618–907), calculates the length of their tenure in each office, and reveals the changing pattern of bureaucratic transfers during the three centuries of the Tang.<sup>6</sup> Unlike Sun, Robert Hartwell concentrates on a particular group of officials – those in the state’s financial administration – in Song times. By tracing their careers, he shows that finance officials in the early centuries of the dynasty followed a career path that valued specialization in a specific branch of government, which was subsequently replaced by a new career pattern that emphasized generalism.<sup>7</sup>

In the field of network analysis, the political elite has also drawn considerable scholarly attention. Most studies treat the elites as nodes and focus on their relationships, interactions, and the power and influence they derived from their network positions.<sup>8</sup> A few other studies, however, construct networks where the nodes are the positions held by the elites and the edges indicate transfers between these positions. These studies seek to understand the operational logic of the bureaucracy. Josef Woldense, for example, traces the movements of subordinates in government during the reign of the last Ethiopian emperor. The results show that by shuffling his officials frequently, this dictator successfully “reconciled the tradeoff between suppressing potential rivals and encouraging expertise for the proper functioning of the state apparatus.”<sup>9</sup>

Whether they adopt formal network analysis or not, studies of bureaucratic transfers and career patterns are instructive for uncovering the rationale of operation in a bureaucracy. These studies, however, are also limited in several ways. They often encounter the difficulty of determining which offices to focus on and base their decisions on current knowledge of the bureaucracy they study, at the risk of leaving out important but understudied offices. Because of the difficulty of producing a manageable dataset for analysis, these studies also find it inevitable

---

*istration* (Honolulu: University of Hawaii Press, 1987); Deng Xiaonan 鄧小南, *Songdai wenguan xuanren zhidu zhu cengmian* 宋代文官選任制度諸層面 (Shijiazhuang: Hebei jiaoyu chubanshe, 1993).

- 6 Sun Guodong 孫國棟, *Tangdai zhongyang zhongyao wenguan qianzhuan tuijin yanjiu* 唐代中央重要文官遷轉途徑研究 (Hongkong: Longmen shudian, 1978). See also similar studies by Lai Ruihe, *Tangdai jiceng wenguan* 唐代基層文官 (Beijing: Zhonghua shuju, 2004); *Tangdai zhongceng wenguan*; and *Tangdai gaoceng wenguan* 唐代高層文官 (Beijing: Zhonghua shuju, 2017).
- 7 Robert M. Hartwell, “Demographic, Political, and Social Transformations of China, 750–1150,” *Harvard Journal of Asiatic Studies* 42.2 (Dec. 1982): 394–425; “Financial Expertise, Examinations, and the Formulation of Economic Policy in Northern Sung China,” *Journal of Asian Studies* 30.2 (Feb. 1971): 281–314.
- 8 For a comprehensive review of this field in the past fifty years, see Franziska Barbara Keller, “Analyses of Elite Networks,” in *The Palgrave Handbook of Political Elites* (London: Palgrave Macmillan, 2018), 135–152.
- 9 Josef Woldense, “The Ruler’s Game of Musical Chairs: Shuffling during the Reign of Ethiopia’s Last Emperor,” *Social Networks* 52 (2018): 154–166.

to largely simplify the complicated myriad of government organs that existed in the real world. Moreover, they often take a long time span as a coherent unit of analysis and therefore lose sight of how bureaucratic practices may directly reflect the specific political and social changes during a particular period.

This study adopts a similar approach but with some improvements. I focus on a specific group of officials, namely those appointed to academic institutes (*guanzhi* 館職)<sup>10</sup> between 1131 and 1164 under the talent-nurturing policy.<sup>11</sup> I use service experience in Song academic institutes as the criterion for defining the study population, but my scope of data collection covers the entire lifespan of each official included in this study. This practice helps avoid the pitfall of studying only the data on a predefined list of offices. To trace changes in the officials' career patterns over time and map congruent political changes, I have also preserved as much temporal information in my dataset as was available in the sources.

Appointees to the academic institutes played a unique role in the Song bureaucracy. They managed the state collection of books and maps, collated literary and classical works, and compiled official histories based on government archives. Having privileged access to state archives and knowledge resources that were typically not open to other officials, they were expected to offer policy advice on state affairs. To some extent, they served as an equivalent to modern think-tanks. More significantly, the Song personnel authorities used academic institutes as a place to foster talent and groom gifted candidates for high-ranking office. Promising junior officials were first selected to serve in academic institutes and then promoted to influential positions, either in central government agencies or in the provinces. Thus, appointment to the academic institutes gave an official the qualifications and experience necessary to reach the upper echelons of officialdom, allowing them to bypass the typical sequence of promotion that ordinary civil servants were subject to.<sup>12</sup> As a result, many Song officials per-

---

10 In Song times, the term *guanzhi* referred both to positions in the scholarly institutes and to the incumbents of these positions, while the latter was also called *guanzhi guan* 館職官.

11 Categorically called "*guanzhi*" in the historical record, posts in academic institutes include a wide range of offices that played a similar role in different periods of the Song dynasty. During the period of 1131–1164 that this study concerns, these posts were all located inside the Imperial Library. Therefore, this article uses "Imperial Library posts" interchangeably with "academic institute posts" in the following discussion. For an example of official proclamations that linked the talent-nurturing policy to the Imperial Library, see Emperor Xiaozong's edict on 1164/III/3. Xiaozong 宋孝宗, "Guan'ge geng bu li'e zhao" 館閣更不立額詔, in *Quan Song wen* 全宋文, ed. Zeng Zaozhuang 曾棗莊 and Liu Lin 劉琳 (hereafter QSW; Shanghai: Shanghai cishu chubanshe, 2006), 234: 5217.173.

12 Once appointments to the academic institutes became a recognized shortcut to high offices, they were inevitably abused by those in power. For contemporary criticisms of such abuses, see Li Geng, *Songdai guan'ge jiaokan yanjiu*, 76–77.

ceived appointments to the academic institutes as an important launchpad to core government positions.<sup>13</sup>

This article posits that the Song state paced different officials' progress along prescribed career paths so that the most talented could climb the ladder of bureaucracy more quickly than others. The talent-nurturing policy served exactly this objective. As it links offices inside and outside the academic institutes, this Song personnel strategy makes an excellent object of study for us to scrutinize the connectedness between offices in the bureaucracy and the logic behind its operations. Therefore, this study examines the talent-nurturing policy and its impacts by constructing a network of bureaucratic transfers. Each node is a post, and the transfer between any pair of posts is represented by a directed edge connecting the two nodes.<sup>14</sup> The weight of each edge signifies the number of instances of such transfers known from the extant historical records. In constructing this network, I first identified all the appointees to academic institutes between 1131 and 1164 and then traced the careers of each appointee.<sup>15</sup> The dataset contains 3,367 records of bureaucratic transfers, which correspond to a total of 197 appointees during this period, generating a network with 1,042 unique nodes and 2,360 unique directed and weighted edges. Two open-source programs, Gephi 0.9.2 and Visone, are employed for visualization and analysis.<sup>16</sup>

- 
- 13 For a discussion of the responsibilities of officials in the academic institutes, see Li Geng, *Songdai guan'ge jiaokan yanjiu*, 2006. For English scholarship on this topic, see John H. Winkelman, "The Imperial Library in Southern Sung China, 1127–1279: A Study of the Organization and Operation of the Scholarly Agencies of the Central Government," *Transactions of the American Philosophical Society* (New Series) 64.8 (1974): 5–61; Xiong Huei-Lan 熊慧嵐, "A Reservoir of Talent: An Analysis of the Career Advancement of Imperial Library Officials during the Southern Song (1127–1279)," *Journal of Song-Yuan Studies* 48 (Sep. 2019): 7–56.
- 14 The Song bureaucratic system was extremely sophisticated, consisting of several parallel components. Three of them were particularly important: "titular offices" (*guan* 官) which determined an official's rank status, "assignments" (*zhi* 職) which usually served as his additional honorary recognition, and "commissions" (*chaiqian* 差遣, or sometimes translated as "functional posts" or "duty assignments") which specified the official's actual duties. In this study, I consider only the "commissions" of an official and refer to them simply as "offices" or "positions." If an official held several commissions concurrently (*jian* 兼), each instance of concurrent appointments is coded in my dataset as a case of a "bureaucratic transfer" but marked as "concurrent" (as opposed to a "regular" transfer between two offices held in succession).
- 15 For a thorough discussion on the source which I used to retrieve these career records, see Appendix 1. Regarding the technical decisions I made to process and converted the data into computer-readable data for analysis, see Appendix 2.
- 16 Bastian M., Heymann S., Jacomy M. "Gephi: an Open-Source Software for Exploring and Manipulating Networks." International AAAI Conference on Weblogs and Social Media. 2009. Ulrik Brandes and Dorothea Wagner, "Visone: Analysis and Visualization of Social Networks," in *Graph Drawing Software*, ed. Michael Jünger and Petra Mutzel (Heidelberg: Springer Heidelberg, 2003), 321–340.

Two statistical measures, modularity and current-flow betweenness centrality, are used. Modularity analysis detects the presence of distinct communities composed of densely connected nodes within a network. A higher modularity value indicates a greater distinction between these communities. Here I use this measure to differentiate groups of offices that officials held at different career stages.<sup>17</sup> Current-flow betweenness centrality of a node measures its frequency of appearing on all the paths linking any pair of nodes in a network.<sup>18</sup> Generally, Song civil servants moved up the bureaucratic ladder, even though the speed of promotion varied from one official to another. The top office one could hope to reach was a state councilor (*zaizhi* 宰執), which marked the peak – though not necessarily the end – of an official’s career. In such a network where lower-ranking offices are linked progressively to higher-ranking ones, a higher value of current-flow betweenness centrality principally indicates that the office played a more influential role in leading its incumbents to the top civilian leadership. In addition, in the network graphs presented below, the size of a node is proportional to its current-flow betweenness centrality, and the thickness of an edge indicates the number of observed transfers between each pair of offices. For visual clarity, all the nodes in the networks are labeled in Chinese, instead of using the lengthy English translations of Song office titles.

## 2. A Gateway to the Core of Governance

In 1127, the Jurchens invaded northern China and forced the Song court to relocate to the south, beginning an era known as the Southern Song (1127–1279). To secure the loyalty of intellectuals to the dynasty, the restored Song court, under the rule of Emperor Gaozong 高宗 (r. 1127–1162), made a symbolic proclamation that he would continue the policy of “nurturing talent” in academic institutes. This policy, reputedly dating back to the dynastic founders, regarded academic institutes as a training ground for fostering candidates for high-ranking court offices, or even the top government leadership. In 1131,<sup>19</sup> Gaozong approved a collective proposal from his leading officials to reinstate “academic institute posts” (*guanzhi*)<sup>20</sup> and soon re-established the Imperial Library, which had been abol-

- 
- 17 For a discussion of modularity and its application, see “Network Visualization and Analysis with Gephi,” <https://noduslabs.com/courses/network-visualization-and-analysis-with-gephi/units/section-1-quick-introduction-to-network-analysis/?try> (accessed September 18, 2020).
  - 18 For a discussion on current-flow betweenness centrality, see M. E. J. Newman, “A Measure of Betweenness Centrality Based on Random Walks,” *Social Networks* 27.1 (2005): 39–54.
  - 19 Years in this article are converted to the Julian calendar, but months and days, if any, follow the Chinese lunar calendar. The dates are expressed as year/month/day. Intercalary months are indicated with the letter I (e.g., I7).
  - 20 Cheng Ju 程俱, “Lintai gushi houxu” 麟臺故事後序, in *Lintai gushi jiaozheng* 麟臺故事校證 (Beijing: Zhonghua shuju, 2000), 218–219.

ished in 1129 due to the war.<sup>21</sup> Nearly all appointees to offices in the Imperial Library in the first six months of its reestablishment<sup>22</sup> moved on to new posts in other government organs within a year of their appointment.<sup>23</sup> By contrast, the Imperial Library carried out limited cultural functions before 1144: it had neither a proper building nor a good collection of books.<sup>24</sup>

The talent-nurturing policy remained the critical component of the Song bureaucratic practice from Gaozong's reign onwards. Gaozong and his government chose to reestablish the Imperial Library, rather than other scholarly institutes, for grooming promising officials. This decision, as well as to which offices the Library appointees were later transferred, signified how they perceived the dynastic tradition of talent-nurturing policy. These practices developed into a standard protocol for succeeding rulers and their officials of the dynasty. I have shown elsewhere that the number of officials with experience in the Imperial Library steadily increased among the top civilian leadership after 1131, and that this general trend continued throughout the Southern Song.<sup>25</sup> In the following section, I will explain which offices can be considered major posts on the career track of talent-nurturing in the tradition that Gaozong established by using a quantitative analysis of the career records of officials who served in the Imperial Library between 1131 and 1164.

## 2.1 The Mainstream of Career Advancement

Network A is constructed from the career records of officials who were appointed to the Imperial Library between 1131 and 1164 (Figure 1, for the complete version, see Appendix 3). It is comprised of five clusters, which are detected via the Louvain method in Gephi.<sup>26</sup> Three large clusters – colored in blue, green, and yellow – form the main body of the network, supplemented by two small com-

21 Li Xinchuan 李心傳, *Jianyan yilai xianian yaolu* 建炎以來繫年要錄 (henceforth *Yaolu*; Taipei: Academia Sinica, Scripta Sinica Database), 22.1129/4.

22 These officeholders included a Vice Director (*bishu shaojian* 秘書少監), an Assistant Director (*bishu cheng* 秘書丞), a Collator (*jiaoshu lang* 校書郎), and a Proofreader (*zhengzi* 正字), see Xu Song 徐松, *Song huiyao jigao* 宋會要輯稿 (Taipei: Academia Sinica Scripta Sinica Database, 1984; hereafter *SHY*), *Zhiguan* 職官 18.24.

23 The only exception was the proofreader, who first received an internal promotion to the collator in early 1132 and then advanced to an external position later that year. Chen Kui 陳騭, *Nan Song guange lu* 南宋館閣錄 (henceforth *NSGGL*, Beijing: Zhonghua shuju, 1998), 82, 86, 110, and 117.

24 A building was not constructed for the Imperial Library until 1143/12, under the aegis of the Fiscal Supervisorate of Liangzhe. Chen Kui, *Nan Song guange lu*, 2.9.

25 Xiong, "A Reservoir of Talent," 18–25.

26 The network has a modularity of 0.547 when the resolution is set at 1.75. A modularity of 0.4 or above usually indicates the presence of a pronounced community structure in the network. On the Louvain method, see Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre, "Fast Unfolding of Communities in Large Net-

munities (colored in orange and pink in the complete version) at the margins. The horizontal distribution of nodes is roughly in line with the ranks of these offices, with lower-ranking offices on the left and the higher-ranking ones on the right. These clusters constitute the mainstream of the officials' career paths.<sup>27</sup>

The blue and pink clusters each form a circular sector, with the nodes denoting the method of entry into government located at or near the vertices, namely "obtaining a "presented scholar" (*jinshi*) degree via regular examinations" (labeled "進士\_常科"), "graduation from erudite literatus examinations" (labeled "中詞科"), "qualifications via government schools" (labeled "官學"), and "protection privilege" (labeled "恩蔭入仕").<sup>28</sup> Graduation from the regular and erudite literatus examinations and government schools certifies a person's erudition and literary skills, and was deemed a prerequisite for an appointment in the Imperial Library. Of the 197 officials discussed in this study, only one had none of these formal qualifications.<sup>29</sup> Ten officials first entered the officialdom using the protection privilege, but they all managed to gain a *jinshi* degree or acquired equivalent qualifications through the governmental school system later on. The edges in the blue and pink clusters lead through different paths to the green cluster, where Imperial Library offices come into sight. These paths vary in length, ranging from one to three or more steps. This supports the current understanding that the speed of promotion to the Imperial Library varied from one official to another, depending on their individual scholarly accomplishments.<sup>30</sup> The top performers were appointed to the Imperial Library as soon as they had finished one term of service in the local administration, whereas others assumed office in the Imperial Library only after finishing two or more terms in local administration.

The green cluster gathers all the nodes that denote Imperial Library posts, except Vice Director, which appears in the yellow cluster. The green cluster also contains nodes that represent positions held by an official immediately preceding or following his appointment to the Imperial Library. Imperial Library offices

---

works," *Journal of Statistical Mechanics: Theory and Experiment* 10 (2008): 155–68.

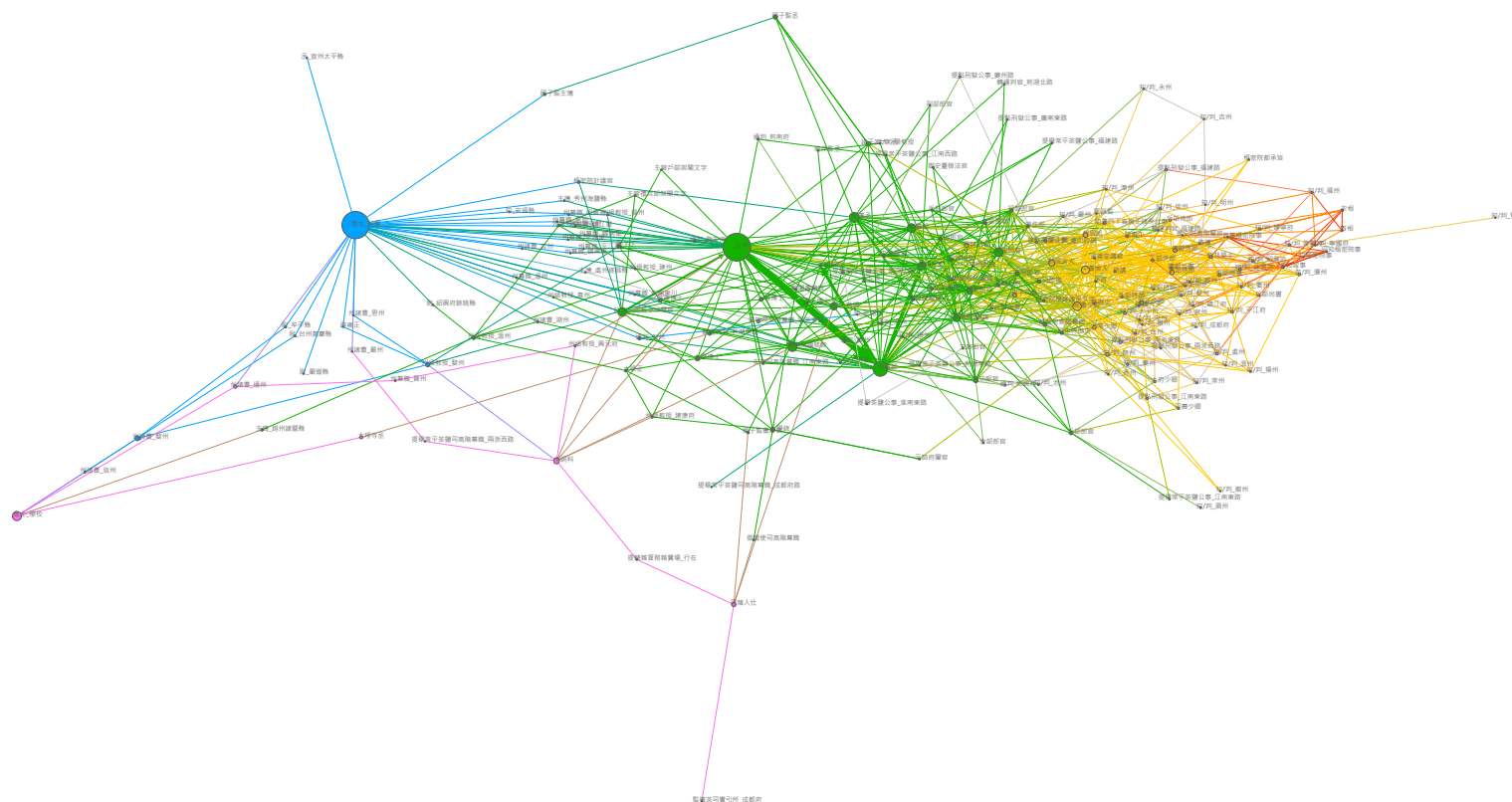
R. Lambiotte, J.-C. Delvenne, M. Barahona *Laplacian Dynamics and Multiscale Modular Structure in Networks*, 2009.

27 Besides these five clusters, in the complete version of Network A, there are also nodes that are scattered across the entire network with weak connections to one another and, sometimes, with no connections at all to any of the five clusters. This means that there are no observed transfers between these posts and those in the mainstream career path. They are evidence of the diverse career experiences of the Imperial Library appointees.

28 For a discussion of different methods of entry and how they are coded in my dataset, see Appendix 2.

29 This man was Emperor Zhezong's brother-in-law, whose appointment to the supervisory position in the Imperial Library was mainly for conferring upon him a distinguished honor, rather than actually serving there. See *Yaolu* 157.478–2 (1157/2/壬子).

30 Zhang Xiqing 張希清, "Songchao gongju shihe shouguan zhidu shulun" 宋朝貢舉釋褐授官制度述論, *Zhongyuan wenhua yanjiu* 中原文化研究 3 (2015): 20–28.



**Fig. 1** Network A. For visual clarity, nodes with a lower current-flow betweenness centrality than that of the Chief Councilor (0.253%) are filtered out. The colors distinguish the communities to which a node is assigned according to modularity analysis.



have the highest scores of current-flow betweenness centrality – not only in this cluster, but also in the network as a whole – and are thus visualized as larger than the others in the graph. This phenomenon is not unexpected, because data on the 197 officials' appointments to and from the Imperial Library are more complete than the career data related to their other offices. Therefore, what truly deserves attention is the non-Library offices with a high value of current-flow betweenness centrality. These include, first of all, the Director of the Bureau of Appointments (*libu si langguan* 吏部司郎官), a leading agency in the Ministry of Personnel, and the Investigating Censor (*jiancha yushi* 監察御史). Other nodes in the green cluster are less noticeable individually, but many are directors and vice directors (collectively known as *langguan* 郎官) of the Twenty-Seven Bureaus (*ershiqu si* 二十七司) in the Department of State Affairs (*shangshu sheng* 尚書省), which were sub-divisions in the executive branch of government. Moreover, all bureaus belonging to the Ministry of Personnel appear in this cluster, but those in other ministries are much less represented.

The edges from the green cluster lead to offices in the yellow cluster, which represents the last constellation of offices in an official's career before he reached the top civilian leadership. The most prominent node, measured by current-flow betweenness centrality, is the Secretariat Drafters (*zhongshu sheren* 中書舍人). This node receives links from the Imperial Diarists in the Secretariat (*qiju sheren* 起居舍人) and those in the Chancellery (*qiju lang* 起居郎) and frequently sends links to the Supervising Secretary (*jishizhong* 給事中) and the Vice Minister of Personnel (*libu shilang* 吏部侍郎). While Ministers and Vice Ministers of the Six Ministries all appear in Network A, those in the Ministry of Personnel stand out in particular.

The orange cluster marks the peak – and sometimes also the terminal stage – of a bureaucratic career. Nodes representing positions in the State Council stand out in this cluster. They include the top government offices from both civilian and military branches, but the civilian leadership is clearly more notable.<sup>31</sup> All the offices in the top civilian leadership appear in this cluster, including the Assistant Chief Councilor (*canzhi zhengshi* 參知政事), the Vice Chief Councilor (*cixiang* 次相, aka the Right Chief Councilor [*youxiang* 右相]), and the Chief Councilor (*shouxiang* 首相, aka the Left Chief Councilor [*zuoxiang* 左相]).<sup>32</sup> By comparison, only some of the supervisory positions in the dynasty's top military lead-

---

31 Top leadership positions in the military branch of government were the supervisory offices in the Bureau of Military Affairs, which were nevertheless filled mostly by civilian officials in Song times.

32 The formal titles of the State Councilors changed frequently during the Song. Whatever their formal titles, they were categorically coded as Chief, Vice, and Assistant Councilors. These distinctions reflect their actual authority and their niche in the bureaucratic hierarchy.

ership – i.e., the Bureau of Military Affairs (*shumi yuan* 樞密院) – appear in the network. Moreover, officials in this study who served in the Bureau of Military Affairs often held top civilian offices concurrently or had only a brief appointment in the Bureau during an interval between two different top civilian appointments.

The aforementioned observations are made based on each node's current-flow betweenness centrality, which adopts an electrical current model to measure the extent of the "current" flowing through a given node between any pair of nodes in the network. In this study, I interpret a high current-flow betweenness centrality as an indication of an office's high degree of influence on the officials' progress toward the top leadership positions.<sup>33</sup> It captures each node's structural prominence in the global structure of the network. To accurately identify the major posts on the 197 officials' career trajectories, I devised a methodology which integrates current-flow betweenness centrality with six other centrality measures for additional information on a node's local significance. I use in-degree centrality of an office to assess how many other offices its incumbents were transferred from, and out-degree centrality to measure how many offices its incumbents were transferred into. Degree centrality of an office ignores the direction of such transfers and simply counts the number of other offices its incumbents were transferred into and out of.<sup>34</sup> To take into account the number of observed transfers in the historical records for each pair of offices, I also include the weighted versions of these three centrality measures. The current-flow betweenness centrality is calculated in Visone,<sup>35</sup> and all the other centrality scores are calculated in Gephi. This generates a total of seven ranking lists for the nodes, one list per centrality measure. Because of a significant overlap among the top-ranking nodes on these lists, this produces a list of 38 unique nodes that are ranked among the top twenty

- 
- 33 The calculation of current-flow betweenness centrality takes into consideration all possible paths between any pair of nodes, although it gives more weight to shorter paths than to longer ones. It therefore differs from the classical definition of betweenness centrality, which considers only the shortest path between each pair of nodes. Current-flow betweenness centrality is more appropriate for the present study because officials took different paths from one office to another in their careers, not necessarily the shortest ones. For a comparison between different betweenness centrality measures, see Newman, "A Measure of Betweenness Centrality Based on Random Walks," 39–54.
- 34 Offices that were usually held concurrently or at the end of bureaucratic careers will have higher in-degrees and lower out-degrees. Conversely, an office with a lower in-degree and a higher out-degree centrality was usually held by officials at the start of their careers. Since available career data is incomplete for some of the 197 officials, a measure of network-wide structural prominence (i.e., current-flow betweenness centrality) may not capture the full significance of the offices they held, which makes it necessary to also include the other six centrality scores that measure the local significance of these offices.
- 35 For a discussion of Visone's algorithm for current-flow betweenness centrality, see Michael Baur, *Software for the Analysis and Visualization of Social Networks* (doctoral dissertation, Universität Fridericiana zu Karlsruhe, 2008), 42.

on at least one of these lists.<sup>36</sup> For each of these nodes, I count the number of lists where it is ranked among the top twenty, and the results are reported in Table 1. If a node is ranked among the top twenty on four or more of these lists, it is considered a major post on the career trajectories of the 197 officials in this study.

Not surprisingly, Imperial Library offices stand out in the top-ranking nodes, in part because I have limited the scope of this study to those officials who had experience in the Imperial Library. Of these offices, the Proofreader, the Collator, and the Assistant Director were the most important, which often marked the start of an official's service in the Imperial Library.<sup>37</sup> Next to them in structural prominence are the Director and Vice Director of the Bureau of Appointments in the Ministry of Personnel, the Secretariat Drafter, the Investigating Censor, Imperial Diarists in the Secretariat and the Chancellery, the Vice Minister of Personnel, the Supervising Secretary, and the Hanlin Academician (*Hanlin xueshi* 翰林學士). Although these offices were scattered across a variety of government organs, many of their incumbents were deeply involved in the making of personnel decisions, which gave incumbents of these offices an opportunity to influence the staffing of the government and its policies.<sup>38</sup> It was precisely these offices that marked the key steps on a talented official's career path towards the top civilian leadership.

Offices responsible for enforcing discipline and conducting performance reviews are also present among the top-ranking nodes. Chief among them are the

---

36 See Appendix 4 for all seven centrality scores of these 38 nodes. I focus on the top twenty offices because a Song official usually held a total of ten to fifteen offices in his entire career. This estimation is based on an average career length of thirty years and a typical tenure of two years or less in each office. Wu Zhihao 吳志浩, "Songdai shiren pingjun siwang nianling kao" 宋代士人平均死亡年齡考, *Zhejiang xuekan* 浙江學刊 4 (2017): 170–181.

37 Xiong, "A Reservoir of Talent," 33–34.

38 Take appointments to the Imperial Library, for example. When a vacancy arose in the Imperial Library, the emperor would instruct senior court officials to nominate candidates. These senior officials, categorically known as the "Two Drafting Groups" (*liangzhi* 兩制) and the "Ministers-in-Attendance" (*shicong* 侍從), included the Hanlin Academicians, Secretariat Drafters, Supervising Secretaries, and Ministers and Vice Ministers, which were exactly those offices represented by the top-ranking nodes in Table 1. The nominees would then take a special examination designed by the Hanlin Academicians or other senior court officials, and the graduates from the examination would be appointed to the Imperial Library. After these appointment decisions were reviewed by State Councilors, formal appointment orders would be drafted by the on-duty Secretariat Drafter, approved by State Councilors, verified by the Department of the Secretariat, and inspected by the on-duty Supervising Secretary in the Chancellery. The Ministry of Personnel was responsible for its execution once the order was finalized. Additionally, at every stage in the above process, censors and remonstrators could step in and voice their objections, if any. By refusing to draft the appointment orders and sending them back for reconsideration, the Secretariat Drafter and the Supervising Secretary also had the power to protest against a decision they deemed inappropriate.

Nodes	正字	校書郎	秘書丞	吏部郎官	秘書郎	秘書少監	中書舍人	著作佐郎	監察御史	起居舍人	禮部郎官	起居郎	禮部侍郎	吏部侍郎	給事中	進士——常科	翰林學士	官學	樞密院編修官	侍講
Times ranked high	7	7	7	7	7	7	7	7	7	7	7	7	7	7	6	5	5	4	4	4

Nodes	詳定一司敕令所屬官	次相	參知政事	直學士院	侍讀	史館屬官	都督府高階幕職	工部侍郎	王府教授	宗正少卿	進士——特賜	中詞科	太學博士	州諸曹——婺州	殿中侍御史	太常少卿	知／判——饒州	大宗正司丞
Times ranked high	3	3	3	3	3	2	2	2	2	2	1	1	1	1	1	1	1	1

**Tab. 1** Summary of top-ranking nodes by seven centrality measures, indicating the number of the seven centrality measures by which the node ranks high. Major posts identified on the officials' careers are highlighted in green. Imperial Library offices are in blue.

directors of various bureaus in the Ministry of Personnel, whose duties included making appointment decisions, conducting regular performance reviews, deliberating on the use of the protection privilege, merits, honors, and promotions. Another group of top-ranking nodes represent the offices in remonstrance and censorial organs, whose incumbents were collectively known as the “speaking officials” (*yanguan* 言官).<sup>39</sup> Among them, the Investigating Censor has the highest score of current-flow betweenness centrality (1.276), followed by the Grand Master of Remonstrance (*jianyi dafu* 諫議大夫, 0.161), the Exhorter (*zhengyan* 正言, 0.152), and the Vice Censor-in-chief (*yushi zhongcheng* 御史中丞, 0.118).

While incumbents of the above offices had a direct role to play in matters pertaining to personnel decisions, the Imperial Diarists and officials in the Imperial Library were also involved in these matters, albeit in an indirect way. Imperial Diarists recorded policy discussions between the emperor and his officials (e.g., state councilors and those who were granted an audience) and among court officials themselves. Appointment decisions were a topic that frequently came up in these discussions. These records provided the sources for the writing of official histories, which was mainly the responsibility of Editors and Assistant Editors in the Imperial Library, with occasional assistance from others.<sup>40</sup> The routine practice of archiving personnel orders and using them for history writing may have had the effect of deterring personnel authorities from appointing unqualified candidates.

In brief, most of the major posts identified in Network A constituted an institutional chain pertaining to personnel management, ranging from nomination and appointment to enforcing discipline and conducting reviews, and to documentation and archival preservation. Although each office had its own distinct duty, the evaluation of an official’s moral character and administrative abilities was a central concern for all of these offices. This signifies that a critical part of preparing an official for civilian leadership in the Song involved cultivating and testing his ability to identify talented candidates (*shiren zhineng* 識人之能). The practice also resonates with contemporary expectations that the fundamental duty of government leaders was to choose the right people to occupy the right offices.<sup>41</sup>

This notion of good leadership was grounded in the talent-nurturing policy, making the selection of academic institute appointees particularly significant as

39 Yu Yunguo 虞雲國, *Songdai taijian zhidu yanjiu* 宋代臺諫制度研究 (Shanghai: Shanghai renmin chubanshe, 2014). Hartman, “Sung Government and Politics,” 103–112.

40 Cai Chongbang 蔡崇榜, *Songdai xiushi zhidu yanjiu* 宋代修史制度研究 (Taipei: Wenjin chubanshe, 1993). Charles Hartman, *The Making of Song Dynasty History: Sources and Narratives* (Cambridge: Cambridge University Press, 2021), 1–19.

41 Zhuge Yibing 諸葛憶兵, *Songdai zaifu zhidu yanjiu* 宋代宰輔制度研究 (Beijing: Zhongguo shehui kexue chubanshe, 2000), 142–150.

this type of post stood at the beginning of the aforementioned office-chain. As Sun Sheng 孫升 (*jinshi* of 1165), a Palace Censor in 1089, remarked, intellectual elites considered the fulfillment of talent-nurturing a key criterion for assessing a state councilor's competency:

The foundation of peace and prosperity lies in seeking out the worthy men and no attainment of state councilors is greater than recommending talents ... State councilors intent on [bringing about peace and prosperity in] the realm must give priority to [the recruitment of] talents ... Our dynastic ancestors established the academic institutes and recruited eminent men from all corners of the realm, favoring them with generous salaries and distinguished honors and appointing them to offices [more prestigious than what] their seniority qualified them for. To test them, [our dynastic ancestors] gave them weighty and demanding assignments at court and in the provinces and observed their behavior and performances. Not until their abilities and achievements were perfected were they employed in high offices. This is the foundation for achieving peace and prosperity ... When our dynastic ancestors promoted officials to the State Council, they always made them recommend several prominent scholars [for offices in academic institutes]. Judging from their recommendations, officials at court and in the provinces could tell whether the newly appointed state councilors were worthy men or not, and future generations could also evaluate the depth of their learning and visions.<sup>42</sup>

Sun's remark underscores that it was not academic institute posts alone, but rather their combination with other essential posts that constituted the model track of "talent-nurturing." These offices figured prominently in the careers of the Imperial Library appointees who benefited from the talent-nurturing policy in the early Southern Song, and a similar pattern is also noticeable in officials' career trajectories in the Northern Song and the Tang (618–907).<sup>43</sup>

The foregoing analysis demonstrates that the revival of the Imperial Library and the reintroduction of the talent-nurturing policy in the early Southern Song served to open the gateway to core leadership positions. In an earlier study, I argued that appointments to the Imperial Library coincided with a preference for training and employing generalists in government service.<sup>44</sup> The analysis in this article takes this argument a step further. It shows that cultivating good leaders in the Song meant not only developing their skills for handling the tasks in each

42 Sun Sheng 孫升, "Qizhao dachen shoujian mingshi zou" 乞詔大臣首薦名士奏, QSW 93: 2021.100.

43 Gong Yanming 龔延明 ed., *Songdai guanzhi cidian* 宋代官制辭典 (Beijing: Zhonghua shuju, 1997), 681–682. Sun, *Tangdai zhongyang zhongyao wenguan qianzhuan tujing yanjiu*, 14–142; Lai, *Tangdai jiceng wenguan*, 13–98; *Tangdai zhongceng wenguan*, 49–92, 127–206; *Tangdai gaoceng wenguan*, 125–144.

44 Xiong, "A Reservoir of Talent," 50.

specific division of government, but also involved assessing and developing their abilities to evaluate the character and competence of their fellow officials.

## 2.2 Through the Lens of Graph Theory

Section 2.1 reveals the mainstream career path of the 197 officials under discussion. In this section, I will turn to divergences in their careers. Based on the positions they held in the Imperial Library, these officials can be divided into three groups (Table 2). Group A, comprised of 146 officials, are those who only held subordinate positions in the Imperial Library; Group B contains 19 officials who only held supervisory positions in the Library; and Group C, consisting of the remaining 32 officials, had experience in both types of positions in the Library. There was a wide gap in rank between the supervisory and subordinate offices in the Imperial Library.<sup>45</sup> The Vice Director of the Imperial Library, for example, was ranked 5b, whereas all the subordinate offices were ranked between 8b and 7b.<sup>46</sup> This raises the question: was the career pattern of the 19 officials who held only supervisory positions in the Library significantly different from that of the other officials?

To answer this question, I partition the data used in section 2.1 into three subsets, one for each group of officials, and then construct a network of bureaucratic transfers using each subset of data. The same methods of analysis used in section 2.1 are applied to each of these networks, which allows us to identify the major posts on the career path of each group. Table 3 compares the top-ranking posts across the three networks. Several of the top-ranking posts identified on the mainstream career path for the entire population of the 197 officials remain important on the career path of each of the three subgroups. These posts include the Investigating Censor (7b), the Imperial Diarist (6b), the Secretariat Drafter (4a), the Supervising Secretary (4a), the Vice Minister of Personnel (3b), and the Vice Minister of Rites (3b). The node denoting the attainment of a *jinshi* degree in regular examinations also figures prominently on the career paths of all three subgroups.

However, these broad similarities aside, the differences are also notable. First, all of the above posts, with the exceptions of the Investigating Censor and the attainment of *jinshi* degree in regular examinations, occupy structural positions

---

45 Offices in the Song bureaucracy were classified into nine ranks, and each rank was further divided into two or four classes. Following the convention, I use Arabic numerals to indicate the rank (1 for the highest rank and 9 for the lowest) and letters to indicate fine gradations inside each rank.

46 The Director of the Imperial Library was ranked at 4a, but the position was left vacant between 1131 and 1164. The Proofreader and Collator were both ranked at 8b, the Secretary and the Assistant Editor 8a, and the Assistant Director and the Editor 7b.

Group	Num. of Officials	Percentage	Notes
A	146	74.11%	Officials who held only subordinate positions. 3 out of 146 served both before and after 1131; 143 served only after 1131.
B	19	9.64%	Officials who held only supervisory positions. All served after 1131.
C	32	16.24%	Officials with experience in both supervisory and subordinate positions. 8 out of 32 held subordinate offices before 1127 and were promoted to supervisory offices after 1131. 24 held subordinate and supervisory offices only after 1131.
Total	197	100.00%	

**Tab. 2** Officials in the Imperial Library between 1131 and 1164 by Service Experience.

that are similar in the careers of Group A and C, but not in that of Group B. Second, although the career paths of the 197 officials sometimes overlapped, they often diverged again soon afterwards. When officials in Group B were promoted, they often moved on to posts that were very different from what officials in Groups A and C were usually transferred to. Third, top leadership offices are conspicuous on the career paths of Groups A and C but absent on that of Group B. These observations are illustrative of the different paths taken by officials who held different types of positions in the Imperial Library.

Scholars have pointed out the existence of multiple paths of career advancement in Song officialdom.<sup>47</sup> These studies each capture a certain aspect of the Song officials' career patterns, but we will still benefit from a more holistic perspective provided by a theoretical model of directed acyclic graphs. This model, known as "the four-continent theory," is based on the work of Serguei Dorogovtsev, et al. Albert-László Barabási describes the model as follows:

Directed networks such as the World Wide Web naturally break down into several easily identifiable continents. In the central core each node can be reached from every other node. Nodes in the IN continent are arranged such that following the

47 Focusing on the bureaucratic practices in the early Northern Song, twentieth-century historian Umehara Kaoru charted the different career paths of officials who took different routes to enter government service. Umehara Kaoru 梅原郁, "Sōdai no bunkai" 宋代の文階, in *Sōdai kanryō seidō kenkyū* (Kyoto: Dōhōsha, 1985), 10–79. Deng Xiaonan has further explicated the divergent career paths in Song officialdom using institutional records and other source materials. Deng Xiaonan, *Songdai wenguan xuanren zhidu zhucengmian*, 181–191.



Node	Bureaucratic Rank	Num. of Centrality Measures by Which It Ranks High		
		Group A	Group B	Group C
進士_常科	Methods of Entry	5	5	5
中詞科		1	—	—
進士_特賜		—	3	1
官學		5	—	—
恩蔭入仕		—	—	1
州諸曹_婺州	9b	1	—	—
州幕職_溫州	8b	—	1	—
州級教授_溫州	8b	1	—	—
詳定一司敕令所屬官	8b	2	1	5
樞密院編修官	8a	6	4	—
太學博士	8b	1	—	—
正字	8b	7	—	7
校書郎	8b	7	—	7
辟雍博士	8a	—	—	1
秘書郎	8a	7	—	1
著作佐郎	8a	7	—	7
秘書丞	7b	7	—	5
王府教授	7b	2	—	1
資善堂講官	7b/7a	—	5	1
監察御史	7b	7	7	7
殿中侍御史	7a	—	7	—
吏部郎官	7a/6b	7	2	7
禮部郎官	7a/6b	7	—	4
考功郎官	7a/6b	—	7	—
司封郎官	7a/6b	2	1	—
祠部郎官	7a/6b	3	—	—
戶部郎官	7a/6b	1	—	—
屯田郎官	7a/6b	—	1	—
左司郎官	6b/6a	—	4	—
右司郎官	6b/6a	—	1	—
起居郎	6b	6	7	5
起居舍人	6b	7	4	7
國子司業	6a	2	—	—
太常少卿	5b	—	7	4
宗正少卿	5b	3	6	—

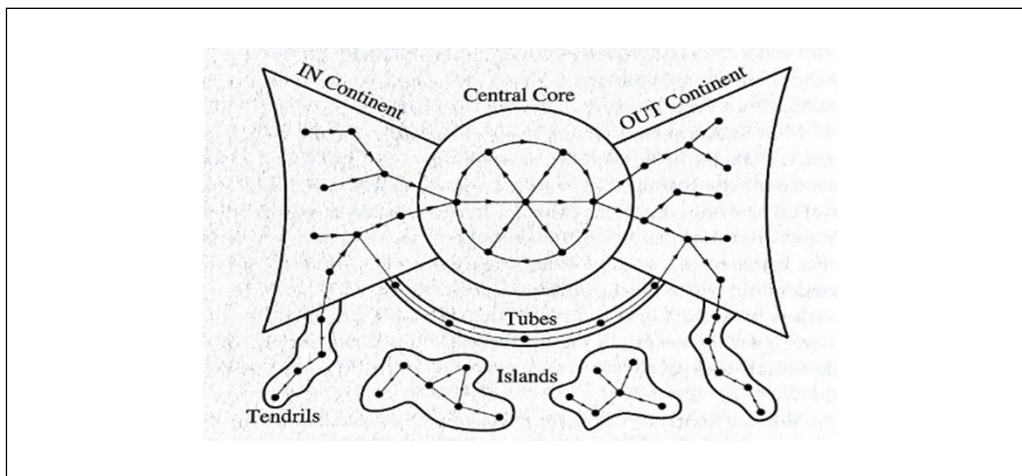
Node	Bureaucratic Rank	Num. of Centrality Measures by Which It Ranks High		
		Group A	Group B	Group C
秘書少監	5b	—	7	7
國子祭酒	4b	—	—	1
中書舍人	4a	7	6	7
給事中	4a	4	—	6
吏部侍郎	3b	7	6	7
戶部侍郎	3b	—	—	2
禮部侍郎	3b	6	7	7
兵部侍郎	3b	—	4	6
工部侍郎	3b	—	6	1
翰林學士	3a	1	3	7
吏部尚書	2b	—	7	3
兵部尚書	2b	—	—	2
參知政事	2a	2	—	2
次相	1b/1a	2	—	4
首相	1b/1a	—	—	3
史館屬官	Ranks vary, depending on officeholders' individual situations	2	—	—
國史院屬官		2	—	—
史館高階屬官		—	2	—
國史院高階屬官		—	—	2
都督府高階幕職		2	—	—
侍講		3	4	4
崇政殿說書		—	—	1
侍讀		2	3	2
直學士院		2	—	3
知/判_潭州		—	—	2
知/判_紹興府		—	7	—
知/判_宣州		—	3	—
知/判_婺州		—	2	—
提舉常平茶鹽公事_荊湖南路		—	1	—

**Tab. 3** Comparison of top-ranking nodes on the career paths of officials in Groups A, B, and C. Offices in the table are arranged in ascending order of bureaucratic rank. Nodes highlighted in blue have similar levels of structural prominence across the three networks, while those in green figure prominently on the career paths of Group A and C, but not on that of Group B. The “—” sign indicates that the node is not among the top twenty by any centrality measure.

links eventually brings you back to the central core but starting from the core doesn't allow you to return to the IN continent. In contrast, all nodes of the OUT continent can be reached from the core, but once you've arrived, there are no links taking you back to the core. Finally, there (are tubes that) directly connect the IN to the OUT continent; some nodes form tendrils, attached only to the IN and OUT continents; and a few nodes form isolated islands that can't be accessed from the rest of the nodes (see Figure 2).<sup>48</sup>

The career pattern of Song officials bears marked similarities to this model. The Song officials were classified into three tiers – “men of selection” (*xuanren* 選人), “capital officials” (*jingguan* 京官), and “court officials” (*chaoguan* 朝官) – which, in spite of how they were called, indicated not whether an official had an assignment at court, in the capital, or elsewhere, but rather functioned as markers of rank and prestige.<sup>49</sup> For a Song official, a promotion in rank from a man of selection to a capital official, known as a “change of status” (*gaiguan* 改官), was of critical importance to his career. An official who failed to rise into the tier of capital officials had no prospect of receiving an appointment beyond the low-ranking local positions. The change of status was also irreversible: once a man of selection became a capital official, he would retain this status and be eligible for more prestigious positions at court and in higher-level local administrations. There is a notable parallel between this practice and the “four-continent theory,” where links between nodes run unidirectionally from the “IN Continent” to the “Central Core.” Furthermore, achieving the status of capital officials was highly competitive, and those who successfully gained this status did so in different ways. As discussed in section 2.1, in the case of those men studied in this article, it was their appointments to the Imperial Library that facilitated their change of status from “men of selection” (the “IN Continent”) to “capital officials.” For other Song officials, however, the change of status could be a much tougher and more complicated process.<sup>50</sup>

- 
- 48 Albert-László Barabási, *Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life*. (New York, NY: Perseus Publishing, 2002; 2014 reprinted), 166. S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, “Giant Strongly Connected Component of Directed Networks,” *Physical Review E* 64.2 (2001). I appreciate Professor Albert-László Barabási for his generosity and consent to using the image of the model of the Four-Continent Theory in this article.
- 49 For a brief discussion of the Song bureaucratic system, see Charles Hartman, “Song Government and Politics,” in *The Cambridge History of China. Vol. 5, Part Two: Sung China, 960–1279*, ed. John W. Chaffee and Denis Twitchett (Cambridge, UK: Cambridge University Press, 2015), 55–71. Hucker translated *xuanren* 選人 as the “Selectmen,” while Hartman translates it as the “men of selection.” I adopt Hartman’s translation here, which I think better captures the meaning of *xuanren* in Song times (i.e., men who were qualified for holding offices and therefore entered the pool of candidates for selection).
- 50 Hukun 胡坤, *Songdai jianju gaiguan yanjiu* 宋代薦舉改官研究 (Shanghai: Shanghai guji chubanshe, 2019), 53–81, 149–219.



**Fig. 2** The Model of the Four-Continent Theory. Image credit: see footnote 48.

A comprehensive examination of the four-continent theory and its potential for the Song bureaucratic system is beyond the scope of this study. It suffices to note that this theory is an invaluable source of inspiration for deciphering the Song bureaucratic practices. Take the talent-nurturing policy, for example. One may conceptualize the entire Song bureaucracy as a network consisting of multifarious offices ranked at different levels and located in diverse government organs. In this network, the talent-nurturing policy provided a mechanism that not only led talented officials out of the IN Continent, but also gave them a shortcut to the higher echelons of government by way of appointments to the Imperial Library. The question remains whether such appointments should best be theoretically conceptualized as a path which leads to the Central Core or as a “tube” which connects the IN Continent directly with the OUT Continent. Questions like this require further research in the future, to compare career records of officials who had experience in the Imperial Library with those who did not.

Although the four-continent theory provides a useful conceptual framework for understanding the Song bureaucracy from a network perspective, adjustments to the theory are also necessary to develop a more sophisticated model that accurately depicts the Song officials’ career paths. For example, Barabási describes the connection between the Central Core and the OUT Continent as unidirectional and irreversible, but this does not map neatly to the Song bureaucratic practices. No path seems to have existed for Song officials to leave the officialdom altogether: those who finished their tenure in the State Council could take up a position in prefectural or circuit administrations or receive a custodian sinecure, but as long as they were alive, they could always be summoned back to the court for high-ranking appointments should the need arise. That Song officials moved frequently between prestigious court appointments, including those to the Imperial Library and the State Council, and local administrative positions suggests

that it may be more instructive to conceive of the Central Core as itself composed of both core and peripheral areas. Movement between offices in these two areas depended heavily on shifting political circumstances. This is the topic to which I will turn now.

### 3. A Decade-Long Setback during the Qin Gui Administration (1140–1155)

Emperor Gaozong and his court revived the talent-nurturing policy to secure the support of scholar-officials for his rule. The mainstream career path charted in section 2 shows that the talent-nurturing policy was clearly in effect during Gaozong's reign and that it opened a gate for appointees in the Imperial Library toward the core leadership positions. On this career path, these appointees were involved, though in different ways, in the Song personnel administration, a practice which tested and developed their abilities to evaluate the character and competence of their fellow officials. However, the degree to which the talent-nurturing policy was implemented changed over the thirty-five years of Gaozong's reign (1127–1162) in tandem with the shifting diplomatic and domestic situations.

During the early decades of the Southern Song, military tension and the diplomatic relationship with the Jurchens posed a tough challenge. Gaozong's attitude toward these issues strongly determined his choice for the top leaders of his administration, who in turn exercised a great influence on the choice of men for other government positions. Modern scholars divide the early Southern Song into three phases.<sup>51</sup> In the first phase, from 1127 to 1140, the Song policy towards the Jurchens wavered between war and appeasement, and the court valued collective deliberation in its decision-making process, a time-honored tradition of the dynasty.<sup>52</sup> In the second phase, between 1141 and 1155/10, Gaozong finally made a resolution to sign a peace treaty with the Jurchens and granted his Chief Councilor Qin Gui exclusive authority to conduct negotiations. In this phase, Qin Gui served as the sole State Councilor with the full support of the emperor, monopolizing the central government. This phase ended with Qin's death in 1155/10. In the subsequent decade (1155/10–1164), the Song administration returned to its tradition of collective deliberation at court. For most of this decade, the Song maintained peaceful diplomatic relations with the Jurchens, but hawkish voices

---

51 For a discussion of the political history of the early Southern Song, see Teraji Jun 寺地遵, *Nan Song chuqi zhengzhishi yanjiu* 南宋初期政治史研究 (Taipei: Daohe chubanshe, 1995); Yu Yunguo 虞雲國, "Shaoxing tizhi yu Nan Song shi zhu wenti" 紹興體制與南宋史諸問題, in *Nandu junchen: Song Gaozong ji qi shidai* 南渡君臣: 宋高宗及其時代 (Shanghai: Shanghai renmin chubanshe, 2019), 1–36.

52 Hartman, "Song Government and Politics," 97–102.

also emerged. War broke out again with the Jurchens in 1161, and peace was not restored until the two parties signed a new treaty in 1164.

This section examines whether the talent-nurturing policy was implemented differently in each of the above three phases. Using these phases as benchmarks, I assign two types of temporal attributes to all edges (Table 4). The first is what I call the “transition attribute,” which indicates when an observed instance of transfer between two offices took place. This attribute has three possible values: prior to 1141 (Phase One), between 1141 and 1155/10 (Phase Two), and after 1155/11 (Phase Three). The other temporal attribute is the “cohort attribute.” To give each edge a cohort attribute, I start with the nodes. I first classify the 197 officials under discussion into three cohorts based on the year of their first appointment to the Imperial Library after its revival in 1131: those first appointed to the Imperial Library between 1131 and 1140, those between 1141 and 1155/10, and those between 1155/11 and 1164.<sup>53</sup> I then give each edge the same cohort attribute as that of the official whose career the edge pertains to.

The distinction between these two types of temporal attributes is important to the following analysis. The life expectancy of Song officials was about fifty-three years and the length of their careers, measured from the time they entered government service to their retirement or death, averaged around three decades.<sup>54</sup> An official’s first appointment to the Imperial Library often took place in the second decade of his career, but some, by virtue of extraordinary academic achievements or out of pure luck, might land a post in the Imperial Library as early as the first decade of their careers. Therefore, some of the 197 officials had careers that spanned two or more phases. Using two temporal attributes for each edge allows us to determine when an appointment to the Imperial Library was made to which cohort of officials and therefore permits a closer examination of how the talent-nurturing policy was implemented in the early Southern Song.

Using the transition attribute, I divide my dataset into three subsets and construct a network (B1, B2, and B3) from each subset of data.<sup>55</sup> Each network corresponds to personnel decisions made in one of the three phases of Gaozong’s reign. Edges in each network are colored differently based on their cohort attribute (blue, yellow, and green for Cohorts One, Two, and Three, respectively). In

---

53 Of the 197 officials who held office in the Imperial Library between 1131 and 1164, some had already served there prior to 1129 when the Imperial Library was abolished. For simplicity, I have ignored appointments to the Imperial Library before 1129 in assigning values for the cohort attribute. Only appointments to the Imperial Library after its revival in 1131 are considered.

54 Wu Zhihao, “Songdai shiren pingjun siwang nianling kao,” 170–181.

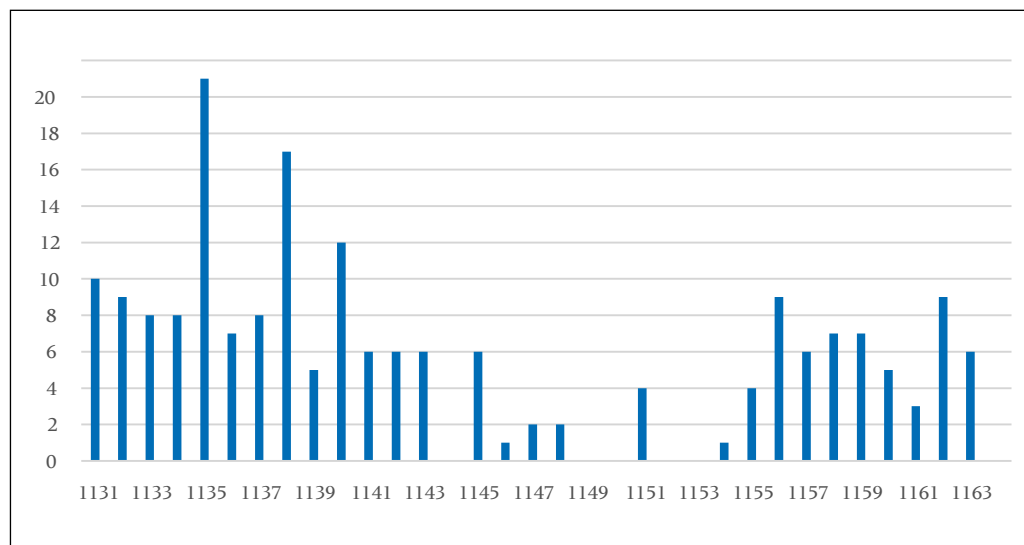
55 Network B1 is constructed from data on appointments before 1141, Network B2 from data between 1141 and 1155/10, and Network B3 from data after 1155/11.

Attr. Type	Attr. Code	Data description
Transition Attribute	Phase One	Transferred to a new Imperial Library post prior to 1141
	Phase Two	Transferred to a new Imperial Library post between 1141 and 1155/10
	Phase Three	Transferred to a new Imperial Library post after 1155/11
Cohort Attribute	Cohort One	Officials first appointed to the Imperial Library between 1131 and 1140. Edges pertaining to this cohort are colored in blue.
	Cohort Two	Officials first appointed to the Imperial Library between 1141 and 1155/10. Edges pertaining to this cohort are colored in yellow.
	Cohort Three	Officials first appointed to the Imperial Library after 1155/11. Edges pertaining to this cohort are colored in green.
<b>The network of bureaucratic transfers is partitioned into three based on the transition attribute of each edge:</b>		
Network B1	Contains only edges with a transition attribute of Phase One.	
Network B2	Contains only edges with a transition attribute of Phase Two.	
Network B3	Contains only edges with a transition attribute of Phase Three.	

**Tab. 4** Partition of network data based on temporal attributes.

what follows, I will compare the patterns of bureaucratic transfers in the three networks and use the mainstream career path identified in section 2.1 as a reference point for evaluating the appointments in each network. These comparisons will reveal whether political conditions in different phases of the early Southern Song (i.e., a government of collective deliberation versus a highly autocratic administration under Gaozong and Qin Gui) had different impacts on the personnel arrangements and the practice of the talent-nurturing policy.

A brief look at the data reveals that the number of officials who received their first appointments to the Imperial Library fluctuated greatly between 1131 and 1164 and that there was a pronounced deviation during the Qin Gui administration. This number started at a total of 105 between 1131 and 1140 (10 years) and declined to merely 38 between 1141 and 1155/11 (15 years) before it rose again to 52 between 1156 and 1164 (9 years). The annual number of new appointees fluctuated just as much (Figure 3): it dropped dramatically after 1141 and went all the way down to zero in some years of Qin Gui's administration. A structural analysis of the networks B1, B2, and B3 provides more details on the differences between these three phases. Using the same methods of analysis as employed in section 2, I have identified, for each network, the nodes that rank among the highest according to at least one of the seven centrality measures (Table 5). The following analysis will be based on the rankings of these nodes in each network, with par-



**Fig. 3** Number of First-Time Appointees to the Imperial Library by Year of Appointment. The vertical axis denotes the annual number of first-time appointees, and the horizontal axis indicates the year of appointment.

ticular attention given to their current-flow betweenness centrality values (see visualization in Figure 4).

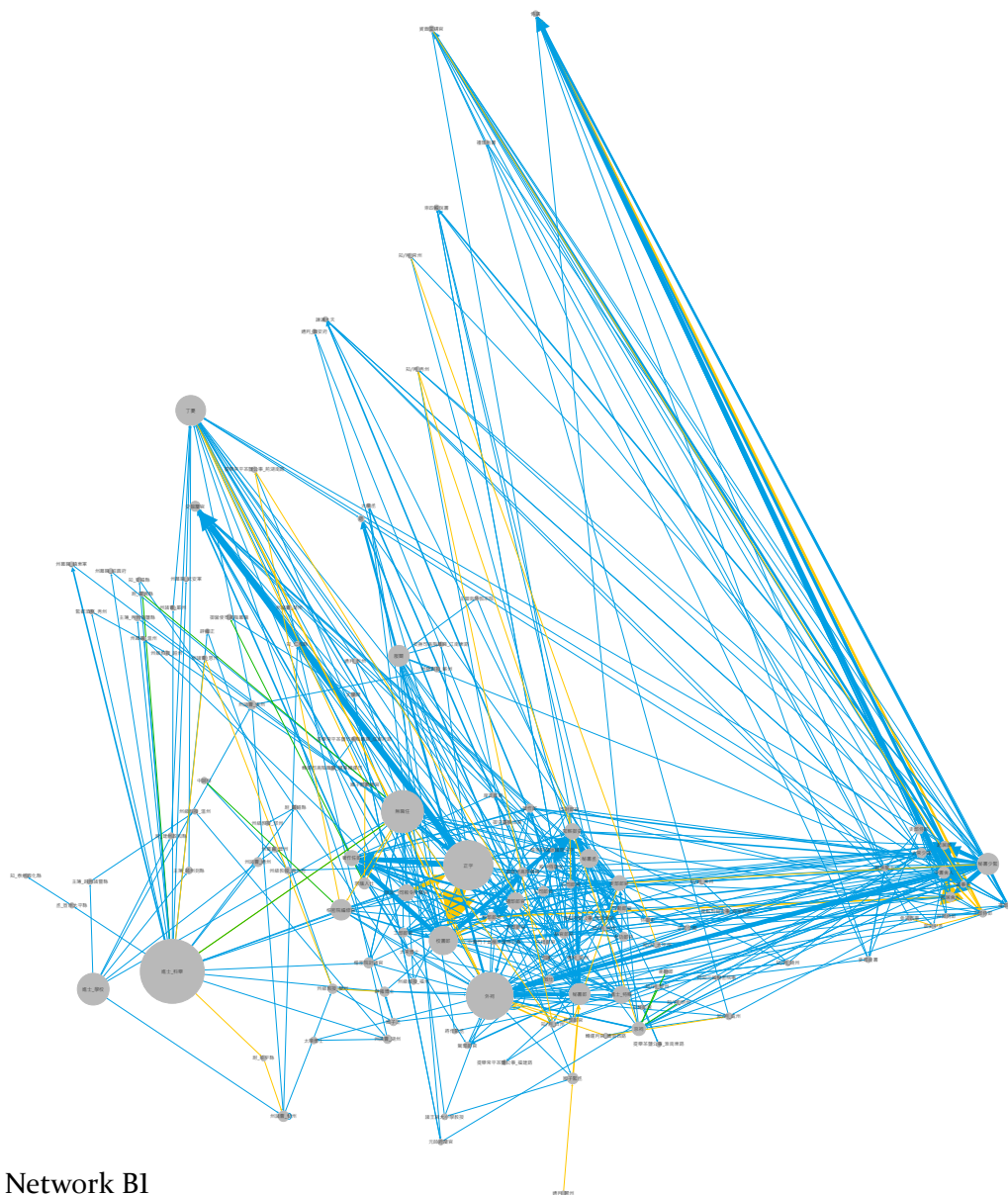
At first glance, the three networks have a noticeable feature in common: “custodian sinecure outside the capital” (*waici* 外祠) and “inactive” (*wuzhiren* 無職任) are among the top nodes in all three networks, and their current-flow betweenness centrality scores are also among the highest. This suggests that the prospect of office-holding was unstable in all three phases of the early Southern Song. The officials in this study frequently left their then-current offices without a new duty assignment, or with merely a sinecure. Such arrangements, as explained in Appendix 1, were not uncommon but implied unusual political situations. What is also worthy of note is the location of these two nodes in each network and what other nodes they are connected to. In Network B3, these two nodes possess many outgoing links from those nodes that denote the high-ranking offices and top leadership positions. In comparison, in Networks B1 and B2, these two nodes are more often linked to nodes that represent low- and middle-ranking offices. This indicates that in Phases One and Two, the officials under discussion often encountered career frustrations at the middle stage of their bureaucratic life. Although they were “talented officials” chosen for nurturing, they failed to receive a smooth promotion to the upper echelons of government, as predicted in the mainstream career path outlined in section 2.1. Instead, they were often given a sinecure outside the capital, relieved of active duty, or appointed to local government positions. This phenomenon becomes more pronounced when transfers from central to local government positions are also taken into consideration.

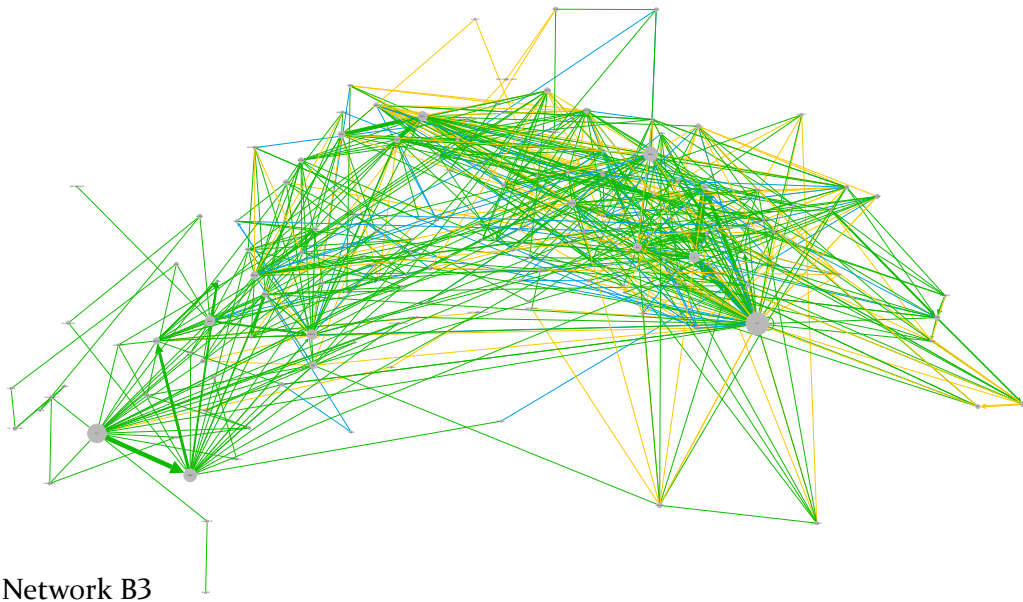
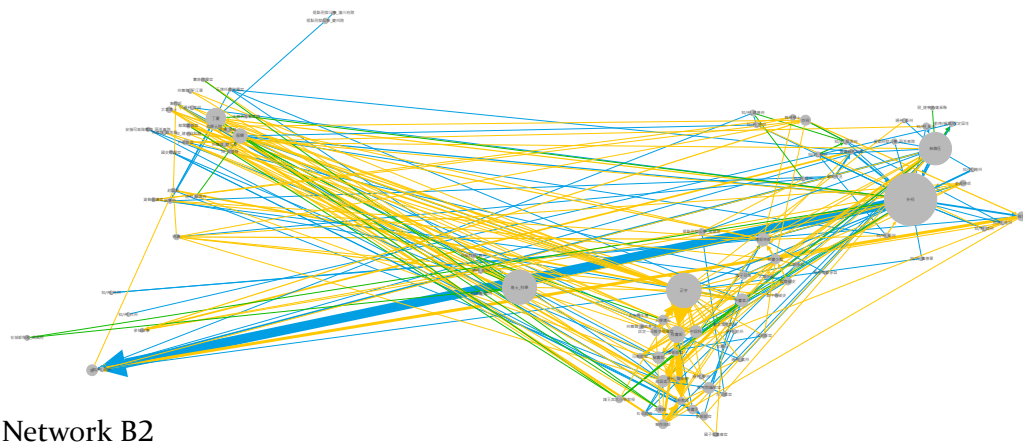


Nodes	Current-Flow Betweenness Centrality Scores (%) in Each Network			Num. of Centrality Measures by Which It Ranks High		
	B1	B2	B3	B1	B2	B3
外祠	4.44	8.59	4.94	7	7	7
正字	4.69	5.55	4.10	7	7	7
無職任	3.97	5.20	3.01	7	7	7
校書郎	2.57	2.40	2.74	7	7	7
中書舍人	0.90	1.77	1.90	7	7	7
秘書少監	1.88	1.27	1.51	7	7	7
起居舍人	—	1.59	1.21	6	7	7
京祠	1.09	1.36	1.61	4	7	7
著作佐郎	1.26	1.17	1.26	7	6	4
吏部郎官	1.24	—	2.06	7	2	7
禮部侍郎	—	1.81	0.98	0	7	7
禮部郎官	1.40	—	—	7	6	0
服闋	1.83	2.06	1.39	4	5	3
監察御史	1.29	—	1.13	7	3	2
秘書郎	1.85	1.59	—	7	5	0
吏部侍郎	—	—	1.36	1	3	7
致仕	—	1.17	1.06	0	6	5
丁憂	2.73	2.97	—	5	5	1
秘書丞	1.58	—	2.03	7	3	1
樞密院編修官	1.80	1.50	—	5	6	0
逝	—	1.49	2.07	0	5	5
進士_科舉	6.16	5.49	—	5	5	0
給事中	—	—	—	3	1	5
起居郎	—	—	1.38	3	0	5
侍講	—	—	—	3	2	3
詳定一司敕令所屬官	1.35	1.12	—	3	5	0
翰林學士	—	—	—	0	2	5
參知政事	—	—	—	0	0	5
次相	—	—	0.97	0	0	5
國子司業	—	—	—	0	5	0

Nodes	Current-Flow Betweenness Centrality Scores (%) in Each Network			Num. of Centrality Measures by Which It Ranks High		
	B1	B2	B3	B1	B2	B3
中詞科	—	1.66	—	0	5	0
進士_學校	2.93	—	—	5	0	0
宗正少卿	—	—	1.10	0	0	4
直學士院	—	—	—	0	2	2
王府教授	—	—	—	0	3	1
進士_特賜	1.45	—	—	4	0	0
知/判_平江府	—	—	—	0	0	3
工部侍郎	—	—	—	0	0	3
侍讀	—	—	—	0	1	2
太學博士	—	1.26	—	0	3	0
國史院屬官	—	—	—	0	0	2
諸王宮大小學教授	—	—	—	0	2	0
史館屬官	—	—	—	2	0	0
資善堂講官	—	—	—	2	0	0
祠部郎官	—	—	—	2	0	0
樞密使	—	—	—	0	0	1
國史院高階屬官	—	—	—	0	0	1
首相	—	—	—	0	0	1
右司郎官	—	—	—	0	1	0
知/判_建康府	—	—	—	0	1	0
太學錄	—	—	—	0	1	0
司勳郎官	—	—	—	0	1	0
太常少卿	—	—	—	0	1	0
勒停/編管/限定居住	—	—	—	0	1	0
簽書樞密院事	—	—	—	0	1	0
恩蔭入仕	—	—	—	1	0	0
都督府高階幕職	—	—	—	1	0	0

**Tab. 5** Top-ranking nodes in networks B1, B2, and B3. Nodes are arranged based on the number of centrality measures by which they rank among the top twenty. The “—” sign indicates that the node is not among the top twenty when ranked by current-flow betweenness centrality measure.





**Fig. 4** Bureaucratic transfers in phases one, two, and three. The median current-flow betweenness centrality is 0.16%, 0.23%, and 0.25% for Networks B1, B2, and B3, respectively. For visual clarity, nodes with a current-flow betweenness centrality lower than the network-wide median are filtered out.

These similarities aside, the structural differences between Networks B1 and B2 also deserve attention. They demonstrate how the talent-nurturing policy was particularly obstructed in Phase Two when Qin Gui dominated court politics. Network B1 features a small cluster of high-ranking court offices, including the Secretariat Drafter and the Vice Minister of several ministries, and some middle-ranking offices in this network, such as the directors and vice-directors of the Twenty-seven Bureaus, also have relatively high values of current-flow betweenness centrality. This suggests that in Phase One, some of the officials who had served in the Imperial Library nevertheless succeeded in climbing up the bureaucratic ladder, even though others encountered career obstacles. Considering that Qin Gui had become the Chief Councilor in 1138, the structural features of Network B1 indicate that Qin's administration continued the talent-nurturing policy in Phase One. However, only some of the offices identified in section 2.1 as characteristic of the mainstream career path of Imperial Library appointees appear in B1. Part of the reason is supplied by Li Xinchuan 李心傳 (1166–1243) and other contemporary observers, who noted that policy debates on war and peace with the Jurchens led many officials to be expelled from the court between 1131 and 1140.<sup>56</sup> Besides, it should also be noted that a decade, from 1131 to 1140, was too short for promising officials in the Imperial Library to reach the peak of their careers and attain high-ranking court positions. Thus, even if an Imperial Library appointee encountered no major setback in his career, his promotions to high offices are not reflected in Network B1 if they came after 1141. Therefore, Network B1 does not provide a full view of the career paths taken by the beneficiaries of the talent-nurturing policy but reflects only the early stage of their careers.

In stark contrast to B1, Network B2 does not contain a cluster of high-ranking court offices at all. The twenty-five years from 1131 to 1155 should have provided sufficient time for a considerable number of Imperial Library appointees – especially those who had entered the Library in Phase One – to advance to the rank of a minister or higher. However, promotions to such offices are not visible in B2. Instead, Network B2 shows that only a few officials appointed to the Imperial Library in Phase One (whose careers are represented with blue edges), and even fewer of those appointed to the Library in Phase Two (yellow edges), successfully made it to high-ranking positions. This reveals that many Imperial Library appointees hit the ceiling in their careers during Qin Gui's administration. Of the high-ranking court offices shown in B2, the Vice Minister of Rites and the Secretariat Drafter are the two nodes with a relatively high value of current-flow betweenness centrality. Yet, no strong ties exist in B2 that link these two nodes to top leadership offices. Conversely, outgoing links from these two nodes

---

56 For some examples, see *Yaolu* 124.681–2 (1138/12/癸酉); 127.723–2 (1139/3/己丑); 134.801–2 (1140/10/辛卯); 137.840–2 (1140/8/壬申). See also Zhu Xi 朱熹, “Fan zhige muji” 范直閣墓記, QSW 253: 5685.205.

often end up, within a few hops, at nodes that denote “custodian sinecure outside the capital,” “inactive,” or local government positions. This means that, in Phase Two, appointees to the Imperial Library often only reached a certain level of the bureaucratic ladder, and were thereafter forced out of the central government.

That the appointees to the Imperial Library met a collective career setback under the Qin Gui administration finds support in another comparison between B1 and B2. As discussed in section 2.1, the directors and vice-directors of the Twenty-seven Bureaus constituted a group of middle-ranking offices by way of which many Imperial Library appointees advanced to higher-ranking court positions. Of these directorships, 19 are present in Network B1. Of these 19 nodes, 16 are among the nodes with values of current-flow betweenness centrality above the median. By comparison, in Network B2, only 12 of these directorships are present, and only seven of them are among the nodes with values of current-flow betweenness centrality above the median (Table 6). This comparison signifies that more Imperial Library appointees in Phase One than in Phase Two successfully landed a job as a bureau director in the Six Ministries of the central government and developed their careers thenceforth. It also echoes an observation from Hong Mai 洪邁 (1123–1202), who criticized Qin Gui for deliberately expelling officials from the court during his administration.<sup>57</sup>

Intriguingly, the structure of Network B2 shows the presence of a major career impediment for all appointees to the Imperial Library, regardless of their relationship with Qin Gui. This observation implies that the Qin Gui administration not only prevented its critics from attaining influential positions in government, but may have also viewed its political allies as potential rivals and forced them out of the court once they reached certain high-level positions. What differs is, the administration’s allies and officials who avoided challenging it openly may have been able to leave the court more decently than its critics.<sup>58</sup> The result was that only very few officials had the chance to rise to the top leadership under the Qin Gui administration. Most of these men acted as Qin’s lackeys, not as a check on his power.<sup>59</sup>

---

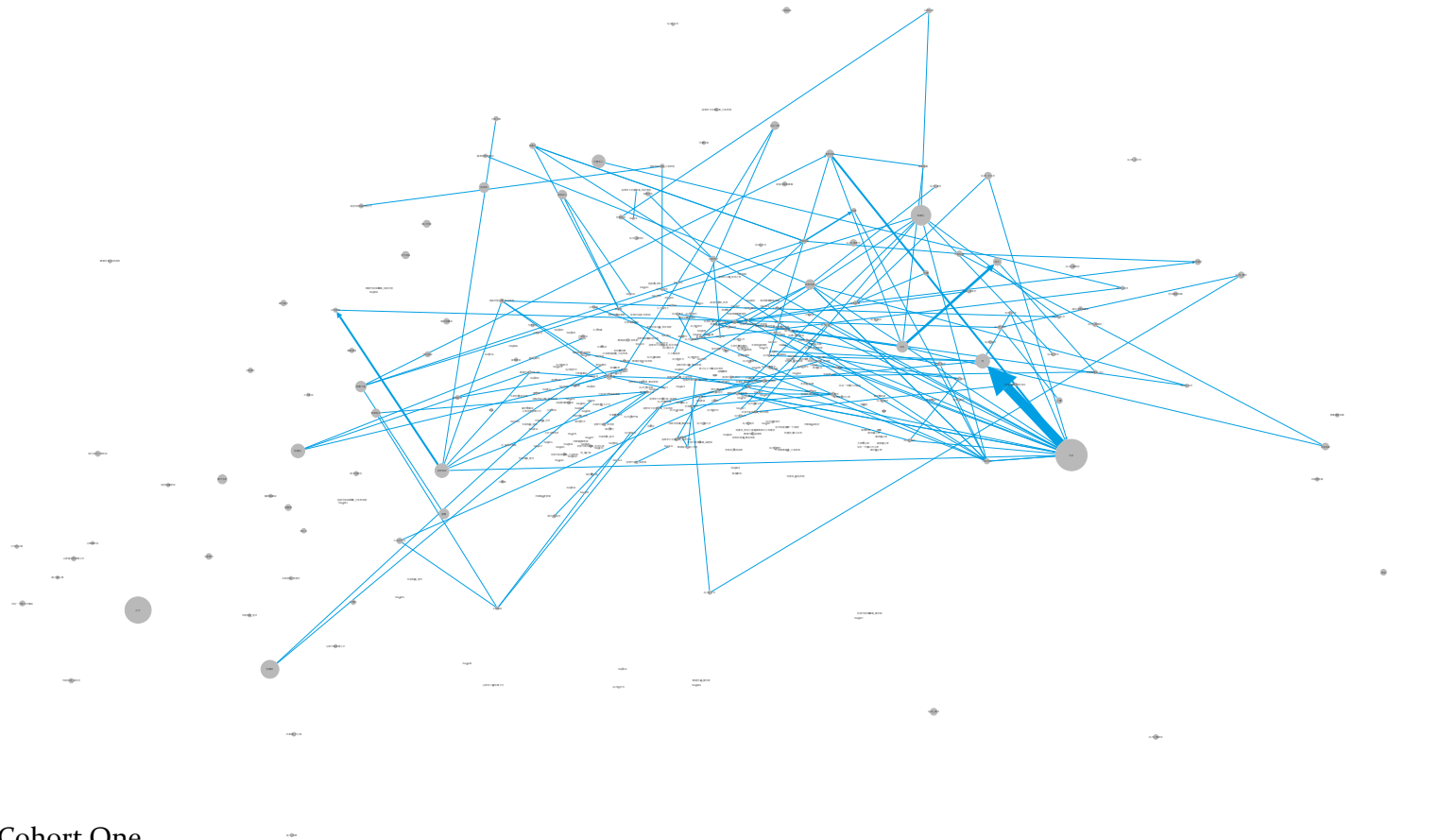
57 Hong Mai, “Langguan yuanshu” 郎官員數, in *Rongzhai sanbi* 容齋三筆 (Beijing: Zhonghua shuju, 2005), 5.484.

58 Some examples include Wang Ciweng 王次翁 (1079–1149) and Song Zhicai 宋之才 (1090–1166). Tuotuo 脫脫 et al., *Song shi* 宋史 (Taipei: Academia Sinica Scripta Sinica Database, 1984), 380.11709. Xue Jixuan 薛季軒, “Song shilang zhicai xingzhuang” 宋侍郎之才行狀, *QSW* 258: 5796.50–51.

59 Teraji Jun, *Nan Song chuqi zhengzhishi yanjiu*, 299–316, 319–327.

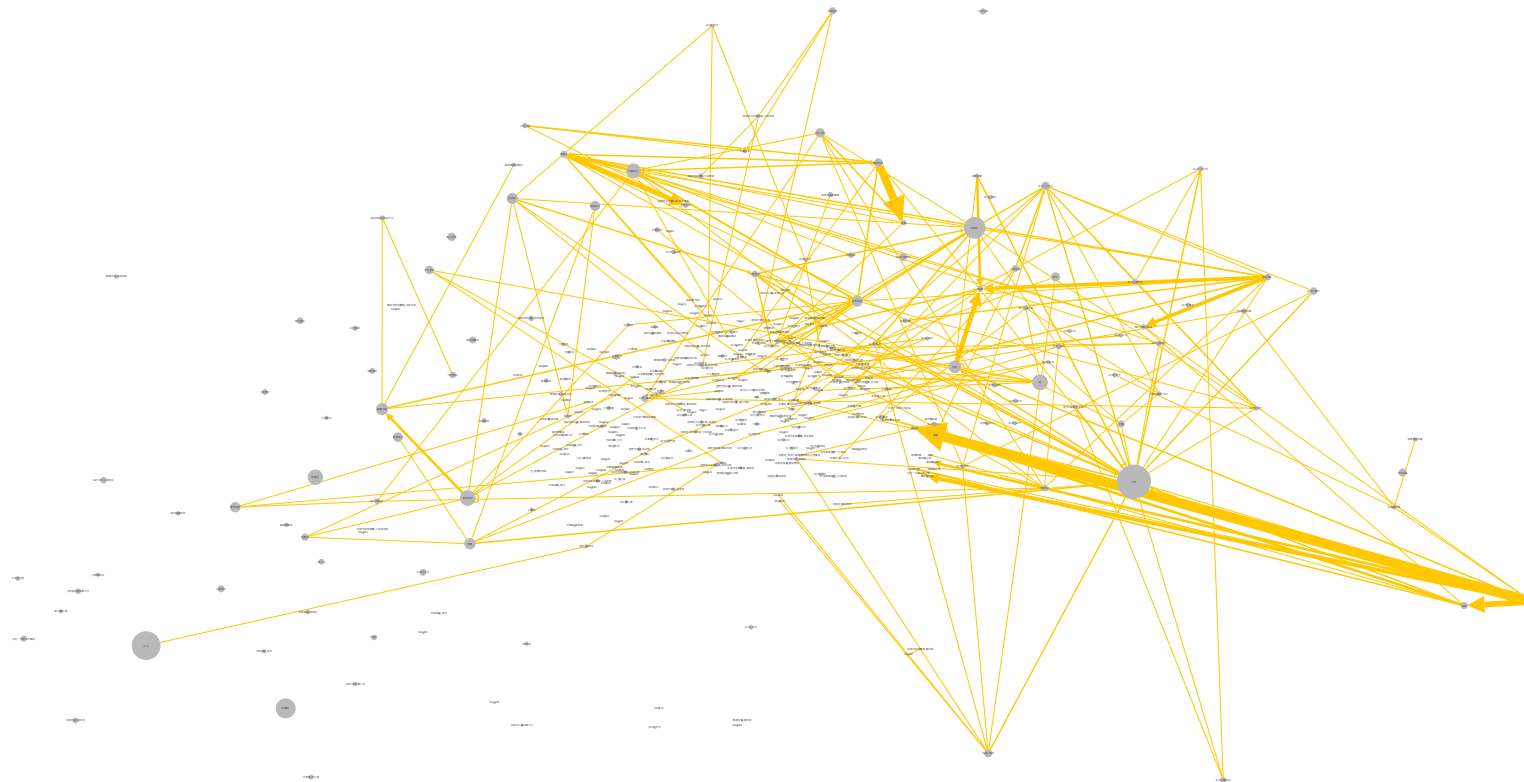
Network B1			Network B2		
Post	Current-Flow Betweenness Centrality (Median = 0.16)		Post	Current-Flow Betweenness Centrality (Median = 0.23)	
	Value	Ranking in B1		Value	Ranking in B2
禮部郎官	1.40	14	吏部郎官	1.04	22
吏部郎官	1.24	18	禮部郎官	0.78	29
左司郎官	0.74	25	戶部郎官	0.55	40
都官郎官	0.66	32	右司郎官	0.52	45
考功郎官	0.65	33	司勳郎官	0.48	52
工部郎官	0.64	35	左司郎官	0.47	53
司勳郎官	0.61	38	司封郎官	0.47	55
祠部郎官	0.60	39	考功郎官	0.20	108
右司郎官	0.54	43	駕部郎官	0.12	144
司封郎官	0.47	49	兵部郎官	0.12	156
駕部郎官	0.38	59	祠部郎官	0.11	181
兵部郎官	0.29	76	屯田郎官	0	195
屯田郎官	0.26	85			
度支郎官	0.26	89			
刑部郎官	0.25	93			
主客郎官	0.21	109			
戶部郎官	0.13	170			
金部郎官	0.07	237			
倉部郎官	0	273			

**Tab. 6** Directorships/vice-directorships of the twenty-seven bureaus in networks B1 and B2. Posts with a current-flow betweenness centrality higher than the network-wide median are highlighted in blue.

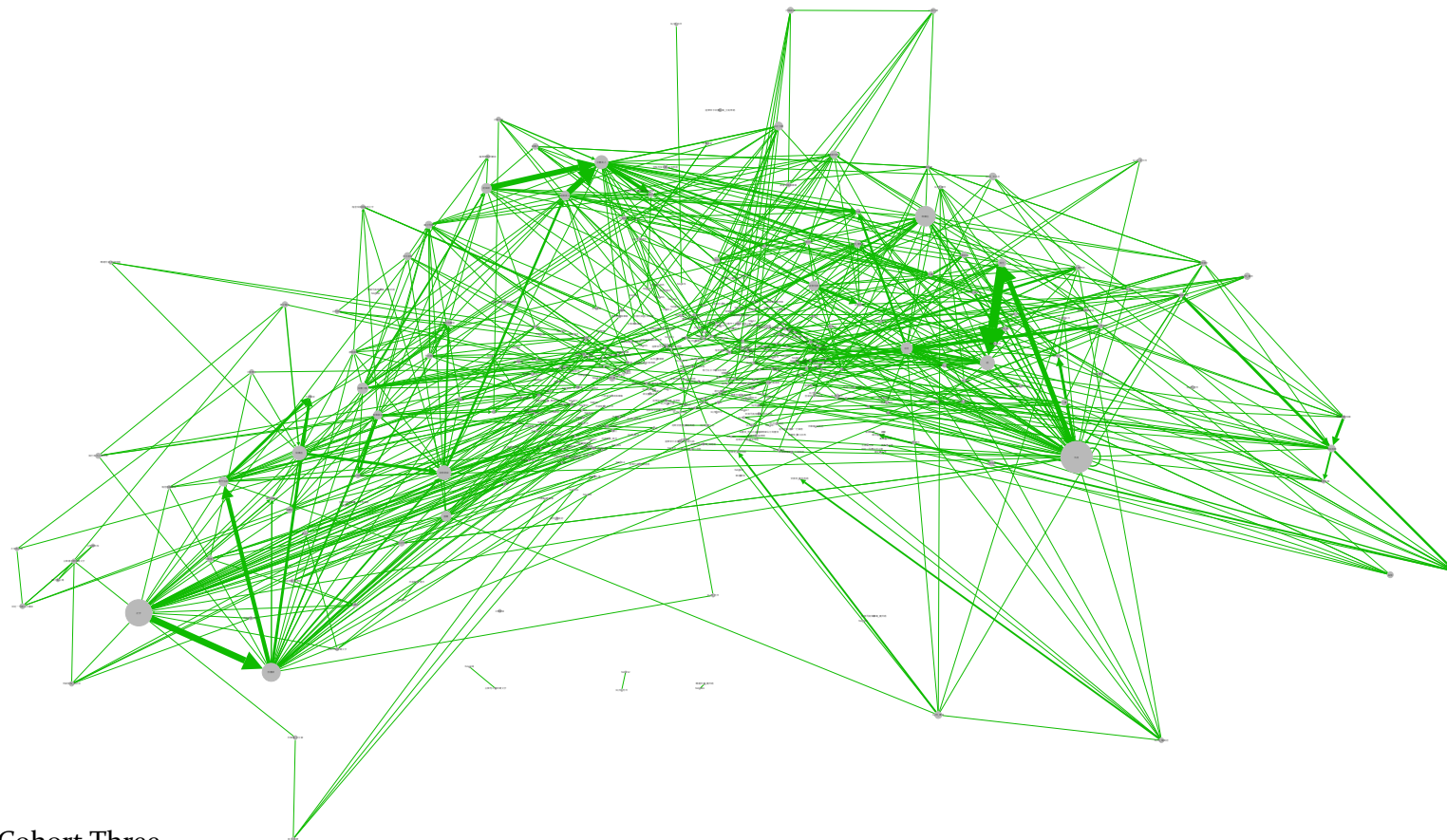


Cohort One





Cohort Two



**Fig. 5** Comparison of bureaucratic transfers in phase three between the three cohorts. Edges are colored based on cohort attribute (blue, yellow, and green for Cohorts One, Two, and Three respectively).

These structural features of Network B2 must be considered along with two other observations. First, as Figure 3 has illustrated, the number of appointments to the Imperial Library declined significantly, sometimes down to zero, between 1144 and 1155. Second, the position of the Imperial Diarist in the Chancellery was left vacant most of the time after 1143 and when not vacant, was usually filled only with officials on probation.<sup>60</sup> Likewise, the Secretariat Drafter and the Supervising Secretary were also left vacant after 1147 and 1150, respectively.<sup>61</sup> All of these suggest that the Qin Gui administration deliberately obstructed any attempt to groom promising candidates for high-ranking offices, as intended by the talent-nurturing policy discussed in section 2, to ensure Gaozong's and Qin Gui's monopoly of power and their exclusive control over personnel decisions.

The structure of Network B3 shows that career obstacles encountered by the Imperial Library appointees in Phase Two were largely removed in Phase Three, when the Qin Gui administration came to an end and the Song court returned to its tradition of collective deliberation.<sup>62</sup> Compared to Networks B1 and B2, Network B3 contains more nodes that represent the top leadership positions in both civilian and military branches of the government. Of these nodes, the Assistant Chief Councilor and the Vice Chief Councilor rank among the highest, according to at least five centrality measures (Table 5). This demonstrates that more officials in Phase Three than in the two earlier phases successfully made it to top leadership positions.

In spite of the differences discussed above, officials who received their first appointments to the Imperial Library in any of three phases seem to have followed, more or less, a similar career path in Phase Three, unless their careers ended prematurely due to death or other reasons. In Figure 5, I partition Network B3 into three subgraphs based on the cohort attribute of the edges. Under the talent-nurturing policy, an appointment to the Imperial Library supposedly placed an official on a fast track to high-ranking offices. However, the graphs in Figure 5 show that as late as the third phase (1155/10–1164), there were still officials from the first two cohorts who were just beginning to receive a promotion to mid-level offices (e.g., directors and vice-directors of the Twenty-seven Bureaus); some were even reappointed back to or received internal transfers inside the Imperial Library after being ousted from the court, and continued to linger there. This also implies that after receiving their first appointments to the Imperial Library be-

---

60 Hong Mai, "Langguan yuanshu," in *Rongzhai sanbi*, 5.484.

61 *Yaolu* 156.197–2 (1147/12/丙辰); 172.420–2 (1156/3/己巳); 176.479–2 (1157/2/丁未).

62 The career impediments were removed upon Qin Gui's death in 1155/10, owing much to officials who made an effort and managed to regain control over decisions concerning the Imperial Library appointments. A full discussion of this will be provided in my forthcoming doctoral thesis.

tween 1131 and 1155, most of these officials saw their careers interrupted or stagnant for many years until the death of Qin Gui.<sup>63</sup>

In brief, my analysis in this section demonstrates that although Gaozong and his court revived the policy of talent-nurturing, its implementation was not smooth but depended heavily on changing political conditions. The analysis also shows the positive correlation between the format of government operation and the career development of the beneficiaries of the talent-nurturing policy. The policy functioned as expected to facilitate its beneficiaries' careers when the tradition of collective deliberation prevailed at the Song court, but was disrupted when Qin Gui dominated court politics. The causal factors behind this phenomenon will be explored in future stages of this project.

#### 4. Conclusions and Prospects

This article reports the findings from an experimental project that uses network analysis to investigate the operations of the early Southern Song bureaucracy. As officials were transferred from one office to another, such transfers established connections between these offices. This creates a network of offices, linked by the observed instances of transfers, which provides invaluable information for deciphering early Southern Song bureaucratic practices and the officials' career paths. It lends itself naturally to analysis as a directed graph. Rich documentation of the officials' bureaucratic transfers in the extant historical records also provides abundant data for such analysis.

By studying the career paths of the 197 officials who served in the Imperial Library between 1131 and 1164, this article examines the talent-nurturing policy in the early Southern Song. Allegedly a dynastic tradition, the policy was intended to prepare talented officials for the top state leadership by first appointing them to the Imperial Library. These appointments, in theory, put these talented officials on a fast track to the top leadership. By analyzing their careers, the study has identified a series of positions held by these officials after they left the Imperial Library and were on their way to the upper echelons of officialdom. It reveals that many of the posts on this fast-track career path were associated with the responsibility of evaluating the character and competence of other officials and had a direct or indirect bearing on personnel decisions at the Song court. This reflected the expectation in Song times that the top leaders of the government should pos-

---

63 Interruption means that an official was deprived of his rank or relieved of active duty. Stagnation means that an official did not receive promotions as quickly as would normally be expected. The graph for Cohort Two inevitably includes a few cases where the officials were appointed to the Imperial Library in the last years of the Qin Gui administration and therefore naturally started off from the middle-level offices in Phase Three.

sess the capability of “recognizing talents,” and in practice service in the Imperial Library – and later in other positions on the fast-track career path – gave these would-be leaders a marked influence on personnel decisions.

This study also compares the officials’ careers in three different periods between 1131 and 1164. The results reveal that the revival of the talent-nurturing policy in the Southern Song went through a tortuous process. With Gaozong’s support, the court steered away from it for more than a decade under the Qin Gui administration. To silence critics of its appeasement policy with the Jurchens, the Qin Gui administration created obstacles that prevented many appointees to the Imperial Library from advancing to high-ranking positions. Instead, these officials sank into a prolonged period of career stagnation that lasted until Qin Gui’s death.

This article demonstrates that network analysis, combined with a prosopographical approach, holds great promise for understanding the operations of the Song bureaucracy. Nonetheless, several questions remain unanswered and must await future research. For example, this study reports how the talent-nurturing policy was implemented in the early Southern Song, but it remains unclear whether there were notable differences between the policy in the Southern Song and its alleged precedent in the Northern Song. Meanwhile, the distinctive features of the Imperial Library appointees’ career paths under the talent-nurturing policy will only stand out in comparison to the careers of other officials who never gained experience in the Library. These questions suggest that network analysis of the Song bureaucratic practices will benefit immensely from synchronic and diachronic comparisons.

The present study also demonstrates the methodological potential of using network analysis to explore the politics and institutions in other periods of imperial Chinese history, given that bureaucracy in these periods shared many important features with that in the Song. However, this methodology also faces challenges. Network theory and relevant mathematical concepts are developed in natural and social sciences, and statistical tools in this field are devised for analyzing specific types of network activity, which have structural commonalities with the personnel flow in a bureaucracy, but nevertheless each also has its own distinctive features.<sup>64</sup> When network concepts and statistical tools are applied in this study, I seek to open a dialogue between network theory and history by mapping the similarities between them. I have done so by translating observed patterns of personnel transfers in a historical bureaucracy into network language and then deriving historical meanings from the results of network analysis. This study demonstrates that such dialogue between network theory and historical studies is productive. However, it requires further research to understand to what

---

64 Stephen P. Borgatti, “Centrality and Network Flow”, *Social Networks* 27 (2005), 55–71.

extent the network approach can help scholars capture more nuanced historical phenomena and whether it is necessary to develop more innovative analytical concepts and tools that are tailored to specific topics in history, such as the personnel flow in historical organizations.

## 5. Appendices

### Appendix 1: Sources for the Dataset

The data used in this study is collected in two steps. The first step is to identify all the officials who were appointed to the Imperial Library between 1131 and 1164. The second step is to reconstruct, as much as possible, the entire career of each of these officials based on all the appointments he had received.

The key sources used in the first step include two dated rosters of Imperial Library officials. One roster is preserved in the *Records of the Imperial Library in the Southern Song* (*Nan Song guange lu* 南宋館閣錄, hereafter NSGGL) by Chen Kui 陳騭 (1128–1203), and the other is retrieved from the *Comprehensive Survey of Song Dynasty Capital- and Court-Rank Officials* (*Songdai jingchaoguan tongkao* 宋代京朝官通考) compiled by modern historian Li Zhiliang.<sup>65</sup> I began with the first roster and used it to create a preliminary list of Song individuals who were recruited into the Imperial Library between 1131 and 1164. I then used the second roster for supplemental information. A total of 197 Imperial Library appointees are collected from these rosters. For each person on this name list, I built a profile using the biographical information provided in the NSGGL and the Database of Civil Service Examination Graduates in All Dynasties (*Lidai jinshi dengke shujuku* 歷代進士登科數據庫).<sup>66</sup>

In the next step, I collected the data on bureaucratic transfers for each of these 197 officials. The types of data I harvested include: (1) the appointee's name; (2) the year when each appointment was announced or when the appointee assumed office, depending on the information available in sources; (3) the position the appointee left; (4) the new position he took up; (5) concurrent appointments, if any, given in the personnel order; and (6) reasons for the transfer, if mentioned.

---

65 Chen Kui 陳騭, *Nan Song guange lu* 南宋館閣錄 (Beijing: Zhonghua shuju, 1998). Li Zhiliang 李之亮, "Bishu sheng" 秘書省, *Songdai jingchao guan tongkao* 宋代京朝官通考 (Chengdu: BaShu shushe, 2003), 4: 328–588.

66 The profile for each official includes his style name, birth and death years (if known), native or resident prefecture, year and method of entry into government (*chushen* 出身), whether he was *jinshi* and/or held another academic degree (and if so, the year he received it, the subject in which he specialized, and his ranking), year of his first appointment to the Imperial Library and the office he held in the Library, and whether he has a biography or an epitaph.

When relevant data is reported in multiple sources, I prioritize what is documented in the official's biographical sources, including his account of conduct (*xingzhuang* 行狀), tomb epitaph (*muzhi ming* 墓誌銘), Spirit Path Inscription (*shendao bei* 神道碑), or his short biography (*zhuan* 傳) in the *Complete Prose of the Song* (*Quan Song wen* 全宋文, hereafter QSW) and the Database of Song Dynasty Epitaphs (*Songdai muzhiming shujuku* 宋代墓誌銘數據庫, hereafter MZK). Information collected from these biographical sources is checked against personnel orders recorded in the *Chronological Record of Important Events since the Jianyan Era* (*Jianyan yilai xinian yaolu* 建炎以來繫年要錄, hereafter Yaolu) and the *Recovered Draft of the Collected Essential Documents of the Song Dynasty* (*Song huiyao jigao* 宋會要輯稿, hereafter SHY). For those officials who do not have a biography in the extant historical record, their career data is gleaned, to the extent possible, from Yaolu and SHY. I also made use of an official's biography in the dynastic history of the Song (*Song shi* 宋史, hereafter SS) and its supplement (*Song shi yi* 宋史翼, hereafter SSY) but only when relevant information is absent in the other sources.<sup>67</sup> The resulting dataset is further enriched with postings data from the China Biographical Database (CBDB) 中國歷代人物傳記資料庫 and with data from rosters of officials in the digitized local gazetteers of the Song and Yuan dynasties found in the Scripta Sinica Database 漢籍電子文獻資料庫. This aggregate dataset was checked for internal consistency and duplicate data was discarded.

## Appendix 2. Definition of Nodes

To convert historical records of bureaucratic transfers into network data involves judgment and interpretation. During this process, I have inevitably drawn on my knowledge of the Song institutional history, but I have tried, whenever possible, to keep this knowledge at bay so that it will not predetermine the structure of the network. Therefore, for example, I take a neutral attitude towards each bureaucratic transfer, without judging whether it was a demotion or promotion. Nevertheless, I made several important decisions in building the dataset and a full disclosure is in order.

First, to examine the impact of the methods of entry (*chushen* 出身) on an official's career under the talent-nurturing policy of the early Southern Song, I have coded each method of entry as a node in the dataset, as if it were an office in the bureaucracy. In Song times, men could enter government service by

---

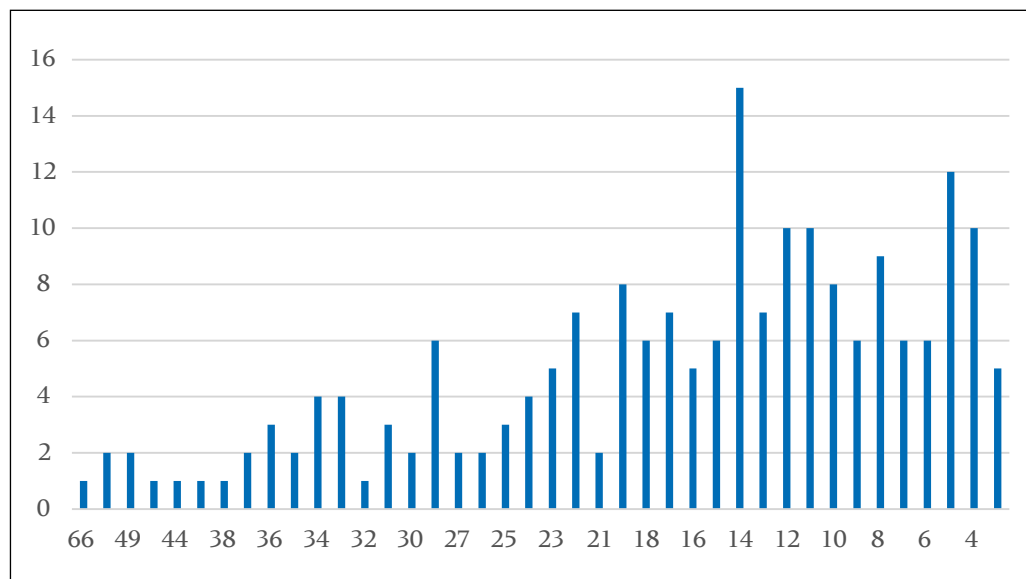
67 Biographies in the dynastic history of the Song are heavily edited and abridged. Therefore, they are given low priority in this study. Compiled by Lu Xinyuan 陸心源 (1834–1893), *Song shi yi* collects biographies of Song people that are not included in the dynastic history. The biographies in *Song shi yi* provide fewer details than what one typically finds in a person's account of conduct or epitaph, but modern scholars agree that these biographies are more reliable than those in the dynastic history. See Wu Boxiong 吳伯雄, "Qianyan" 前言, in *Song shi yi* 宋史翼, ed. Lu Xinyuan (Huzhou: Zhejiang guji chubanshe, 2017), 1–24.

Groups	Num. of Officials	% of Total	Num. of Officials in the group with fewer than nine* appointment records
Having biographical materials in both QSW/MZK and SS/SSY	32	16.2%	4
Having biographical materials in QSW/MZK but not in SS/SSY	10	5.1%	0
Having biographical materials in SS/SSY but not in QSW/MZK	46	23.4%	0
Nine or more appointment records in <i>Yaolu</i> and/or <i>SHY</i> , but no biographical materials in both QSW/MZK and SS/SSY	45	22.8%	0
Fewer than nine appointment records in <i>Yaolu</i> and/or <i>SHY</i> , and no biographical materials in both QSW/MZK and SS/SSY	64	32.5%	44
Total	197	100%	48

\* The criterion of “nine appointment records” comes from the interquartile range (IQR) of the distribution of the 197 officials’ appointment record amounts, which falls between 9 and 23, by applying the box-and-whisker plot for observation. A box-and-whisker plot is a standardized way of displaying the distribution of data based on a five-number summary (“minimum”, first quartile (Q1), median, third quartile (Q3), and “maximum”). The first quartile (Q1/25th Percentile) means the middle number between the smallest number (not the “minimum”) and the median of the dataset while the third quartile (Q3/75th Percentile): the middle value between the median and the highest value (not the “maximum”) of the dataset. Thus, the term “interquartile range (IQR)” refers to the 25th to the 75th percentile, which is approximately the same as the middle 50% of a nearly normal distribution. The interquartile range can be used to an indicator of the majority situation, compared to individual instances. This definition is cited from Michael Galarnyk, “Understanding Boxplots,” posted on the website of “Toward Data Science,” <https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51>, retrieved on 17 April 2020.

**Tab. 7** Sources of biographical data for the present study. Officials in this table are grouped by the sources that contained their biographical data.





**Fig. 6** Number of officials by the number of surviving appointment records. The horizontal axis indicates the number of extant appointment records, and the vertical axis denotes the number of officials with a given number of surviving appointment records.

different methods, among which the most important were obtaining a “presented scholar” (*jinshi*) degree from the civil service examinations (*keju* 科舉), passing the erudite literatus examination (*cike* 詞科),<sup>68</sup> graduating from government schools (*guanxue* 官學), and using the protection privilege (*enyin* 恩蔭), which allowed officials above a certain rank to sponsor their relatives for appointment to low-ranking offices. Those who took the first three routes were considered men with “formal qualifications” (*you chushen* 有出身) for government service,<sup>69</sup> which were conventionally regarded as a prerequisite for an Imperial Library post. In rare cases, when prominent scholars without such qualifications

68 Guan Qin 管琴, *Cike yu Nan Song wenxue* 詞科與南宋文學 (Beijing: Peking University Press, 2018), 104–115. Although men who passed the erudite literatus examination were also granted a *jinshi* degree, this study codes graduation from the erudite literatus examination separately from the attainment of a *jinshi* degree in the regular civil service examination.

69 The historical records distinguish between a variety of degrees awarded in the civil service examinations and government schools, including “metropolitan graduates with honors” (*jinshi jidi* 進士及第), “regular metropolitan graduates” (*jinshi chushen* 進士出身), “associate metropolitan graduates” (*tong jinshi chushen* 同進士出身), “graduates with honors from the Upper Hall” (*shangshe jidi* 上舍及第), “regular graduates from the Upper Hall” (*shangshe chushen* 上舍出身), “associate graduates from the Upper Hall” (*tong shangshe chushen* 同上舍出身). My dataset ignores these fine distinctions.

(*wu chushen* 無出身) were considered for appointment to the Imperial Library, they were often conferred a *jinshi* degree, by imperial grace, prior to the appointment. In the dataset, I labeled five kinds of methods of entry observed in the 197 officials' biographies, include obtaining a *jinshi* degree in regular examinations (labeled “進士\_常科”), obtaining a *jinshi* degree by imperial grace (labeled “進士\_特賜”), graduation from erudite literatus examination (labeled “中詞科”), graduation from government schools (labeled “官學”), and protection privilege (labeled “恩蔭入仕”).

Second, the career of an official could be disrupted or ended in two scenarios. In one scenario, an official could leave officialdom, temporarily or permanently, due to mourning obligations (*dingyou* 丁憂), punishment, retirement, or death.<sup>70</sup> In the other scenario, an official remained in the officialdom but without a substantive commission. Some of them were given a custodian sinecure at a Buddhist or Daoist temple (*ciguan* 祠官).<sup>71</sup> Others were not even given custodian sinecures, and they are labeled in my dataset simply as 無職任 (*wu zhiren*, lit. “inactive”).<sup>72</sup> These scenarios could occur at any stage of a Song official's bureaucratic life, disrupting his career or ending it altogether. Given that the first goal of this study is to reconstruct the mainstream career path for the Imperial Library appointees and identify the major posts they passed through toward top leadership, I have treated these scenarios as “noise nodes” and left them out of my analysis in section 3.<sup>73</sup> In section 4, however, these nodes are included in the analysis be-

70 Temporary leaves caused by punishment included suspension of official status (*leting* 勒停), imposition of residential restriction (*juzhu* 居住), surveillance (*bian'guan* 編管), and so on. Permanent leaves were those due to retirement (*zhishi* 致仕), expungement (*chuming* 除名) from the officialdom as a severe punishment, or death. Although these “permanent” leaves usually marked the end of an official's career, it was possible for those in retirement or those expunged to return to the officialdom on the emperor's order.

71 The temples were scattered across the country. Categorically, those located in the capital were called *jingci* 京祠 while others in the provinces *waici* 外祠. However, since these officials were not required to be physically present at these temples, the actual locations of these temples do not matter. Therefore, I only make a distinction between capital sinecures from provincial ones.

72 Song officials might become “inactive” by choice or on government orders due to administrative remissness, criticisms from remonstrance organs, and political strife. The difference between a custodian sinecure and an inactive status was that holders of the sinecures had a salary while inactive officials did not. Historical sources use a variety of terms for officials who became inactive, such as *ba* 罷, *chu* 黜, *quguan* 去官, and so forth, but my dataset does not distinguish between these terms. It should be noted that this study ignores the period when an official spent awaiting his next appointment (*daique* 待闕), because this was a common phenomenon in the Song officialdom. As the number of officials increased but the number of vacancies in government did not increase in tandem, the waiting period was sometimes as long as several years. This study does not consider an official-in-waiting as “inactive.”

73 In total, 37 unique nodes and 244 unique edges are filtered out.

cause the goal of this section is to understand how court politics disrupted the careers of the Imperial Library appointees.

Third, in a few cases, offices with different titles are lumped together and coded as a single node, because they had similar ranks and duties and were often mentioned without distinction in Song sources. For example, the director (*langzhong* 郎中) and vice director (*yuanwailang* 員外郎) in each of the twenty-seven bureaus of the Department of State Affairs are coded simply as *langguan* 郎官 of that bureau. Likewise, functionary assistants (*muzhi* 幕職) in each circuit administrations<sup>74</sup> are grouped into two categories – the senior (*gaojie muzhi* 高階幕職) and the ordinary (*shuguan* 屬官) – and each category is treated as a single node in the network graph, regardless of their specific duties and exact titles. This preserves the ambiguity in the historical record and also avoids unnecessary complexities in the dataset. This affects only about 5% of a total of 3367 edges and therefore has only limited influence on the structure of the network.

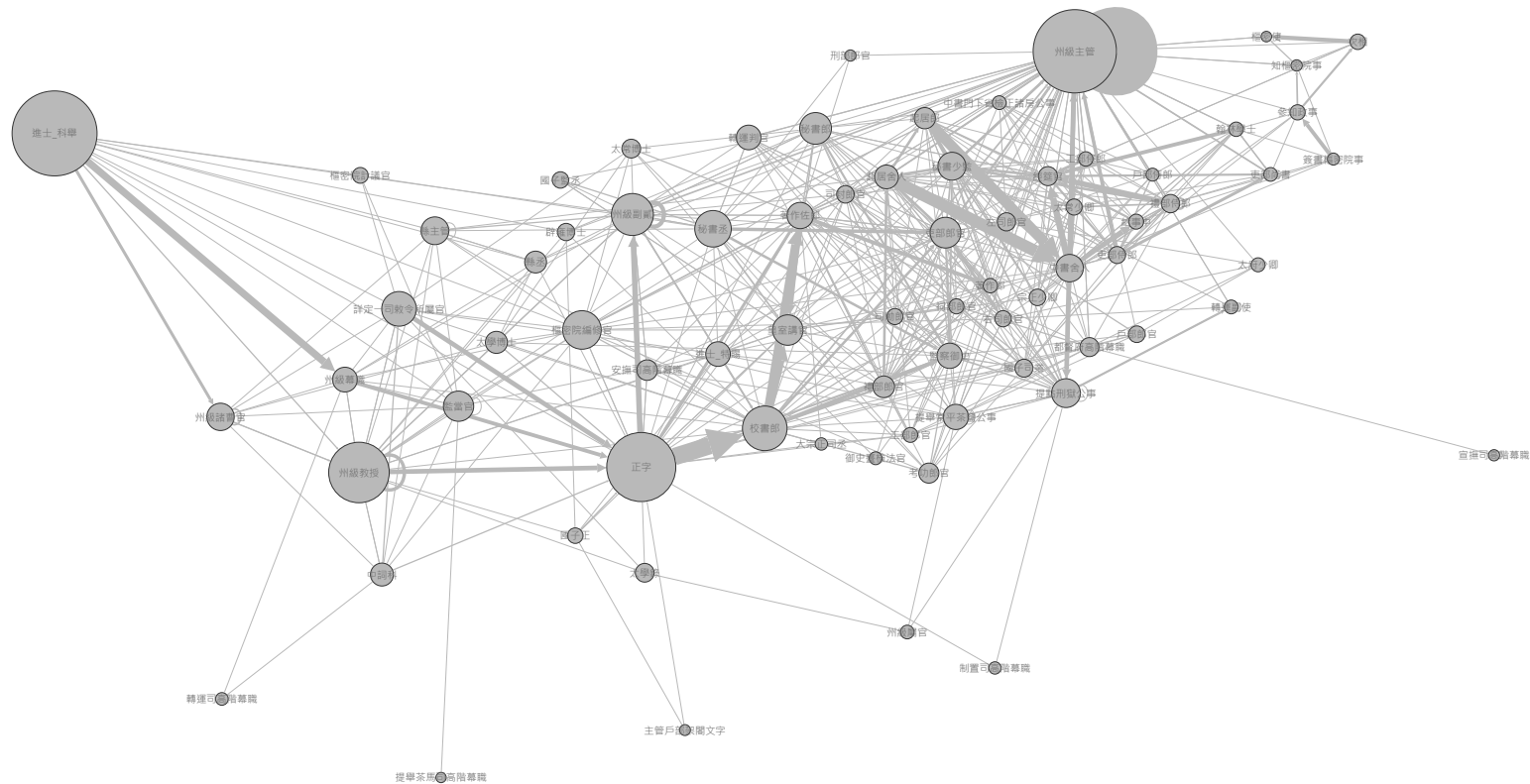
Fourth, I code offices with the same title but in different administrative divisions as different nodes. For instance, the prefect of Lin'an (*zhi Lin'an fu* 知臨安府) and that of Fuzhou (*zhi Fuzhou* 知福州) are treated as two different nodes in the dataset. Two considerations are behind this decision. To begin with, these offices, despite having the same title, were handled differently in the Song bureaucratic practices. The Song government divided its territory into three administrative levels (i.e., circuits, prefectures, and counties), and the prefectures were further classified into four types (*fu* 府, *zhou* 州, *jun* 軍, *jian* 監) and several different grades according to their administrative functions and importance. Counties were likewise classified into several grades according to the size of their population. The type and grade of a prefecture or county were an important consideration of the Song personnel authorities when they made appointment decisions. Officials who were held in high regard and had a good performance record, for example, were more likely to receive an assignment in the more highly ranked places.

Moreover, the Song bureaucratic system had a pyramidal structure, with more offices with the same titles (e.g., prefects) at the bottom and fewer such offices (e.g., ministers) at the top. In 1208, for instance, there were 295 prefects across the country but only six ministers in the central government. If one counts all offices with the same title as a single node, regardless of the places they administered, and calculates, say, their degree centrality in the network of bureaucratic transfers, the node denoting the prefect would almost inevitably have a much

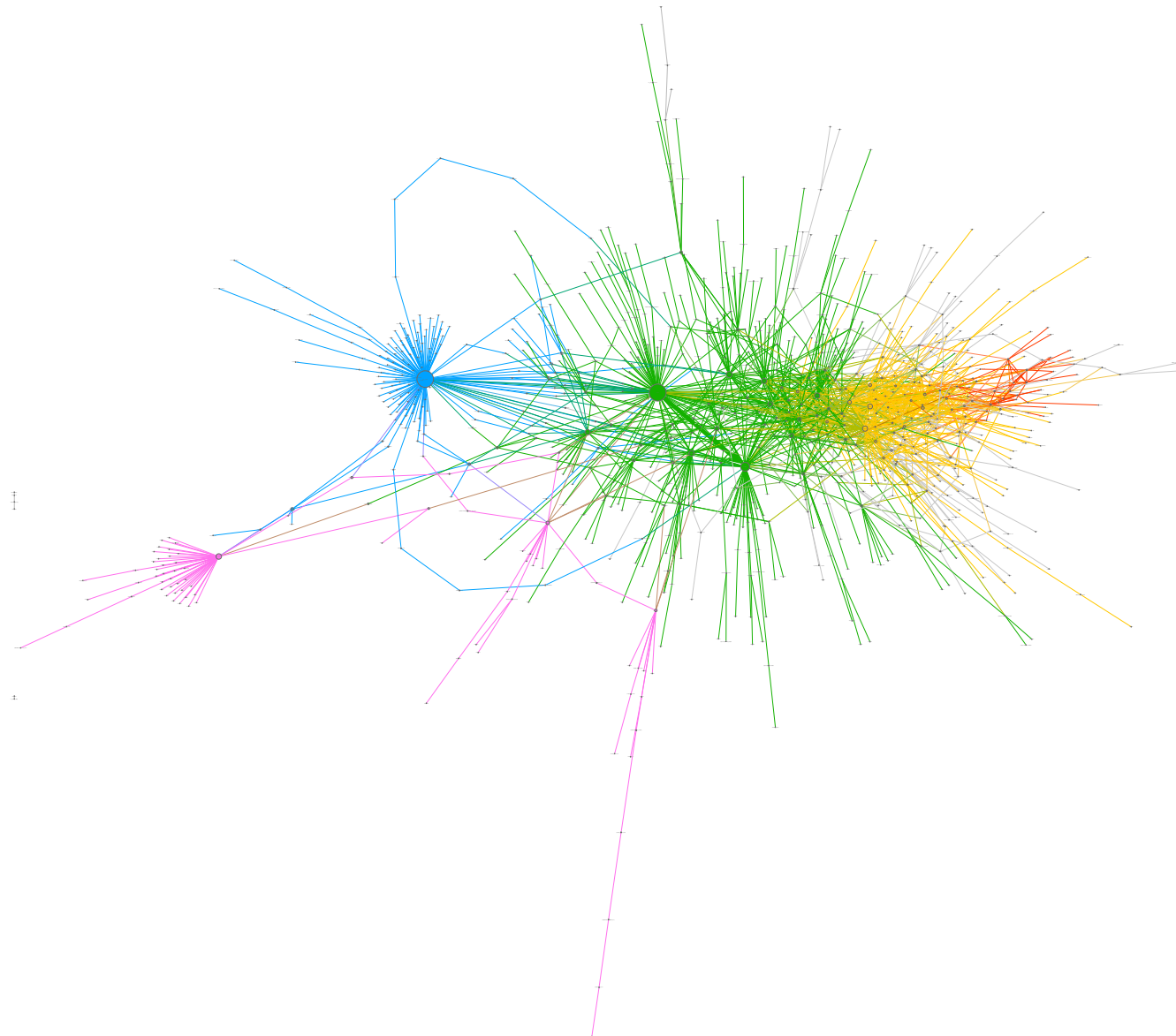
---

74 Each circuit in the Song was governed by four intendancies: fiscal, military judicial, ever-normal granaries, headed by a commissioner, a vice commissioner, and in some cases an assistant commissioner, who led a group of functionary assistants with diverse titles but overlapping duties.

higher centrality score than the node denoting a minister. A graph like this (Figure 7) may provide a useful illustration of institutional features of Song bureaucracy, because it reveals the hierarchy of offices (with lower-ranking offices on the left-hand side and higher-ranking ones on the right-hand side) and the sequence of offices that officials went through in their careers. However, this is not the topic under discussion in this article.



**Fig. 7** Network Z. Nodes with a current-flow betweenness centrality below the median are filtered out for visual clarity.

**Appendix 3: Complete Graph of Network A**

#### Appendix 4: Centrality Scores for the Thirty-Eight Top-Ranking Nodes

The “—” sign indicates that the node is not among the top twenty when ranked by the relevant centrality measure.

Nodes	In-degree	Out-degree	Degree	Weighted In-degree	Weighted Out-degree	Weighted Degree	Current-Flow Betweenness (%)	Num. of Centrality Measures by Which It Ranks High
正字	64	40	104	87	108	195	5.52	7
校書郎	28	37	65	71	103	174	2.83	7
秘書丞	23	18	41	26	38	64	1.83	7
吏部郎官	25	30	55	43	55	98	1.61	7
秘書郎	18	23	41	24	36	60	1.57	7
秘書少監	36	27	63	48	60	108	1.56	7
中書舍人	25	35	60	84	84	168	1.44	7
著作佐郎	14	27	41	46	57	103	1.40	7
監察御史	19	19	38	45	40	85	1.28	7
起居舍人	20	24	44	42	64	106	1.05	7
禮部郎官	15	20	35	30	29	59	1.04	7
起居郎	15	19	34	30	51	81	0.78	7
禮部侍郎	15	24	39	33	50	83	—	7
吏部侍郎	19	22	41	33	40	73	—	7
給事中	—	15	28	28	33	61	—	6
進士_常科	—	130	132	—	134	136	5.40	5
翰林學士	—	16	28	—	34	52	—	5
官學	—	41	41	—	41	—	1.67	4

Nodes	In-degree	Out-degree	Degree	Weighted In-degree	Weighted Out-degree	Weighted Degree	Current-Flow Betweenness (%)	Num. of Centrality Measures by Which It Ranks High
樞密院編修官	18	15	33	—	—	—	1.60	4
侍講	18	—	—	63	—	66	—	4
詳定一司敕令所屬官	18	—	27	—	—	—	1.51	3
次相	—	—	—	—	36	46	—	3
參知政事	—	—	—	26	—	46	—	3
直學士院	15	—	—	44	—	—	—	3
侍讀	14	—	—	33	—	—	—	3
史館屬官	14	—	—	—	—	—	—	2
都督府高階幕職	14	—	—	—	—	—	—	2
工部侍郎	—	—	—	—	25	—	—	2
王府教授	—	—	—	41	—	—	—	2
宗正少卿	—	15	—	—	—	—	—	2
進士_特賜	—	—	—	—	—	—	1.27	1
中詞科	—	—	—	—	—	—	1.00	1
太學博士	—	—	—	—	—	—	0.80	1
州諸曹_婺州	—	—	—	—	—	—	0.80	1
殿中侍御史	—	—	—	—	—	—	—	1
太常少卿	—	—	—	—	—	—	—	1
知/判_饒州	—	—	—	—	—	—	—	1
大宗正司丞	—	—	—	—	—	—	—	1



## References

### Databases and Visualization Tools

- Academia Sinica 中央研究院. *Scripta Sinica Database* 漢籍電子文獻資料庫. Taipei: Academia Sinica, 1984. <https://hanchi.ihp.sinica.edu.tw>
- Bastian M., Heymann S., Jacomy M. *Gephi*, 2008. <https://gephi.org/>
- Brandes, Ulrik et al. *Visone*. 1996. <https://visone.ethz.ch/>
- Gong, Yanming 龔延明, Database of Civil Service Examination Graduates in All Dynasties (*Lidai jinshi dengke shujuku* 歷代進士登科數據庫), 2016. <http://examination.ancientbooks.cn/docDengke/>
- Harvard University, Academia Sinica, and Peking University. *China Biographical Database*, 2018. <https://projects.iq.harvard.edu/cbdb>
- Zhonghua shuju 中華書局. *Songdai muzhiming shuju ku* 宋代墓誌銘數據庫, 2016. <http://inscription.ancientbooks.cn/docShike/>

### Primary Sources

- Chen, Kui 陳騭. *Nan Song guan'ge lu* 南宋館閣錄. Beijing: Zhonghua shuju, 1998.
- Cheng, Ju 程俱. *Lintai gushi jiaozheng* 麟臺故事校證. Beijing: Zhonghua shuju, 2000.
- Hong, Mai 洪邁. *Rongzhai suibi* 容齋隨筆. Beijing: Zhonghua shuju, 2005.
- Li, Xinchuan 李心傳. *Jianyan yilai xinian yaolu* 建炎以來繫年要錄. Taipei: Academia Sinica Scripta Sinica Database, 1984.
- Lu, Xinyuan 陸心源. *Song shi yi* 宋史翼. Beijing: Zhonghua shuju, 1991.
- Tuotuo 脫脫 et al. *Song shi* 宋史. Taipei: Academia Sinica Scripta Sinica Database, 1984.
- Xu, Song 徐松. *Song huiyao jigao* 宋會要輯稿. Taipei: Academia Sinica Scripta Sinica Database, 1984.
- Zeng, Zaozhuang 曾棗莊, and Liu Lin 劉琳 eds. *Quan Song wen* 全宋文. Shanghai: Shanghai cishu chubanshe, 2006.
- Zhonghua shuju 中華書局 ed. *Song Yuan fangzhi congkan* 宋元方志叢刊. Taipei: Academia Sinica Scripta Sinica Database, 1984.

### Secondary Sources

- Barabási, Albert-László. *Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life*. New York, NY: Perseus Publishing, 2002; 2014 reprinted.
- Bastian M., Heymann S., Jacomy M. "Gephi: an Open-Source Software for Exploring and Manipulating Networks." International AAAI Conference on Weblogs and Social Media, 2009.
- Baur, Michael. *Software for the Analysis and Visualization of Social Networks*. Doctoral dissertation, Universität Fridericiana zu Karlsruhe, 2008.

- Blondel, Vincent D, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Le-febvre. "Fast Unfolding of Communities in Large Networks." *Journal of Statistical Mechanics: Theory and Experiment* 10 (2008): 155–68. DOI: 10.1088/1742-5468/2008/10/P10008
- Borgatti, Stephen P. "Centrality and Network Flow." *Social Networks* 27 (2005): 55–71. DOI:10.1016/j.socnet.2004.11.008.
- Brandes, Ulrik and Dorothea Wagner. "Visone: Analysis and Visualization of Social Networks." In *Graph Drawing Software*, edited by Michael Jünger and Petra Mutzel, 321–340. Heidelberg: Springer Heidelberg, 2003. DOI: 10.1007/3-540-45848-4\_47.
- Cai, Chongbang 蔡崇榜. *Songdai xiushi zhidu yanjiu* 宋代修史制度研究. Taipei: Wenjin chubanshe, 1993.
- Deng, Xiaonan 鄧小南. *Songdai wenguan xuanren zhidu zhu cengmian* 宋代文官選任制度諸層面. Shijiazhuang: Hebei jiaoyu chubanshe, 1993.
- Dorogovtsev, S. N., J. F. F. Mendes, and A. N. Samukhin. "Giant Strongly Connected Component of Directed Networks," *Physical Review E* 64.2 (2001). DOI:10.1103/PhysRevE.64.025101.
- Galarnyk, Michael. "Understanding Boxplots," posted on the website of "Toward Data Science," <https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51> (accessed April 17 2020).
- Gong, Yanming 龔延明 ed. *Songdai guanzhi cidian* 宋代官制辭典. Beijing: Zhonghua shuju, 1997.
- Guan, Qin 管琴. *Cike yu Nan Song wenxue* 詞科與南宋文學. Beijing: Peking University Press, 2018.
- Hartman, Charles. "Introduction: Bureaucratic Institutes and Information Networks," *Journal of Song-Yuan Studies* 48 (Sep. 2019): 1–6. DOI: 10.1353/sys.2019.0014
- Hartman, Charles. "Sung Government and Politics." In *The Cambridge History of China. Vol. 5, Part Two: Sung China, 960–1279*, edited by John W. Chaffee and Denis Twitchett, 19–138. Cambridge, UK: Cambridge University Press. DOI: 10.1017/CHO9781139193061.003
- Hartman, Charles. *The Making of Song Dynasty History: Sources and Narratives*. Cambridge: Cambridge University Press, 2021. DOI: 10.1017/9781108877176
- Hartwell, Robert M. "Demographic, Political, and Social Transformations of China, 750–1150," *Harvard Journal of Asiatic Studies* 42.2 (Dec. 1982): 365–442. DOI: 10.2307/2718941
- Hartwell, Robert M. "Financial Expertise, Examinations, and the Formulation of Economic Policy in Northern Sung China," *Journal of Asian Studies* 30.2 (Feb. 1971): 281–314. DOI: 10.2307/2942916
- Hucker, Charles O. *A Dictionary of Official Titles in Imperial China*. Stanford: Stanford University Press, 1985.
- Hu, Kun 胡坤. *Songdai jianju gaiguan yanjiu* 宋代薦舉改官研究. Shanghai: Shanghai guji chubanshe, 2019.

- Jin, Zhongshu 金中樞. "Songdai guan'ge de jianzhi yu zhishi kao" 宋代館閣的建置與職始考. In *Songdai de xueshu he zhidu yanjiu jinian xuanji* 宋代的學術和制度研究紀念選集, 339–387. Xinbei, Taiwan: Daoxiang chubanshe, 2016.
- Keller, Franziska Barbara. "Analyses of Elite Networks." In *The Palgrave Handbook of Political Elites*, edited by Heinrich Best and John Higley, 135–152. London: Palgrave Macmillan, 2018. DOI: 10.1057/978-1-137-51904-7\_11
- Kracke, Edward A. Jr. *Civil Service in Early Sung China, 960–1067, with Particular Emphasis on the Development of Controlled Sponsorship to Foster Administrative Responsibility*. Cambridge, Mass: Harvard University Press, 1957.
- Lai, Ruihe 賴瑞和. *Tangdai gaoceng wenguan* 唐代高層文官. Beijing: Zhonghua shuju, 2017.
- Lai, Ruihe 賴瑞和. *Tangdai jiceng wenguan* 唐代基層文官. Beijing: Zhonghua shuju, 2004.
- Lai, Ruihe 賴瑞和. *Tangdai zhongceng wenguan* 唐代中層文官. Beijing: Zhonghua shuju, 2011.
- Lambiotte, R., J.-C. Delvenne, and M. Barahona. "Laplacian Dynamics and Multiscale Modular Structure in Networks," *IEEE Transactions on Network Science and Engineering* 1.2 (2015): 76–90.
- Li, Changxian 李昌憲. "Songdai wenguan tiezhi zhidu" 宋代文官貼職制度, *Wenshi* 文史 30 (1982): 109–135.
- Li, Geng 李更. *Songdai guan'ge jiaokan yanjiu* 宋代館閣校勘研究. Nanjing: Fenghuang chubanshe, 2006.
- Li, Zhiliang 李之亮. *Songdai jingchao guan tongkao* 宋代京朝官通考. Chengdu: BaShu shushe, 2003.
- Liu, Pujiang 劉浦江. "Deng Guangming luezhuan" 鄧廣銘略傳. In *Xiangnian Deng Guangming* 想念鄧廣銘, edited by Zhang Shilin 張世林, 21–26. Beijing: Xinshijie chubanshe, 2012.
- Lo, Winston W. *An Introduction to the Civil Service of Sung China: With Emphasis on Its Personnel Administration*. Honolulu: University of Hawaii Press, 1987.
- Newman, M. E. J. "A Measure of Betweenness Centrality Based on Random Walks," *Social Networks* 27.1 (2005): 39–54. DOI: 10.1016/j.socnet.2004.11.009
- Nodus Labs. "Network Visualization and Analysis with Gephi," <https://noduslabs.com/courses/network-visualization-and-analysis-with-gephi/units/section-1-quick-introduction-to-network-analysis/?try> (accessed September 18, 2020).
- Sun, Guodong 孫國棟. *Tangdai zhongyang zhongyao wenguan qianzhuan tuijin yanjiu* 唐代中央重要文官遷轉途徑研究. Hongkong: Longmen shudian, 1978.
- Teraji, Jun 寺地遵. *Nan Song chuqi zhengzhishi yanjiu* 南宋初期政治史研究. Taipei: Daohe chubanshe, 1995.

- Twitchett, Denis C. *The Writing of Official History under the Tang*. Cambridge: Cambridge University Press, 1992. DOI: 10.1017/CBO9780511572678
- Umehara, Kaoru 梅原郁. "Civil and Military Officials in the Sung: The Chi-lukuan System," *Acta Asiatica* 50 (1986): 1–30.
- Umehara, Kaoru 梅原郁. *Sōdai kanryō seidō kenkyū* 宋代官僚制度研究. Kyoto: Dōhōsha, 1985.
- Winkelman, John H. "The Imperial Library in Southern Sung China, 1127–1279: A Study of the Organization and Operation of the Scholarly Agencies of the Central Government," *Transactions of the American Philosophical Society* (New Series) 64, no. 8 (1974): 5–61. DOI: 10.2307/1006199
- Woldense, Josef, "The Ruler's Game of Musical Chairs: Shuffling during the Reign of Ethiopia's Last Emperor." *Social Networks* 52 (2018): 154–166. DOI: 10.1016/j.socnet.2017.07.002
- Wu, Zhihao 吳志浩, "Songdai shiren pingjun siwang nianling kao" 宋代士人平均死亡年齡考. *Zhejiang xuekan* 浙江學刊 4 (2017): 170–181. DOI:10.16235/j.cnki.33-1005/c.2017.04.022
- Xiong, Huei-Lan 熊慧嵐. "A Reservoir of Talent: An Analysis of the Career Advancement of Imperial Library Officials during the Southern Song (1127–1279)." *Journal of Song-Yuan Studies* 48 (2019): 7–56. DOI: 10.1353/sys.2019.0015
- Yu, Yunguo 虞雲國. *Nandu junchen: Song Gaozong ji qi shidai* 南渡君臣: 宋高宗及其時代. Shanghai: Shanghai renmin chubanshe, 2019.
- Yu, Yunguo 虞雲國. *Songdai taijian zhidu yanjiu* 宋代臺諫制度研究. Shanghai: Shanghai renmin chubanshe, 2014.
- Zhang, Xiqing 張希清. "Songchao gongju shihe shouguan zhidu shulun" 宋朝貢舉釋褐授官制度述論, *Zhongyuan wenhua yanjiu* 中原文化研究, no. 3 (2015): 20–28.
- Zhuge, Yibing 諸葛憶兵. *Songdai zaifu zhidu yanjiu* 宋代宰輔制度研究. Beijing: Zhongguo shehui kexue chubanshe, 2000.



The Journal of  
**HISTORICAL  
NETWORK  
RESEARCH**

PETER K. BOL

# From Kinship to Collegiality: Changing Literati Networks, 1100–1400

*Journal of Historical Network Research* 5 (2021) 87–113

**Keywords** Song dynasty, Yuan dynasty, China Biographical Database

**Abstract** This is study of changes in kinship and scholarly association in Wuzhou, a prefecture in the middle of Zhejiang province. The geographic extent of literati kinship connections became increasingly local from the twelfth century on, but paradoxically the kinship connections across the prefecture declined as well. However, at the same time cross-prefecture scholarly connections among literati increased, becoming the new foundation for literati group solidarity.

With social network analysis, it is possible to track social change at scale. An anecdote can be the starting point for an explanation of historical change, but it is data – the amassing of anecdotal information – that proves the case. This article relies on data from the China Biographical Database (discussed elsewhere in this issue by Michael Fuller and Wang Hongsu), which is a highly structured relational database of biographical data and the largest ever created for the study of Chinese history. I use this data not only to confirm a well-known argument about elite social change, but also to explore an unexpected paradox.\*<sup>1</sup>

This study is based on biographical data from the prefecture of Wuzhou (Jinhua from the mid-fourteenth century onwards) in mid-Zhejiang province in southeastern China. The prefecture, a diamond shape 74 miles from north to south and 92 miles from east to west, was made up of seven counties at the time; it had about 155,000 registered households in the mid-twelfth century and 216,000 by the end of the thirteenth. In the late 1120s, when the capital of Song China fell to the Jurchens and was driven from the north China plain to Hangzhou in the southeast, Wuzhou and many other places suddenly found themselves near the capital with easy access to the court and information. Wuzhou literati, the very small segment of the population whose education made them eligible to compete for official appointments in the civil service examinations, saw greater success than it had ever known, and success bred success. However, Wuzhou, like the rest of the Song state in the south, fell to Mongols in the 1270s, and remained under foreign rule for the next century.

“Literati” is a translation of the term *shi*, a term that had been used since antiquity to denote the national political-social elite. During the course of the Song dynasty (960–1279) shared views of what defined men as *shi* shifted from birth and pedigree to merit as represented by learning, an attainment that written examinations were intended to assess. To translate *shi* as literati recognizes this change. From the government’s perspective, the civil service examination system was a highly selective mechanism for choosing men for office. In some places, Wuzhou for example, the pass:fail ratio was set at 1:200. But from the perspective of those competing, participation alone had its rewards: meeting other men from the prefecture with a similar education, gaining access to local officials, meeting potential marriage partners, to name but a few. Above all, participa-

---

\* **Acknowledgements:** My thanks to Song Chen for his suggestions.

**Corresponding author:** peter\_bol@harvard.edu

1 “The China Biographical Database.” The database is accessible online and is freely available to download from the website, <https://projects.iq.harvard.edu/cbdb/>. For the structure, logic, and goals of the database, see the article in this issue by Michael A. Fuller and Wang Hongsu. For a detailed explanation of all features the database see Michael A. Fuller, *The CBDB User’s Guide*, (Cambridge, MA: China Biographical Database Project, 2020), [https://projects.iq.harvard.edu/files/cbdb/files/users\\_guide\\_20200927.pdf](https://projects.iq.harvard.edu/files/cbdb/files/users_guide_20200927.pdf).

tion became evidence that one was one of the *shi*, one of the literati, and thus a member of the national elite even while his life was lived out in his home locale. Although the number of candidates kept increasing and the proportion of those who passed kept decreasing, families were recompensed for their investment in educating their sons for twenty or thirty years.<sup>2</sup> However, this equation of examination participation with literati status became problematic when the Mongols, upon conquering the south in the 1270s, discontinued recruitment through literary examinations, and although they restored them in 1315, the numbers were so limited that it was no longer the best way to civil office.

Literati learning produced the sources for our knowledge of elite social associations and kinship. There is a widespread, but incorrect, notion that the education of civil service examination candidates consisted of memorizing the Confucian Classics and writing essays on passages from the Classics in a rigid format. Education meant learning to read and write in the first place and, as a student advanced, acquiring more knowledge of the textual tradition (the Classics and commentaries, historical writings, philosophical texts, and literary traditions) as well as mastering the art of composition in a variety of literary genres (poetic forms, rhapsodies, letters, inscriptions, prefaces, epitaphs, memorials, edicts, legal judgments, etc.).<sup>3</sup> Literary writings were typically “occasional;” that is, they were composed for or at a specific social occasion or event, but “occasional” understates the importance of literary writing for social communication and bureaucratic practice. All literati learned what they needed to know to participate in schools and exams, but some gained fame as “scholars”, as *xuezhe*, or “those who learn.” Being a “scholar” includes much of what we would expect – writing commentaries on the Confucian Classics and other texts, philosophical lectures, historical studies, statecraft treatises, local histories, and so on – but it almost always included producing occasional writings, often at the request of others, which in turn would be compiled by his students or descendants in the form of a literary collection. Although we might see literati who gained a following for their literary skill – whose compositions were sought after and served as models for others – as writers, they were “scholars” in the parlance of the time. Their literary collections, as well as the surviving collections of occasional writings of everyone else, are thus repositories of the literary exchanges that created the social networks of literati life.

---

2 The classic study of the Song examination system is John W. Chaffee, *Thorny Gates of Learning*. On the role of the system in certifying social status see Peter K. Bol, “The Sung Examination System”.

3 Although less attentive to the literary aspect of literati learning, the breadth of Song learning is evident from the articles in Wm. Theodore de Bary and John W. Chaffee, *Neo-Confucian Education*. The definitive history of the late imperial examination system is Benjamin A. Elman, *Cultural History of Civil Examinations*.

One genre of occasional writing, the epitaph, is the most widely-available source of what we know about kinship networks.<sup>4</sup> Epitaphs – biographical records placed in the grave of the deceased, but also included in the author's literary collection – give three generations of the patriline of the deceased, sometimes including wives and their parentage, the descendants of the deceased at the time the epitaph was written, and usually included the deceased spouses and the husbands of female descendants. Official titles of men and women are usually included. Getting a well-known person to write an epitaph added luster to the reputation of the deceased, so famous scholars were thus likely to write many epitaphs. Social historians have used these to make important arguments about how family strategies changed over time.

The general principle that literati families followed in selecting marriage partners for their sons and daughters was to treat marriage as an alliance between families that had the capacity to be of service to each other. Thus, in a family that saw government service as the family occupation, for example, a father who was an official (or aspired to be one) served his family best by marrying his children to the children of a man of equal or better status. Under such circumstances, where he was from would matter less than the rank he held. But what was thought to be of value in making a marriage was not set in stone. Social historians have argued that around the time the Song moved south, literati family strategies shifted. In essence, families shifted from pursuing national bureaucratic marriage, which would create allies within the national bureaucracy, to local marriage, which would cement relations with other powerful local families.<sup>5</sup> There are four interlocking ways of accounting for this. Institutionally, the examination system encouraged education and participation, but because the number of official positions available did not increase, literati families had ever less prospect of placing their sons in office and thus became ever less able to marry nationally; they began to conclude that local marriage alliances were vital for their descendants' well-being. Politically, there were two reasons: first, the intense fac-

---

4 For the period under discussion genealogies have potential value, as the work by Zhong Chong noted below has shown. However, that research largely depended on genealogies held in the family.

5 This was proposed by Robert Hartwell, "Demographic, Political, and Social Transformation of China, 750–1550." The local elite thesis was given substance in an intensive local study by Robert Hymes; see Robert P. Hymes, *Statesmen and Gentlemen*; "Marriage, Descent Groups, and the Localist Strategy in Sung and Yuan Fu-chou." I find this convincingly demonstrated, but some have argued against a "localization" of the elite; see Bao Weimin, "Jingyingmen difanghual ma?". Hymes's contention that this went together with the withdrawal of the state from local society should be understood as the Southern Song court's abandoning of the activist policies of late Northern Song. Beverly Bossler, in a study of Song Wuzhou families and chief councilor families, has shown that matching status was still the operative family strategy but that in Southern Song one was much more likely to find suitable partners locally; see Beverley Bossler, *Powerful Relations*. Hymes takes all this into account in Robert Hymes, "Sung Society and Social Change".



tionalism toward the end of the Northern Song period (960–1126) forced out certain families that had seen government as their family occupation; and second, the activist policies that had immensely extended the state's interference in society, economy, and education and had guided the court from the 1070s to the 1120s were abandoned as the reconstituted dynasty in the south focused its attention and resources on national defense, granting the families with bureaucratic experience but no career prospects greater leeway to play dominant roles in local society. Geographically, the move south left southern literati families in place, where they had already been for generations and, if they did attain office, gave them local bases to return to, whereas the northerners who remained with the Song state were now refugees. The counties with a strong literati presence, as measured by examination degrees, were in four areas (modern Zhejiang, Jiangxi, Fujian, and southern Jiangsu provinces) that had not been decimated by warfare (although some areas, including Wuzhou, had suffered from the Fang La rebellion in 1120). This meant that over time, families with histories of government service had gradually increased, making it easier to find marriage partners of similar status locally. A demonstrable result of these four developments was that local literati became increasingly interested in local society and their own role in its welfare and history, something that had not been apparent from their occasional writing during the Northern Song.

## 1. The Data

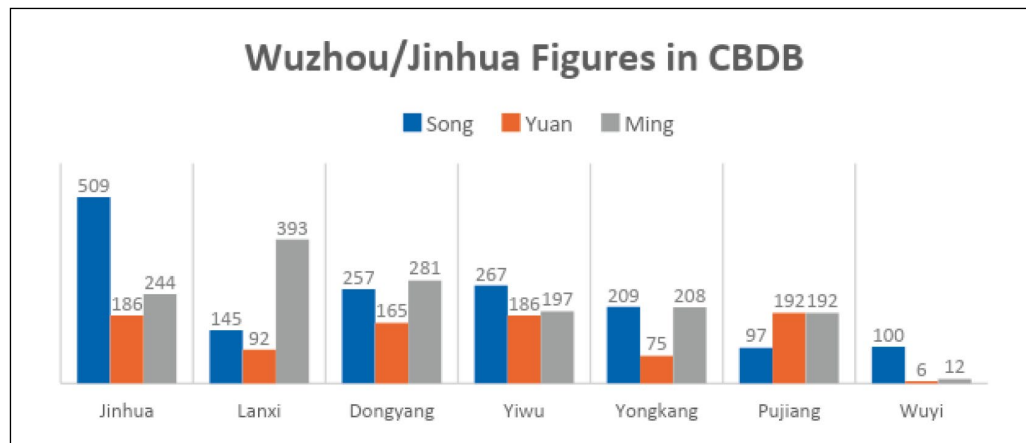
Beginning in the twelfth century, the number of Wuzhou literati competing in the civil service examinations steadily increased; some were extraordinarily successful, with four becoming chief councilors. The number of literati who became scholars and left behind literary collections and books also increased, and this continued to be the case into the late fourteenth century. This has provided a trove of data on kinship and social relationships, a good proportion of which has been included in the China Biographical Database (CBDB). For this study, all available kinship data for Wuzhou persons from Northern Song through Yuan has been entered into the CBDB.

At the time of writing, the CBDB has data on over 470,000 persons, mainly from the seventh through nineteenth centuries. Given how data is extracted from texts and the collection priorities, the amount of data varies by person and period. For the Tang, Song, and Yuan dynasties in particular, there is strong information on literary exchanges, kinship, careers, native places, and writings. Wuzhou persons in the CBDB are given in table 1.

The bar chart below shows the distribution by county and by dynasty. This is not evidence of changes in population, but of the number of persons who entered the historical record. One can see from this that Lanxi, for example, produced relatively more people of note during the Ming period.

Period	Total	With Index Year	Jinhua	Lanxi	Dong-yang	Yiwu	Yong-kang	Pu-jiang	Wuyi	Tangxi after 1471
Tang-Five Dynasties	86	67								
Song	1,584	1,375 (N. Song 229) (S. Song 1,146)	509	145	257	267	209	97	100	
Yuan	916	611	186	92	165	186	75	192	6	
Ming	1,556	1,371	244	393	281	197	208	192	12	29
Qing	316	164	33	40	125	32	20	60	3	3
ROC	20	20	8	2	4	2	4			
TOTAL	4,478	3,608	980	672	832	684	516	541	121	32

**Tab. 1** The number of people in the CBDB listed with Wuzhou as their place of affiliation (*jiguan* 籍貫) by dynasty, the number with index years, and the distribution over counties. The concept of “index year” is explained below. Northern Song 960–1126, Southern Song 1127–1276, Yuan 1270–1368, Ming 1368–1644, Qing 1644–1911, Republic of China 1911–1949.



**Chart 1** Distribution of Wuzhou persons in CBDB by county and dynasty.

Epitaphs provide the birth and death dates of the principals, however, this data is lacking for many of the people who appear in the CBDB. In most cases, however, it is possible to use other data to locate a person in time. As Fuller and Wang explain, to accomplish this the CBDB uses a heuristic, the “index year,” a person’s putative 60<sup>th</sup> year. Using index years allows one to base queries on persons we can be confident are from a given time period.

## 2. The Spatiality of Kin Relations

In order to test the theory that marriage networks narrowed, I designed a query to discover the spatial distribution of the kin relations of Wuzhou literati. In this case the results will include in and out migrants. The CBDB kinship query has four parameters, as explained in greater detail by Fuller and Wang:

- Number of Ancestral Generations
- Number of Descendent Generations
- Number of Collateral Links (number of horizontal or sibling links; a father's brother has one unit of "collateral" distance, a wife's sister has one unit of "marriage" distance and one unit of collateral distance).
- Number of Marriage Links (number of links defined by marriage; e.g. a wife's sister's husband has two units of "marriage" distance)

In all but one case I kept the collateral and marriage links constant at 1 and varied the vertical ancestral and descendant generations from 3 to 0. The results are in table 2.

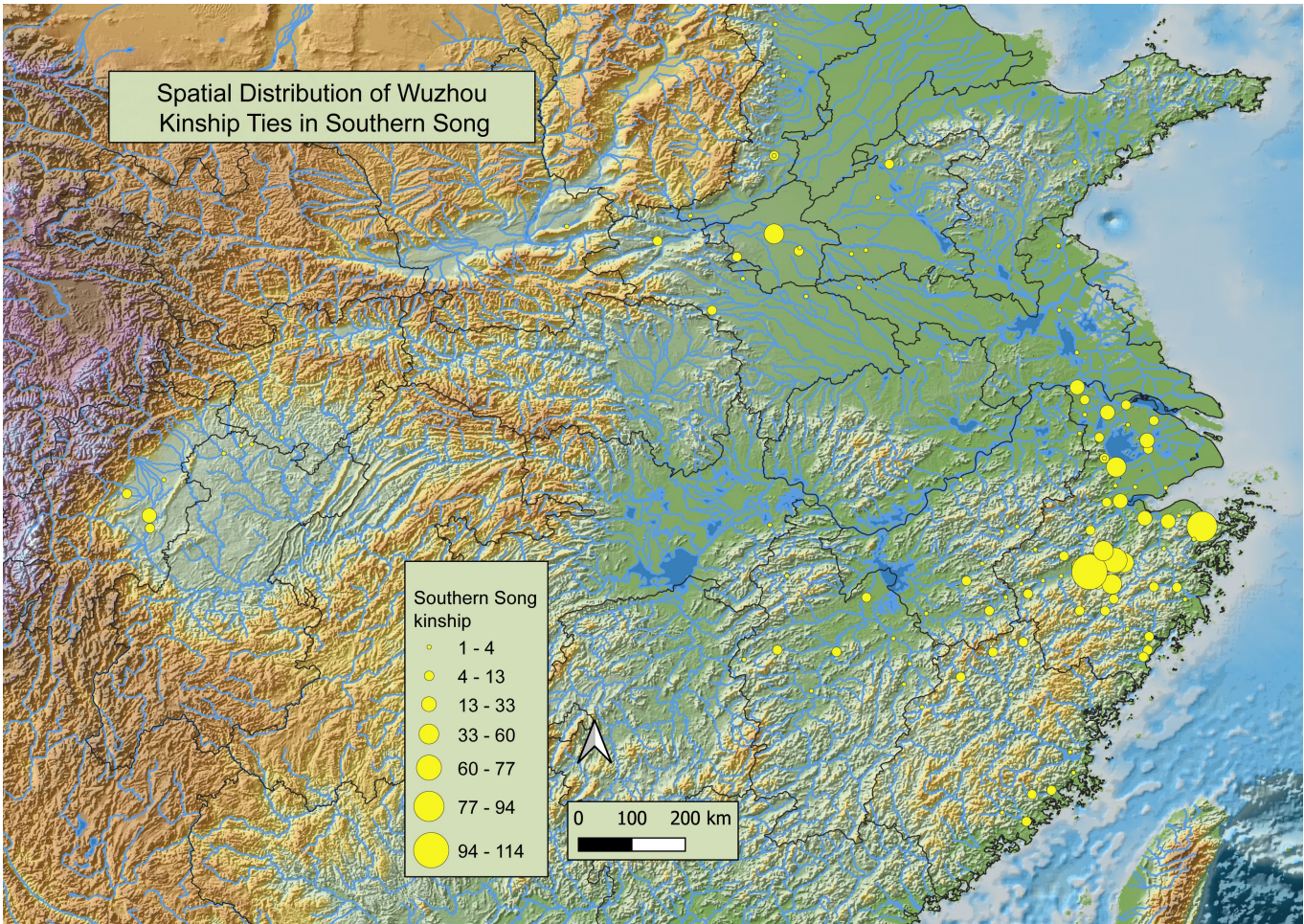
Two ways of interpreting this chart stand out. First, the Southern Song pattern is not so different from Northern Song that we would posit a change in family strategy. Second, Yuan kinship is extremely localized. Marriage did become more localized, but this is only obvious for the Yuan period. The maps below compare the spatial distribution of Southern Song and Yuan kinship with the parameters of 2-2-1-1. Ultimately, marriage did become more localized. The lack of a significant difference between Northern Song and Southern Song may be the result of refugees from the north settling in Southern Song Wuzhou. This did not happen when the Yuan conquered the Song in the 1270s.

Marriages were arranged to benefit the partner families. A narrowing of the geographic range thus means that families were able to find suitable partners locally. But what were the criteria for suitability? To put the question in another way, what were the interests that led families to see others as suitable partners, and how did these change over time? There were practical economic and security interests – families wanted affines who were at least as wealthy; there were status concerns – official status mattered, but there were also cultural concerns – in Southern Song, being well educated offered the prospect of examination success and shared values. We would expect, therefore, that as marriage partners become more local, there would be ever greater intermarriage between local literati. To test this I graphed the 2-2-1-1 kinship networks for Southern Song and Yuan. First, I identified the "giant component", namely the largest group of nodes that were connected together. Second, I labeled the persons in the component to see if different families were involved. Third, I colored the nodes by county to see if families from different counties were intermarrying. The Southern Song giant component had 1681 nodes, 57.43% of the records returned in the query. In the

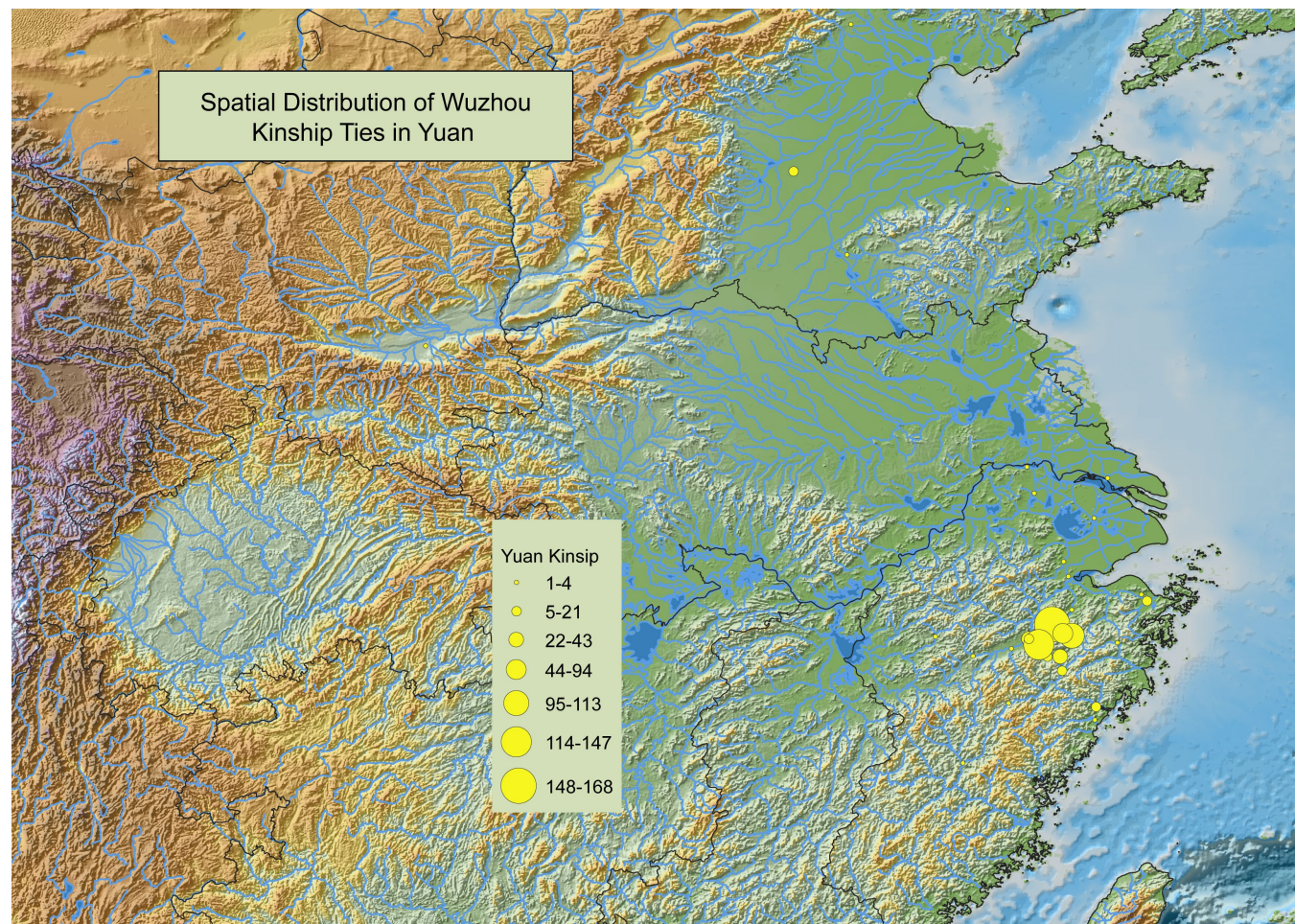
Period	Query	Total records returned	Total records with addresses	Total unique counties	Records with Wuzhou addresses	% of kin from Wuzhou = F/J	Number of persons on the query list
N Song	3-3-1-1	2249	1736	115	776	0.45	230
N Song	2-2-1-1	1218	919	80	545	0.59	230
N Song	1-1-1-1	524	388	43	284	0.73	230
N Song	0-1-1-1	368	264	31	222	0.84	230
N Song	0-0-1-1	46	23	10	16	0.70	230
N Song	0-0-0-1	22	4	2	4	1.00	230
S Song	3-3-1-1	5446	4211	203	1945	0.46	1146
S Song	2-2-1-1	3504	2526	139	1491	0.59	1146
S Song	1-1-1-1	2021	1414	89	1050	0.74	1146
S Song	0-1-1-1	1343	870	52	754	0.87	1146
S Song	0-0-1-1	319	164	29	141	0.86	1146
S Song	0-0-0-1	185	45	19	34	0.76	1146
Yuan	3-3-1-1	2510	2086	52	1885	0.90	611
Yuan	2-2-1-1	2021	1674	38	1545	0.92	611
Yuan	1-1-1-1	1363	1114	29	1042	0.94	611
Yuan	0-1-1-1	1057	877	24	846	0.96	611
Yuan	0-0-1-1	293	213	18	200	0.94	611
Yuan	0-0-0-1	150	88	15	80	0.91	611

**Tab. 2** Results of CBDB kinship queries concerning Song and Yuan persons with index years. The order is the number of Ancestral-Descendant-Collateral-Marriage generations or links. For Northern Song the 0-0-0-1 kinship results are too small to be meaningful. There is some duplication when two persons may have two different kinds of relationships, for example when ego's mother's brother is also his wife's father. However, this did not significantly alter the results. Of the 5446 records of the Southern Song 3-3-1-1 kinship query, there were only 45 duplicates (.008%), when the duplicates were stripped there were 4211 records with addresses and 203 unique addresses; of 4211 records with addresses, there were 1927 with Wuzhou addresses (45.7%, rounded to 46%).





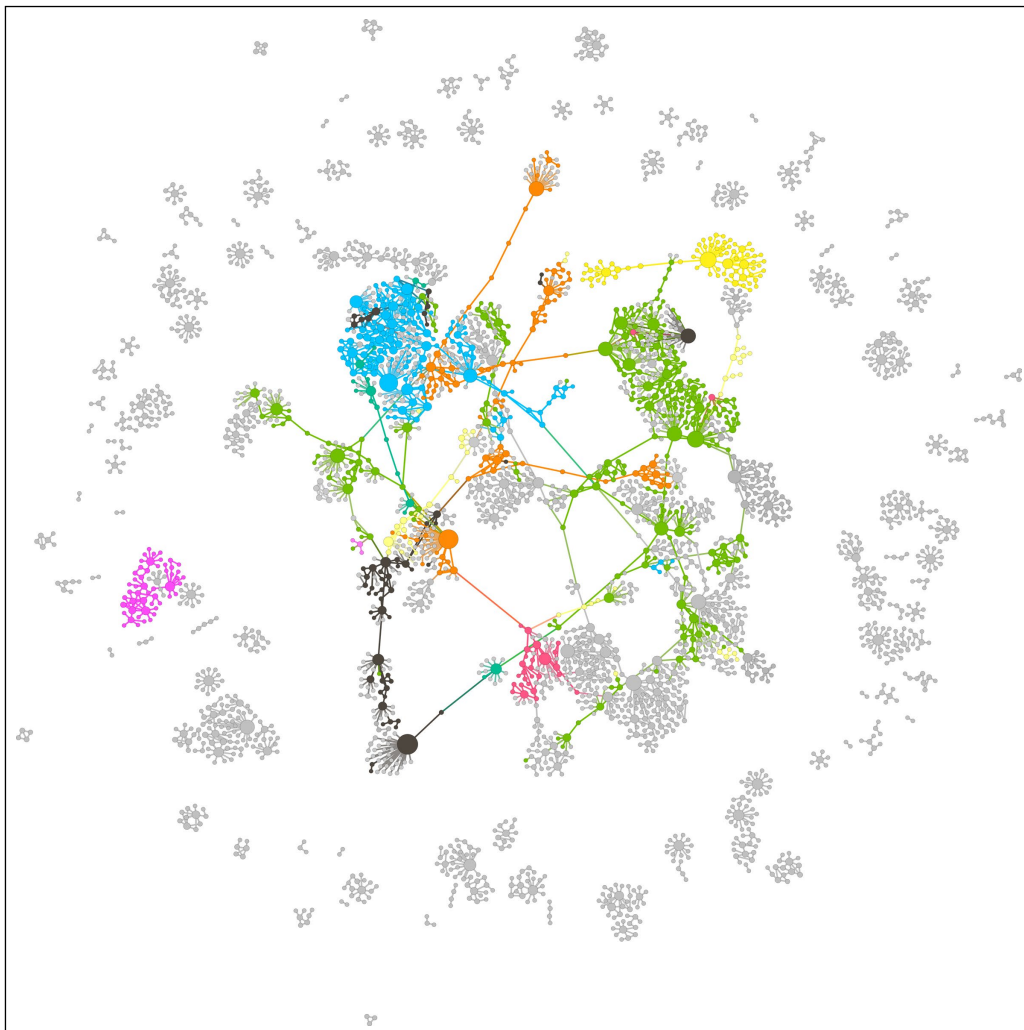




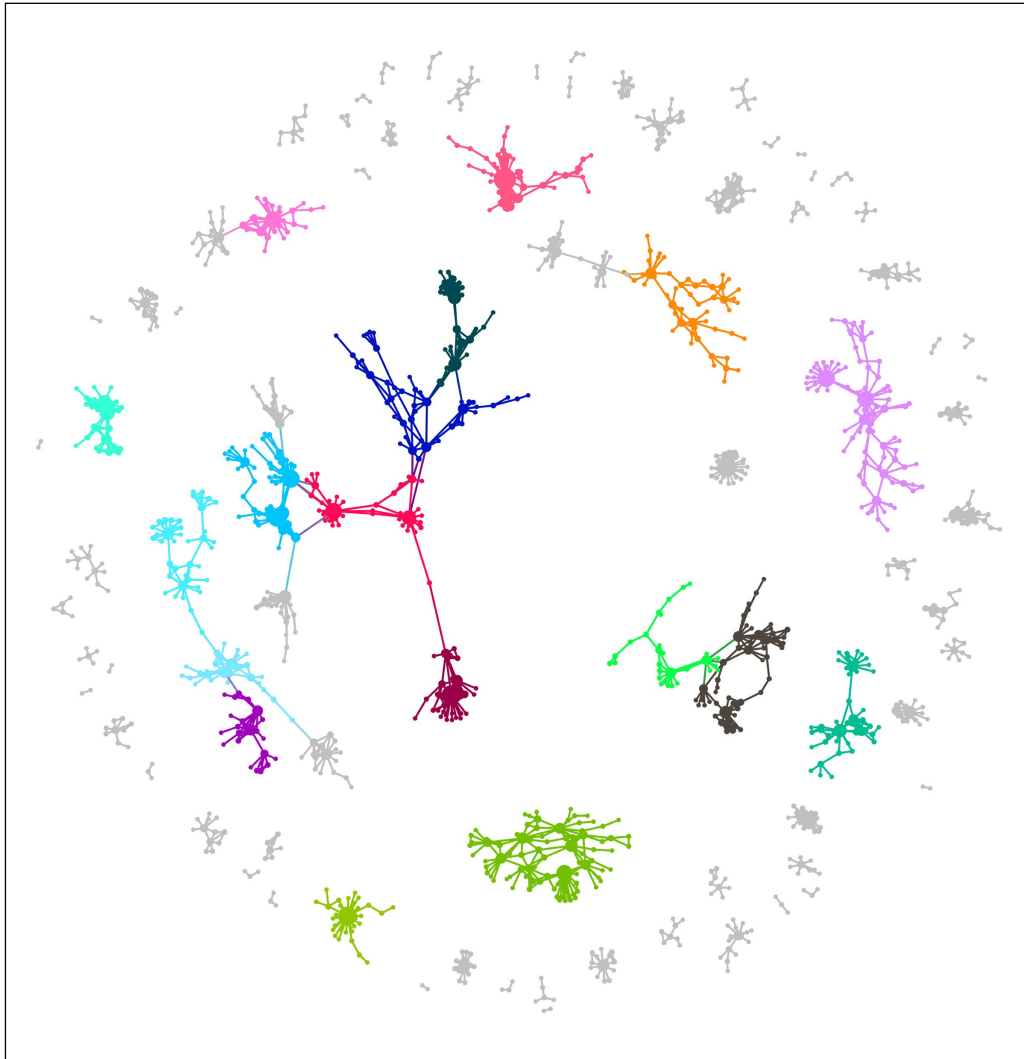
**Map 1 and 2** Spatial Distribution of Southern Song and Yuan Wuzhou persons in CBDB as reflected in a 2-2-1-1 kinship query. This level of kinship was chosen to reflect the extent of a given person's likely acquaintances. All maps by the author, using China Historical GIS layers.

figure below, Wuzhou counties are distinguished by color; grey represents a non-Wuzhou county.

Figure 2 shows the giant component of Yuan kinship. The differences are manifest. First, the giant component is much smaller, with 242 nodes out of 1321, only 18% of all nodes. Second, the graph largely segments by county, with only the northwestern quadrant showing some cross-county marriage. This is even more apparent when the other components are examined: all but one of them



**Fig. 1** The “giant component” with 1681 nodes in Southern Song 2-2-1-1 kinship relations. 56% of the records returned in the query are connected together. Wuzhou counties are colored. Unknown and non-Wuzhou counties are in grey. There is a degree of division by county and there is evidence of cross-county marriage. Graph generated with Gephi.



**Fig. 2** The giant component of Yuan 2-2-1-1 kinship is composed of 242 nodes out of 1321. The nodes for persons from Wuzhou counties are colored. Others are grey.

are county specific. In this case, the giant component is small enough that we can look more closely at the surnames involved.

What is surprising is that the records for three of the four subcomponents are not only from a single county, as shown in figure 2, but that within the county they are also predominantly from a single surname. This means that the database had the fact of a marriage but lacked data on the affines. Yuan literati in Wuzhou were not intermarrying with well-documented families. This is a puzzle.



### 3. If Not Kinship, Then What?

The examination system provided a state-sanctioned way in which local men could be recognized as literati, come to know each other, have contact with local officials, and learn from each other, even if they did not pass the examination. The foregoing shows that kinship had been an important way in which literati families established relationships with each other during the Southern Song. Before 1315, Yuan did not recognize literati through an examination system and, moreover, there is no evidence for a significant prefecture-wide marriage network. We know that literati continued to exist, but what held them together?

Accounting for the disappearance of the larger kinship network is not essential to this study, but Zhong Chong's intensive study of the genealogies of 169 lineages of 75 surnames in the North River basin in Dongyang County in Wuzhou suggests a likely explanation.<sup>6</sup> First, genealogies in this particular area were not being compiled for the first time until the thirteenth century with a high point in the fourteenth prior to the founding of the Ming, and almost none were compiled for the first time after that. Second, almost all the persons from which a genealogy began were officials or degree holders, although it can be shown in some cases that kin were already living in the locale prior to the apical ancestor of the genealogy. Third, the lineages that appear after 1400 were, with the exception of a few small ones, all branches of lineages established outside the basin. Fourth, the formation of these genealogy-based lineages went together with the advent of single-surname villages. Once single-surname villages emerged in this area, they continued; indeed, they have dominated the countryside until the present day. How they came about is not hard to fathom. Someone who became an official acquired more land – some of which they could hold tax-free. They then instituted the practice of maintaining a genealogy as a means of ensuring unity and mutual aid among their descendants, giving their kin a good reason to live together, even if they rarely held property in common. If there was adequate land and they avoided natural and human disasters, over time the estate would become a village. Descent groups can be tracked through their genealogies; generations increasing in number is a good sign that they lived in the same area, sometimes segmenting into multiple villages, and saw the advantages of kinship solidarity. Many members would likely spend their lives farming, some were merchants, and a very few might make a career as an official, teacher, or government clerk (in Yuan). Under such circumstances, the security of the growing descent group was best served by intermarriage with surrounding descent groups like

---

6 Zhong Chong, *Bei Jiang pendi*. See the map, pp. 112–113, showing the location of the lineages and tying segmented descent groups to the home lineage. For a bibliography of all the editions of all the genealogies used, see pp. 240–276. For an introduction to his methods and major findings, see “Zhejiang Dongyang shi Beijiang.”

their own.<sup>7</sup> Prefaces to genealogies written in Yuan began to promote the idea that the genealogy-based descent group – not the bureaucratic success of an individual – was not only part of the moral enterprise of social cohesion and unity, but was also the way in which literati could serve the common good of society from the bottom up.<sup>8</sup>

But how did literati in Wuzhou maintain their identity as literati under the circumstances? This became a particularly pressing issue during the Yuan period, first when there was no civil service examination system and thereafter when it was reduced to a minor avenue of recruitment. There are anecdotes about families with a history of service and scholarship who stopped making the effort. We can imagine that from a practical point of view, ambitious people might see greater value in lineage building and controlling land and water. No doubt some proportion of Wuzhou families with literati backgrounds did exactly this. What could those who thought, for whatever reason, that being literati mattered do?

The answer in Wuzhou was learning, and Wuzhou was notable during the Yuan as a place where literati learning flourished. There was also a bit of an incentive: the household (not the wider descent group) with a son registered at a county school or a state-recognized academy was relieved of labor service, something that set them apart from the merely well-to-do, although keeping a son in school for decades was also expensive. Chinese literati from the south had very little prospect of advancing to high office under the Mongols, but for those seeking office recommendation, not examination, it was the most promising route, and although for many this meant appointment as a teaching official or academy master, such a post was quite respectable. However, the recommendation required recognition. We might expect that in Yuan, literati would have depended more heavily on learning to create social networks that would thus provide recognition, and that they would seek access to the nearest, best-known, and best-connected literati in the locality. Collegiality would replace kinship and networks of collegiality were likely to take shape locally.

---

7 Zhong has found examples of this. Even today, marriage relations continue between the Yaxi Lu 雅溪盧 family of Luzhai 盧宅 and the nearby Li family of Lizhai 李宅. The seven volume Lu genealogy is now in the Harvard Yenching Library; *Yaxi Lu shi*.

8 Peter K. Bol, “Local History and Family in Past and Present.”

#### 4. Collegiality

In the following, I test these two propositions. Networking through learning was not new. Was this only a matter of degree? Or was something else changing as well?

First, I will explain the CBDB's approach to social associations. CBDB codes about 300 different kinds of associations. The CBDB network query allows queries by selected associations. The output also reports relationships between the persons discovered in the query. For example, a query that reports that A, B, C, and D were the students of X will also report if A had relationships with B, C, or D. Thus, a query about the associations of a group of people will reveal multiple networks and subnetworks. Some, such as friendship and teacher-student relationships are attested to by a third party, but most are evidenced by literary exchanges. The CBDB allows the user to select which associations to include in a query. For the purposes of this query, I included all codes for the friendship, teacher-student, scholarly affiliation, common membership, academic patronage, literary/artistic associations, and "writings" that evidence associations between people; the latter is the largest category of all. Writings includes subtypes: commemorative texts, epitaphs, prefaces and postfaces, ritual texts, biographical texts, explanatory texts, admonitory texts, correspondence, and various other texts for social occasions.

I began with the lists of persons with Southern Song and Yuan index years used in the kinship query. However, in contrast to the multiple generational kinship queries, the social association was limited to a distance of 1; that is, I included the friend but not the friend's friend, a distance of 2. I wanted to compare two periods of equal length that had some similarities. One was to include a generation that lived through the conquest by the Yuan in the first case and the Ming in the second. Another was that both periods were times of scholarly activity, so that the results would not be skewed in favor of one period. The first group had index years from the year prior to the death of Lü Zuqian in 1181 (as he did much to activate literati learning in Wuzhou), through to 1294, the last year of the conqueror Khubilai's Khan's reign. This was meant to capture the post-Lü Zuqian generations and those who lived through the Song-Yuan transition. The second, 1295 (post-Khubilai) to 1403 (the Yongle usurpation) was meant to capture persons who came of age during the Yuan and those who lived through the Yuan-Ming transition. I limited queries to one node distance and constrained the query to include associates with index years in the same period index year, in order to avoid generating extremely large networks and reduce the number of associations to people from earlier and later periods (for example, a person in 1400 who writes a preface to a book Lü Zuqian wrote in 1170 has formed a meaningful association, but not one that involves another person at the time). However, this also meant that persons whose index years were unknown were excluded from the results. This dramatically reduced the size of the results. For example, a query

of only teacher-student associations in the 1295–1403 list returned 99 instances, but when persons without index years were included the same query returned 178 instances. We should assume that the query results are indicative of relative volume rather than a representation of all relevant data in the CBDB.

Table 3 records the number of persons in the query, the number of persons returned by the query (nodes), and the total number of associations. It shows that there were almost twice as many relations per node in Yuan as in Song, that there was a giant component comprising twice the number of persons, and that there were on average twice as many associations per year. The Yuan literati network was larger and more active than it had been in Song.

In addition to the metrics of size and frequency, we can also show the central figures in the Wuzhou literati learning network and where they came from. The density of the graphs in figures 3 and 4 is impressive, although it makes them less useful. They do, however show major figures in the network. In both cases, we are seeing the “giant component.”

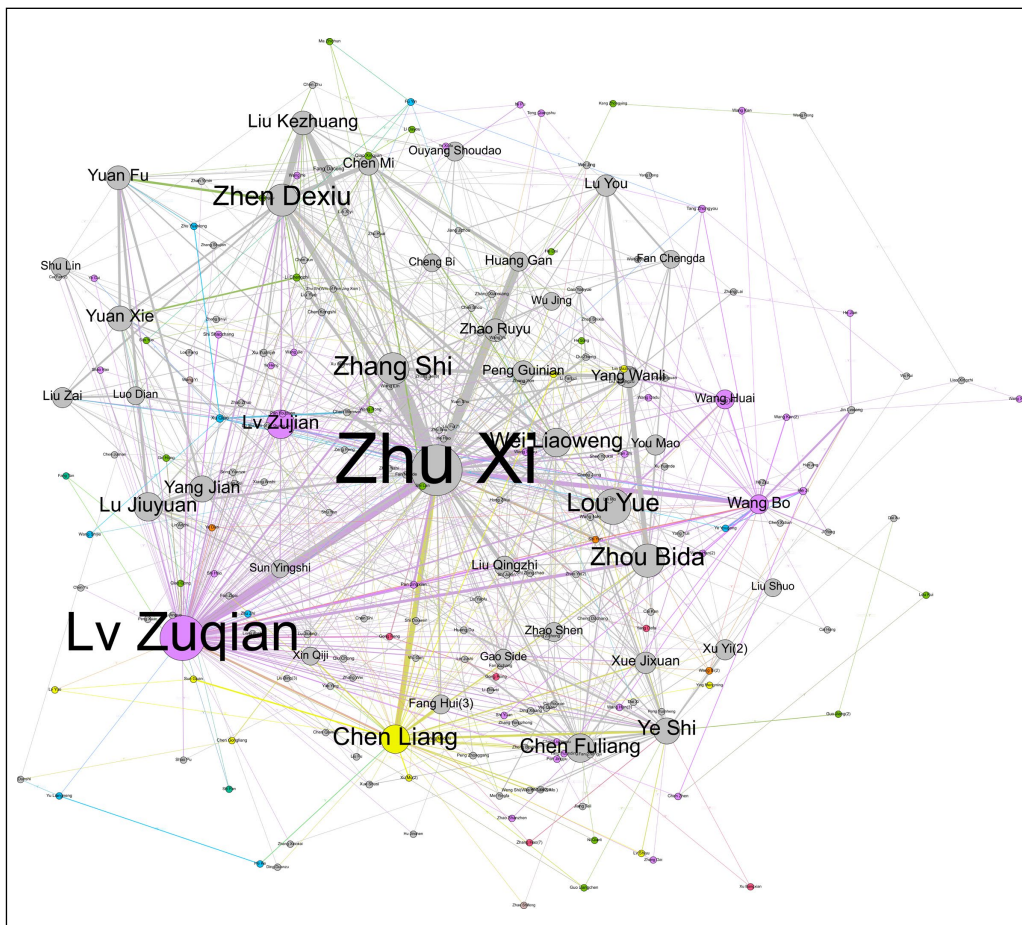
A more useful view, figures 5 and 6, shows those people in the network who have at least seven associations.

Four network algorithms were used to assess relative centrality in the network. The most straightforward is “degree” (or “degree centrality”). This is the total number of persons a figure is connected to as recorded in the CBDB. Slightly more complicated is “weighted degree,” which takes the number of persons one is connected with, but also how many instances of connections he has with each of them. In Southern Song Wuzhou, for example, Lü Zuqian was connected to more people than Zhu Xi (node size) but Zhu Xi’s weighted degree was greater (label size). The third is “betweenness centrality,” which measures the degree to which a person connects persons together who would otherwise be unconnected. For example, suppose there are two separate networks of three people each, and then a seventh person who has connections to both networks. That person is “between” the two and has the greatest “betweenness centrality” of all seven. The fourth is “eigenvector centrality,” which computes the approximate importance of each node as the sum of the centrality values of the nodes that it is connected to. This is a measure of influence within a network.

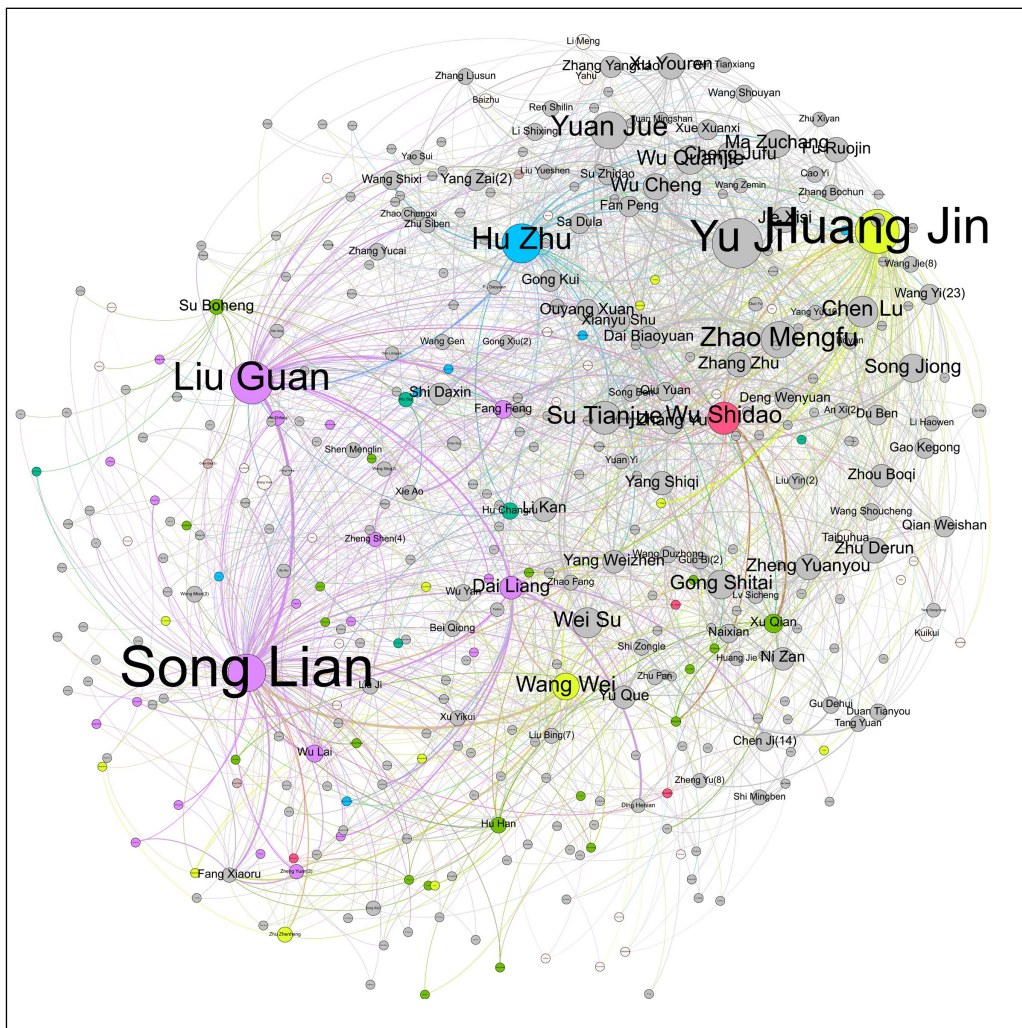
Networks will always have more and less central figures. In this case, I also want to know where the most central figures are from and the degree to which local figures draw together literati from across the prefecture. Comparing Lü Zuqian and Song Lian in table 4, the local figures with the most associates, we see that both had prefecture-wide networks and national networks. But this similarity belies an important difference between Song and Yuan in terms of who dominated local networks. In figures 3–6, the nodes are color coded to differentiate between persons from Wuzhou and others. The difference is more apparent in

Index Year Periods	Persons in query list	Unique persons (nodes)	Associations (edges)	Ave. number of relations per node	Giant component unique persons (nodes)	Giant component edges	Percent-age in the giant component	No. of years	Assoc. per year
1180–1294	947	1180	1019	.9	430	1008	0.36	115	8.86
1295–1403	766	1439	2169	1.5	851	2167	0.60	109	19.89

**Tab. 3** Social associations in Southern Song and Yuan compared.

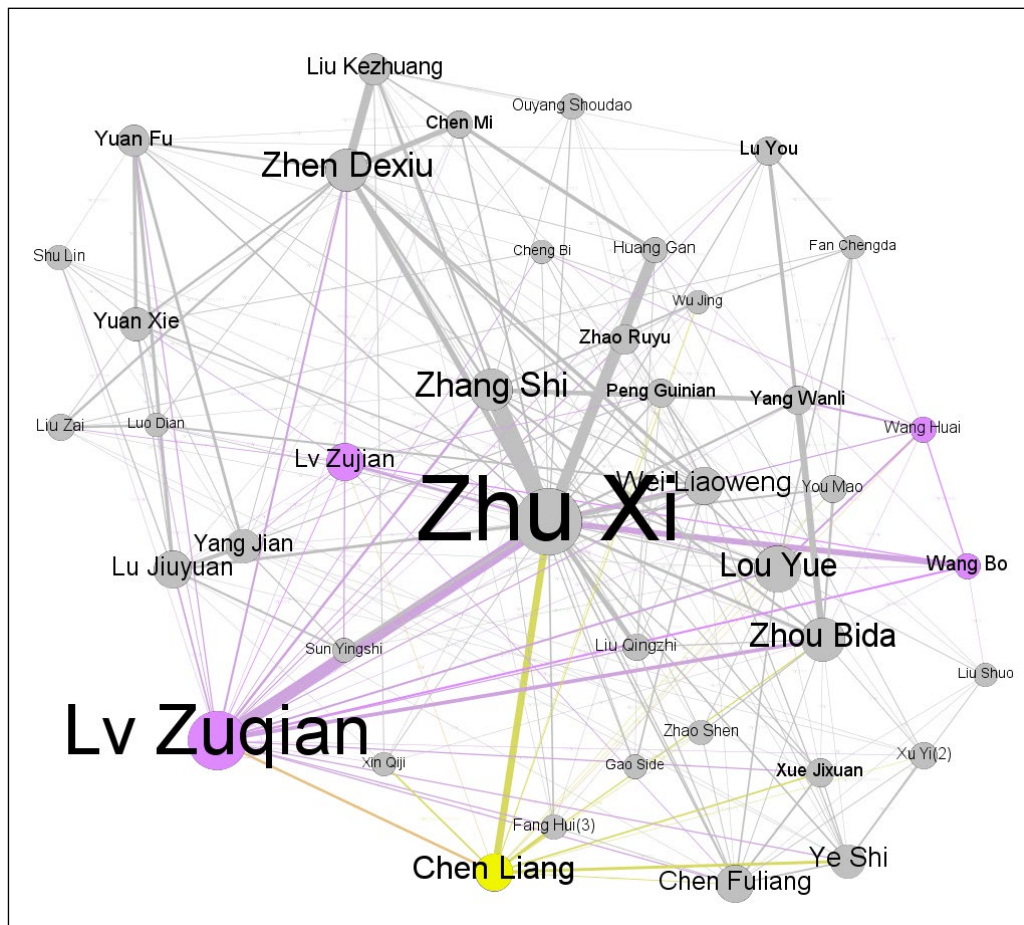


**Fig. 3** The “giant component” in literati networks with associations limited to persons with index years 1180–1294 (left) and 1295–1403 (right). The size of the node represents the number of connections a person has, the label size represents the weighted degree, colored nodes correspond to native counties in Wuzhou, and grey nodes represent people with unknown addresses and non-Wuzhou addresses. The size of the name corresponds to the weighted degree. The thickness of the edge reflects the number of associations between persons.



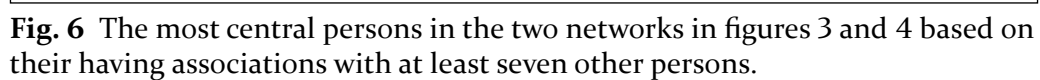
**Fig. 4** The “giant component” in literati networks with associations limited to persons with index years 1180–1294 (left) and 1295–1403 (right). The size of the node represents the number of connections a person has, the label size represents the weighted degree, colored nodes correspond to native counties in Wuzhou, and grey nodes represent people with unknown addresses and non-Wuzhou addresses. The size of the name corresponds to the weighted degree. The thickness of the edge reflects the number of associations between persons.





**Fig. 5** The most central persons in the two networks in figures 3 and 4 based on their having associations with at least seven other persons.





	nodes	edges	addresses	Wuzhou counties	Wuzhou nodes	Wuzhou edges
Lü Zuqian	215	365	70	7	85	123
Song Lian	361	497	93	6	86	142

**Tab. 4** Comparing the networks of Lü Zuqian in Song and Song Lian in Yuan-Ming.

tables 5 and 6, which give the results of the four algorithms for the two periods, listing the 25 highest-ranking persons in each category. Those from Wuzhou are in grey cells. The most obvious finding is that during the first period, the majority of important network figures were from outside Wuzhou.

However, from Yuan into Ming, 1295–1403, local men dominated networks of learning.

Amalgamating all the results, during the first period only 22% of the 25 most central persons were Wuzhou literati, but during the second period the number doubles to 44%. If we observe only the top ten persons in Song-Yuan, then 27% were from Wuzhou, but in the Yuan-Ming period they made up 70%. Taking into account just the top 25 persons, in the Southern Song-Yuan period only 32% of the associations were with Wuzhou literati, versus 65% in the Yuan-Ming period.

Scholar Networks in Wuzhou, 1180–1294					
Degree	Weighted degree	Eigenvector centrality		Betweenness centrality	
Lǚ Zuqian (215)	Zhu Xi (488)	Lǚ Zuqian	1.00	Lǚ Zuqian	0.083
Zhu Xi (125)	Lǚ Zuqian (357)	Zhu Xi	0.74	Zhu Xi	0.034
Chen Liang (78)	Zhen Dexiu (150)	Zhang Shi	0.34	Chen Liang	0.027
Wang Bo (65)	Chen Liang (149)	Chen Liang	0.34	Wang Bo	0.027
Ye Shi (53)	Wang Bo (149)	Zhen Dexiu	0.32	Ye Shi	0.013
Zhang Shi (44)	Zhang Shi (130)	Zhou Bida	0.30	Zhou Bida	0.007
Zhen Dexiu (39)	Ye Shi (120)	Ye Shi	0.30	Lou Yue	0.006
Zhou Bida (36)	Zhou Bida (100)	Lou Yue	0.30	Wei Liaoweng	0.006
Wei Liaoweng (36)	Liu Kezhuang (99)	Chen Fuliang	0.29	Zhen Dexiu	0.005
Lou Yue (35)	Wei Liaoweng (95)	Wei Liaoweng	0.28	Liu Kezhuang	0.004
Lu Jiuyuan (31)	Lou Yue (72)	Wang Bo	0.26	Lǚ Zujian	0.004
Chen Fuliang (30)	Lu Jiuyuan (70)	Lǚ Zujian	0.26	Fang Dacong	0.004
Liu Kezhuang (27)	Huang Gan (64)	Lu Jiuyuan	0.26	Lu You	0.004
Lǚ Zujian (24)	Yang Wanli (62)	Yuan Xie	0.22	Zhang Shi	0.004
Lu You (23)	Chen Fuliang (59)	Yuan Fu	0.20	Yuan Fu	0.003
Huang Gan (22)	Chen Mi (59)	Chen Mi	0.20	Xu Qiao	0.003
Yuan Xie (21)	Yuan Fu (59)	Yang Wanli	0.20	Huang Gan	0.003
Chen Mi (20)	Lu You (53)	Zhao Ruyu	0.19	Jin Lǚxiang	0.003
Yuan Fu (19)	Yuan Xie (53)	Lu You	0.19	Tang Zhongyou	0.002
Yang Wanli (18)	Lǚ Zujian (43)	Liu Kezhuang	0.18	Zhu Zhi	0.002
Yang Jian (18)	Yang Jian (41)	Yang Jian	0.17	He Ji	0.002
Zhao Ruyu (15)	Zhao Ruyu (31)	You Mao	0.16	Lu Jiuyuan	0.002
Liu Zai (14)	Liu Zai (31)	Liu Qingzhi	0.16	Qiao Xingjian	0.002
Fang Dacong (13)	Fang Dacong (29)	Fang Hui	0.15	Chen Fuliang	0.002
He Ji (13)	Liu Qingzhi (29)	Shu Lin	0.15	Wang Shijie	0.002

**Tab. 5** Four measures of centrality of persons in the network depicted in Figure 5 left. Wuzhou persons are in grey cells.

Scholar Networks in Wuzhou, 1295–1403					
Degree	Weighted degree	Betweenness centrality		Eigenvector centrality	
Song Lian (361)	Song Lian (528)	Song Lian	0.20	Song Lian	1.00
Huang Jin (240)	Huang Jin (316)	Huang Jin	0.12	Huang Jin	0.77
Liu Guan (110)	Liu Guan (173)	Liu Guan	0.04	Liu Guan	0.49
Wang Wei (86)	Yu Ji (155)	Wang Wei	0.03	Yu Ji	0.47
Yu Ji (77)	Dai Liang (128)	Dai Liang	0.03	Hu Zhu	0.40
Dai Liang (68)	Wang Wei (126)	Su Boheng	0.03	Zhao Mengfu	0.35
Hu Zhu (64)	Hu Zhu (105)	Xu Qian	0.01	Wu Shidao	0.33
Su Boheng (56)	Yuan Jue(2) (89)	Hu Zhu	0.01	Wang Wei	0.32
Yuan Jue (55)	Su Tianjue (82)	Yu Ji	0.01	Dai Liang	0.30
Wu Shidao (53)	Wu Cheng (82)	Wu Shidao	0.01	Yuan Jue	0.30
Zhao Mengfu (51)	Wu Shidao (81)	Fang Xiaoru	0.01	Wei Su	0.29
Wei Su (42)	Zhao Mengfu (81)	Zhao Mengfu	0.01	Su Tianjue	0.27
Su Tianjue (40)	Wei Su (66)	Hu Han	0.01	Wu Quanjie	0.26
Xu Qian (39)	Su Boheng (65)	Yang Weizhen	0.01	Gong Shitai	0.25
Chen Lu (38)	Wu Quanjie (65)	Wei Su	0.01	Cheng Jufu	0.25
Gong Shitai (36)	Yang Weizhen (62)	Wu Lai	0.00	Chen Lu	0.24
Yang Weizhen (35)	Xu Qian (59)	Hu Changru	0.00	Xu Qian	0.23
Wu Cheng (35)	Chen Lu (59)	Xie Ao	0.00	Li Kan	0.22
Zhang Zhu (34)	Zheng Yuanyou (57)	Gong Shitai	0.00	Ma Zuchang	0.22
Xu Youren (33)	Gong Shitai (55)	Fang Feng	0.00	Xianyu Shu	0.22
Zheng Yuanyou (33)	Dai Biaoyuan (54)	Wu Cheng	0.00	Yang Weizhen	0.22
Cheng Jufu (32)	Cheng Jufu (53)	Zheng Shen	0.00	Zheng Yuanyou	0.21
Ma Zuchang (32)	Zhang Zhu (51)	Cheng Jufu	0.00	Wu Cheng	0.21
Song Jiong (31)	Fang Xiaoru (51)	Zhang Yining	0.00	Zhang Yu	0.21
Wu Quanjie (29)	Xu Youren (50)	Chen Lu	0.00	Shi Daxin	0.20

**Tab. 6** Four measures of centrality of persons in the network depicted in Figure 6 right. Wuzhou persons are in grey cells.

## 5. Conclusion

From Song into Yuan, localization in kinship and learning took place along parallel tracks. Although the spatial distribution of kinship networks in Southern Song was not significantly different from Northern Song, in Yuan, kinship was almost entirely local. Similarly, Southern Song networks of learning formed around a mixture of local and external persons, whereas in Yuan local literati were the central figures. But behind this parallel lies an important difference. The prefecture-wide kinship network of Southern Song disappeared in Yuan when marriages became strictly local. However, there was a network that brought literati together across the prefecture, a network based on literati learning. The difference with the learning network of Southern Song was a difference in size, intensity, and leadership.

The literati were both a social and a cultural formation. These two aspects may have changed in similar ways, but they were not the same. It was possible to have a learning network without a parallel marriage network, and it was possible to have marriage alliances that were based on lineage alliances rather than office and examinations. I think this was a predictable outcome of the new models of decent group solidarity that took shape in Southern Song and had been accumulating more members with every generation. The larger the lineage the fewer the literati or, to put it another way, the larger the lineage the greater the number of members who had to devote themselves to making a living. Learning networks made it possible for some members to transcend their circumstances.<sup>9</sup>

The localization of intellectual leadership was not total. In Southern Song, literati learning was aimed at preparing literati for national roles. Zhu Xi was the most important of the scholars from elsewhere who spread their ideas about learning and self-cultivation by attracting students and correspondents from many places. But why should local scholars have become so dominant in Yuan? The answer is two-fold. First, a need for introductions and recommendations made personal contact important, and having famous people at hand was convenient. Literati like Song Lian, who had already gained fame as a writer and intellectual while living in Wuzhou before he entered the orbit of the Ming dynasty founder, had themselves benefitted from the attention of their seniors. Local scholarship could be self-sustaining. This leads to a second reason: Wuzhou was one of the first places in Yuan where local scholars successfully promoted the idea

---

9 Some lineages made considerable efforts to maintain the literati identity of their lineage, by investing in education and working to keep ties to the state. The Zheng family of Pujiang is a famous example of this and it has continued into the present, although most of its members were farmers. See Chang Jianhua 常建華, “Yuan Ming shiqi yimen Zheng shi ji qi guifan de shehui yingxiang 元明時期義門鄭氏及其規範的社會影響”; Danjō Hiroshi, “Gimon Teishi to Genmatsu”; John W. Dardess, “The Cheng Communal Family.”

that there were local traditions of learning that local literati should care about.<sup>10</sup> This was, in my view, part of an effort by local scholars to persuade local elites to invest in literati learning and accept literati views of moral behavior, good government, and social responsibility at a time when literati could not count on government support. In doing so, they separated collegiality from kinship.

## References

- Bao, Weimin 包偉民. "Jingyingmen difanghualie ma? Shilun Han Mingshi 〈Zhengzhijia yu shenshi〉 yu "difanshi" yanjiu fangfa" 精英們“地方化”了嗎？－試論韓明士〈政治家與紳士〉與“地方史”研究方法. *Tang Yanjiu* 唐研究 II (2005): 653–72.
- Bol, Peter K. "Local History and Family in Past and Present." In *The New and the Multiple: Sung Senses of the Past*, edited by Thomas H. C. Lee. Local History and Family, 307–48. Hong Kong: Chinese University Press, 2004.
- "The Rise of Local History: History, Geography, and Culture in Southern Song and Yuan Wuzhou." *Harvard Journal of Asiatic Studies* 61, no. 1 (2001): 37–76.
- "The Sung Examination System and the Shih." *Asia Major* 3rd ser. 3, no. 2 (1990): 149–71.
- Bossler, Beverley. *Powerful Relations: Kinship, Status, and the State in Sung China (960–1279)*. Cambridge, Ma.: Harvard University, Council on East Asian Studies, 1998.
- Chaffee, John W. *The Thorny Gates of Learning in Sung China: a Social History of Examinations*. Cambridge; New York: Cambridge University Press, 1985.
- Chang, Jianhua 常建華. "Yuan Ming shiqi yimen Zheng shi ji qi guifan de shehui yingxiang" 元明時期義門鄭氏及其規範的社會影響. *Hebei xuekan* 河北學刊 31, no. 2 (2011): 61–67.
- "The China Biographical Database." Harvard University Fairbank Center for Chinese Studies, Academia Sinica Institute of History and Philology, Peking University Center for Research on Ancient Chinese History, 2011–.
- Danjō, Hiroshi 檀上寛. "Gimon Teishi to Genmatsu no shakai" 義門鄭氏と元末の社會. *Tōyō gakuhō* 63 (1981): 299–335.
- Dardess, John W. "The Cheng Communal Family: Social Organization and Neo-Confucianism in Yuan and Early Ming China." *Harvard Journal of Asiatic Studies*, no. 34 (1974): 7–52.
- de Bary, Wm. Theodore, and John W. Chaffee. *Neo-Confucian Education: The Formative Stage*. Berkeley: University of California Press, 1989.
- Elman, Benjamin A. *A Cultural History of Civil Examinations in Late Imperial China*. Berkeley: University of California Press, 2000.

---

10 See the discussion of Wu Shidao and his *Record of Honoring Our Locale* in Peter K. Bol, "The Rise of Local History."

- Fuller, Michael A. *The CBDB User's Guide*. Cambridge, MA: China Biographical Database Project, 2020. [https://projects.iq.harvard.edu/files/cbdb/files/users\\_guide\\_20200927.pdf](https://projects.iq.harvard.edu/files/cbdb/files/users_guide_20200927.pdf).
- Hartwell, Robert. "Demographic, Political, and Social Transformation of China, 750–1550." *Harvard Journal of Asiatic Studies* 42, no. 2 (1982): 365–442.
- Hymes, Robert P. "Sung Society and Social Change." In *Cambridge History of China*, Vol. 5, Part Two, edited by John W. Chaffee and Denis Twitchett, 526–664. Cambridge: Cambridge University Press, 2015.
- "Marriage, Descent Groups, and the Localist Strategy in Sung and Yuan Fu-chou." In *Kinship Organization in Late Imperial China, 1000–1940*, edited by Patricia Buckley Ebrey and James L. Watson, 95–136. Berkeley: University of California Press, 1986.
- *Statesmen and Gentlemen: The Elite of Fu-Chou, Chiang-hsi, in Northern and Southern Sung*. Cambridge: Cambridge University Press, 1986.
- Yaxi Lu shi jiasheng 雅溪盧氏家乘. 7 vols. Dongyang: Luzhai xiuzhi weiyuanhui 盧宅修志委員會, 2001.
- Zhong, Chong 鍾翀. *Beijiang pendi: zongzu, juluo de xingtai yu fasheng shi yanjiu* 北江盆地: 宗族、聚落的形態與發生史研究. Beijing: Shangwu yinshuguan, 2011.
- "Zhejiang Dongyang shi Beijiang pendi zongzu de xingcheng yu zhan-kai-dongnan Zhongguo zongzu yu zongzu cun fasheng zhi lishi dilixue kaocha" 浙江東陽市北江盆地宗族的形成與展開 – 東南中國宗族與宗族村發生之歷史地理學考察. In *Zouru lishi de shenchu: Zhongguo dongnan diyu wenhua guoji xueshu yantaohui lunwenji* 走入歷史的深處: 中國東南地域文化國際學術研討會論文集, edited by Wu Songdi 吳松弟, Lian Xiaoming 連曉鳴 and Hong Zhenning 洪振寧, 361–380. Shanghai: Shanghai renmin chubanshe, 2011.

CÉCILE ARMAND/CHRISTIAN HENRIOT

# From Textual to Historical Networks: Social Relations in the Biographical Dictionary of Republican China

*Journal of Historical Network Research* 5 (2021) 114–153

**Keywords** Biographical dictionary, China, Cooccurrence; elites, NLP

**Abstract** In this paper, we combine natural language processing (NLP) techniques and network analysis in order to systematically map the individuals mentioned in the *Biographical Dictionary of Republican China*, thus revealing its underlying structure. We depart from previous studies due to the distinction we make between the subject of a biography (bionode) and the individuals mentioned within a biography (object-node). We examine whether the bionodes form sociocentric networks based on shared attributes (provincial origin, education, etc.). Our major contribution consists of annotating the links between individuals in order to: (1) question the assumption that word cooccurrences equate to actual relations; (2) define a more accurate classification of relationships among elites in republican China. We demonstrate that political and professional relations in this population outweigh the types of social ties commonly accepted in scholarship on modern day China. We ultimately develop a method that can be applied to similar corpora in a critical and comparative perspective.



## 1. Introduction\*

A biographical dictionary is by definition a work centered on individuals whose lives provide the backbone of distinct biographical narratives. The amount and scope of information in such works fall short of the breadth and depth of full biographical works, in which the life of an individual is minutely described and usually closely intermeshed with the social, political, economic events of the times. Even with the best of efforts and the intention to offer a macro-reading of historical events – which was an explicit goal of the editors of the *Biographical Dictionary of Republican China* (BDRC) – the format of more or less short biographical notes necessarily curtailed this ambition.<sup>1</sup> This holds especially true for the social relations and contacts that an individual had in the course of his/her life. In the condensed biographical notes that make up a dictionary, all the related historical actors are reduced to brief and often unique mentions in the body of the text. Moreover, due to the involvement of many contributors – vs. a single author in a biographical work – such mentions are unsystematic, with no apparent rationale as to the selection of the people included beyond the subjects of the biography.

In this paper, we propose a systematic mapping of all the individuals whose names appear in the biographical notes in order to reveal the networks underlying the BDRC. We argue that the links between individuals in the biographical texts create an interlinked reference network of the biographical texts.<sup>2</sup> This network can in turn be used to examine to what degree the cooccurrence of names is constitutive of relations between individuals and whether these relations can be further qualified. We follow in the footsteps of previous experiments on the relevance of network analysis into exploring a world of word cooccurrences (named individuals) in biographical texts and establish the existence of actual social networks based on these named entities.<sup>3</sup> Our approach, however, differs from pre-

---

\* **Acknowledgements:** We would like to acknowledge the contribution of Pierre Magistry (Aix-Marseille University) and Baptiste Blouin (Aix-Marseille University) who prepared the data used in this paper, advised on the methods, and organized the annotation workflow. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 788476).

**Corresponding author:** Cécile Armand, IrAsia, Aix-Marseille University, cecile.armand@gmail.com

- 1 Howard L. Boorman and Richard C. Howard, *Biographical Dictionary of Republican China* (New York: Columbia University Press, 1967), I, vii.
- 2 Christopher N. Warren et al., "Six Degrees of Francis Bacon: A Statistical Method for Reconstructing Large Historical Social Networks," *Digital Humanities Quarterly* 010, no. 3 (July 12, 2016).
- 3 Matje van de Camp and Antal van den Bosch, "The Socialist Network," *Decision Support Systems* 53, no. 4 (November 2012): 761–69; Matje van de Camp and Antal van den Bosch, "A Link to the Past: Constructing Historical Social Networks," in *Proceedings of*

vious studies in the distinction that we make between the two different kinds of population in the dictionary: those who were the subject of a biography, and those who were merely mentioned in a biography. There is a considerable imbalance in the wealth of information on each group. The latter is reduced to simply a name, except when they belonged to the group of individuals biographed. We argue that this distinction is necessary when analyzing the data in network analysis.

This paper is structured as follows. Section 1 describes how we built a reference network from robustly recognized person-to-person cooccurrences with the highest possible accuracy. In our case, we built the network as a directed network in which the nodes are individuals, and when individual B is mentioned in the biography of A, we added a directed edge from A to B. We examine this network of cooccurrences (textual links) in the first section of the paper to study the underlying structure of the BDRC and propose an alternative reading of the dictionary and its population on a global scale. In section 2, we explore whether sociocentric subnetworks form on the basis of the specific attributes (provincial origins, education, etc.) that we extracted from the biographies. In section 3, we shift the point of observation from the study of networks of cooccurrences to that of social networks. To this end, we enriched the network of cooccurrences with annotations that qualified the nature of relations, in order to: (1) distinguish mere cooccurrences from actual social relationships; and (2) build subnetworks based on the nature of relations, which we can compare with the corresponding subnetworks of attributes.

## 2. The BDRC as a network of cooccurrences

In this section, we proceed in two steps. First, we describe the workflow for extracting named entities from the BDRC and building the reference network of cooccurrences. Second, we experiment with various methods (global and local metrics, pruning tables, clustering) in order to analyze its structure.

---

*the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, WASSA '11 (Stroudsburg, PA, USA: Association for Computational Linguistics, 2011), 61–69; Minna Tamper, Eero Hyvönen, and Petri Leskinen, “Visualizing and Analyzing Networks of Named Entities in Biographical Dictionaries for Digital Humanities Research,” EasyChair Preprints (EasyChair, April 8, 2019); Pablo Aragon et al., “Biographical Social Networks on Wikipedia: A Cross-Cultural Study of Links That Made History,” in *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration – WikiSym '12* (Eighth Annual International Symposium, Linz, Austria: ACM Press, 2012), 1.*

The BDRC consists of four volumes published between 1967 and 1971<sup>4</sup> and has served generations of China historians. It was produced under the editorship of Howard L. Boorman, with contributions from around 100 different authors. Although the biographies eventually passed through the hands of a small group of editors, the first thing that our digital forensics revealed was the inconsistencies in the vocabulary used to describe individuals, positions, and institutions. The four volumes contain 589 individual biographies of unequal length – from 576 to 18,000 tokens – that feature “eminent Chinese” of the Republican period (1912–1949). This constitutes a very small sample of the Chinese Republican elites, by any standard, and the criteria for selecting this group of historical figures have proved debatable. Nonetheless, a great number of people (3,178) are mentioned in these 589 biographies, which come under three main categories: family members, authors, and other individuals.

As a rule, the biographers provided information on the family of the biographed individuals, usually starting the biographical notes with the genitors or those who raised them, if known. The latter may not be the same as the genitors, due to death or adoption. Each biography thus starts with birth and childhood, with a discussion of the family background. In most cases, the biographers cited only the name of the father, almost never that of the mother, even in the case of prominent families. The father and mother of people of humble origin were simply not named. Generally, at the end of each biography, there is also often, but not always, a list of the biographed character’s direct family members, namely wife/wives and children. Most of the time, the information provided is sketchy, especially for the wives, except when they were themselves prominent figures, socially, intellectually, or politically (the Song sisters, Ding Ling, etc.).

Within the category of authors, we grouped all the individuals who wrote about the biographed character and whose works are cited in the biography. Some of these are people who were effectively in contact with the biographed person in the course of his/her life. This is the case with former students who compiled and edited the writings of their former mentor, or sometimes next-of-kin (son-in-law, nephew, etc.). The majority of such works, however, were produced *ex post facto* by individuals who were unrelated to the person. Finally, the group of authors also includes historians and professional biographers who wrote extensive monographs or papers on the major figures in the BDRC.

To index the content and to identify all the named entities in the text, we processed the 589 biographies with Stanford CoreNLP.<sup>5</sup> Data extraction produced a raw file of 3,178 persons that listed all the biographed persons and all the individ-

---

4 Howard L. Boorman and Richard C. Howard, *Biographical Dictionary of Republican China* (New York: Columbia University Press, 1967–1971).

5 <https://stanfordnlp.github.io/CoreNLP>

uals mentioned in their biographies. The high number of cooccurrences is indicative of the wealth of data available in the BDRC beyond the 589 biographed persons, something that one may perceive through conventional reading but will fail to embrace to its full extent. The number of individuals mentioned in each biography varies greatly, from 124 for Jiang Jieshi (蔣介石) to just one for Li Yizhi and Ma Buqing. Based on this list, we built a directed network linking each biographed person (hereafter “bionode”) to the individuals mentioned in his/her biography (hereafter “object-nodes”).<sup>6</sup> The direction of arrows indicates whether an individual mentions (outgoing edges) or is mentioned by (incoming edges) another individual.<sup>7</sup> We counted each pair of individuals only once, even if an individual is mentioned several times in a biography. Given the high number of cooccurrences and the nature of the BDRC – a collection of individual biographies – what is the relational structure of the dictionary? Is it merely an aggregate of multiple ego-networks, or does it form an interconnected global network?

The network of cooccurrences generated from the extracted data comprises 3,254 nodes and 9,524 edges. It is made up of a total of 11 components, with one giant component (3,177 nodes with 9,377 edges) and ten disconnected components. All of the latter, except one, are in fact isolated ego-networks built around one single bionode. The exception is a small component consisting of two small ego-networks that centered on Kang Cheng (Ida Kahn) and Shi Meiyu (Mary Stone), respectively. In fact, these two figures were the first Chinese female physicians trained in the United States at the end of the 19<sup>th</sup> century. Both received the help and support of the same female missionary (Gertrude Howe), through whom their ego-networks are interconnected and form a small component.<sup>8</sup> The other ego-networks revolve around individuals with very specific profiles: some had careers that unfolded mostly before the Republican era (Ye Changchi, Wang Ganchang) or in religious organizations (Wei Zhuomin, Zheng Hefu), while others were scientists who either spent a lot of time abroad or made most of their career outside of China (physicists Wu Jianxiong, Qian Xuesen), or whose profile fully diverged from the “mainstream population” in the BDRC (Li Yizhi, Pei Wenzhong). It cannot be said that these individuals were poorly connected since what we catch here are just mentions of names in their biographies. What can be said is that neither they nor the individuals named in their biographies were re-

- 
- 6 We borrowed this distinction between bionodes and nodes from Henrike Rudolph, “Structures of Empowerment: A Network Exploration of the Collective Biographies of Women Activists in Twentieth-Century China,” in *Knowledge, Power, and Networks: Elites in Transition in Modern China*, ed. Cécile Armand, Christian Henriot, and Huei-min Sun (Leiden: Brill, forthcoming).
- 7 The extraction process was done in R-studio. The data and the script are available on Git-Lab (<https://gitlab.com/enpchina/brelations>).
- 8 Connie Anne Shemo, *The Chinese Medical Ministries of Kang Cheng and Shi Meiyu, 1872–1937: On a Cross-Cultural Frontier of Gender, Race, and Nation* (Bethlehem: Lehigh University Press, 2011).

lated to any of the nodes in the main component. Why were these individuals selected if they seemed rather off the mark? The probable answer lies between the editors' decision to feature "representatives" of different sectors of society and the availability of source materials.

Within the main component, how and to what extent are the biographies interconnected? How much do they rely on object-nodes to be interconnected? Do the latter contribute to the connectedness of the global network? Previous studies of biographical dictionaries have generally focused only on the biographed individuals and their relations.<sup>9</sup> In this paper, we move a step further and compare the entire network of cooccurrences with the network consisting only of bionodes. Once the object-nodes are excluded, the number of disconnected components increases to 16, with 15 isolated individuals. The network of bionodes, however, remains highly connected. If we compare the global metrics of the two networks, the density increases tenfold (entire network = 0.002/bionode network = 0.028) and the clustering coefficient multiplies by a factor of four (entire network = 0.083/bionode network = 0.337).<sup>10</sup> In both networks, there is still a high degree of connectedness given the considerable number of nodes in each.

Beyond global metrics, we use various centrality measures in order to examine the relative position of nodes and bionodes: edge count (number of neighbors); indegree (number of incoming edges); outdegree (number of outgoing edges); and betweenness centrality. The edge count displays a long-tailed distribution, with a minimum of one tie and a maximum of 367 (Jiang Jieshi).<sup>11</sup> We defined six thresholds as shown in Table 1.

It can be observed from Table 1 that 2,154 individuals are mentioned only once in the BDRC. These nodes are each linked to only a single biography and are strictly related to the life of this individual. Although the vast majority becomes part of the main component of the BDRC network by virtue of their association with at least one bionode, their presence makes sense only in relation to this particular individual. These individuals all came from the object-node category, except for Li Yizhi (engineer, 1882–1938), who happened to have the lowest number of edges among the bionodes. The group of 496 individuals with 2 to 5

---

9 Aragon et al., "Biographical Social Networks on Wikipedia"; van de Camp and van den Bosch, "The Socialist Network"; Tamper, Hyvönen, and Leskinen, "Visualizing and Analyzing Networks of Named Entities in Biographical Dictionaries for Digital Humanities Research."

10 Technically, network density shows how densely the network is populated with edges. It is a value between 0 and 1. A network which contains no edges and solely isolated nodes has a density of 0. In contrast, the density of a clique is 1. The clustering coefficient measures the ratio of the number of edges between neighbors and the maximum number of edges that could possibly exist in the network.

11 We have not included isolated nodes (13) in this table.

Edge count	Number of bionodes	Percentage	Number of nodes	Percentage	Total
100–367	13	2,2%	0	0,0%	13
50–99	40	6,8%	0	0,0%	40
25–49	153	26,0%	2	0,1%	155
6–24	345	58,6%	51	1,9%	396
2–5	37	6,3%	459	17,2%	496
1	1	0,2%	2153	80,8%	2154
Total	589	100,0%	2665	100,0%	3254

**Tab. 1** Edge count.

ties includes 459 object-nodes and 37 bionodes. The latter represents a group of individuals who are either of lesser historical importance, more peripheral (scientists), or those with a shorter lifespan. Within the network of only bionodes, a significant percentage (19 percent) of bionodes also have only this range of ties.

In the next group of 396 individuals with between 6 and 24 ties, we find mostly bionodes (345) who directly connect to one another. However, it also includes 51 object-nodes who appear in a good number of biographies. This category comprises a wide range of profiles and includes both Chinese and foreign elites.<sup>12</sup> Among the foreigners, we can distinguish two groups of people that are highly connected, although they are from different social circles – American philosophers or military advisers, and Soviet/Comintern agents. In both cases, this is about foreign experts involved in Chinese politics. The Chinese in this group include political figures from the late imperial period (emperor Guangxu, Li Hongzhang, Zeng Guofan) and intellectual figures from the Republican period (Mao Dun, Zhang Junmai, Yan Huiqing, Ding Wenjiang), some of whom had political connections. Zhang Zhidong appears in the biographies of individuals of a similar generation with whom he had direct (Sheng Xuanhuai) or indirect contact (Shen Jiaben, Yan Fu), but was also a figure of inspiration to younger people (Guo Bingwen).

12 By order of importance: Zhang Zhidong (23), Li Hongzhang (18), general George Marshall (17), Zhang Junmai (Carson Chang), Mao Dun (16 each), John Dewey and Yan Huiqing (W. W. Yen) (15 each), Wu Chaoshu (C. C. Wu) and Zeng Guofan (14), Emperor Guangxu (13), Komintern agent Gregory Voitinsky, warlord Lu Yongxiang, and the Indian writer Rabindranath Tagore (12 each).

Outside of bionodes, the two individuals with the highest number of edges are Joseph Stalin (30) and Michael Borodin (27), both due to their direct or indirect role in Chinese politics. Stalin is mostly mentioned in the biographies of members of the Chinese Communist Party (CCP) (17), but also in relation with major political figures of the Guomindang (Jiang Jieshi, Jiang Jingguo, Song Ziwen, etc.) and left-wing personalities (Song Qingling, Liao Chengzhi, Deng Yanda, etc.). Borodin, in contrast, is barely mentioned in the biographies of CCP figures, despite his close interaction with communist leaders in China as the main Comintern representative. His network includes most of the major Guomindang figures, from Sun Zhongshan to Jiang Jieshi, which reflects his activity and direct relations with these individuals during the mid-1920s in Guangzhou.

Finally, the top three categories with more than 50 edges include exclusively bionodes. This group consists of the main figures of the Guomindang (Sun Zhongshan, Jiang Jieshi, Wang Jingwei, Hu Hanmin), two communist leaders (Mao Zedong, Zhou Enlai), all the main warlords (Duan Qirui, Zhang Zuolin, Feng Yuxiang, Wu Peifu, Zhang Xueliang), the transitional figure of Yuan Shikai, and two intellectuals (Liang Qichao, Hu Shi).

In order to better understand the underlying structure of the network of bionodes, we applied Marilyn Levine's method of dissecting the "hairball" by using pruning tables.<sup>13</sup> We take the edge count as the criterion for pruning the network. As shown on the pruning tables and the pruned graphs below, no significant change occurs until we remove the bionodes with 25 ties or more. From this point on, the number of ties and nodes in the network is more than halved at each step. In the final step (Graph F), there remains only the four pivotal figures in the BDRC – Yuan Shikai, Sun Zhongshan, Jiang Jieshi, and Mao Zedong – i.e., the four major leaders that shaped the conventional narrative of the Republican period. This is just a preliminary exploration – a more systematic utilization of the pruning method will help to penetrate more deeply the structure of such complex networks.

The hierarchies based solely on edge count, however, do not take into account the directed nature of the network. In order to refine our analysis, we need to distinguish between incoming and outgoing edges. To this end, we selected the individuals with an outdegree above 20 and calculated the ratio between indegree and edge count. Table 3 presents the results for the top 25 bionodes ranked by edge count (i.e., bionodes with more than 70 mentions in the BDRC). The ratio between indegree and edge count (last column) serves as a general indicator of how often an individual was mentioned in other biographies. It reinforces the impression that individuals with a very high rate of indegree are those who play an

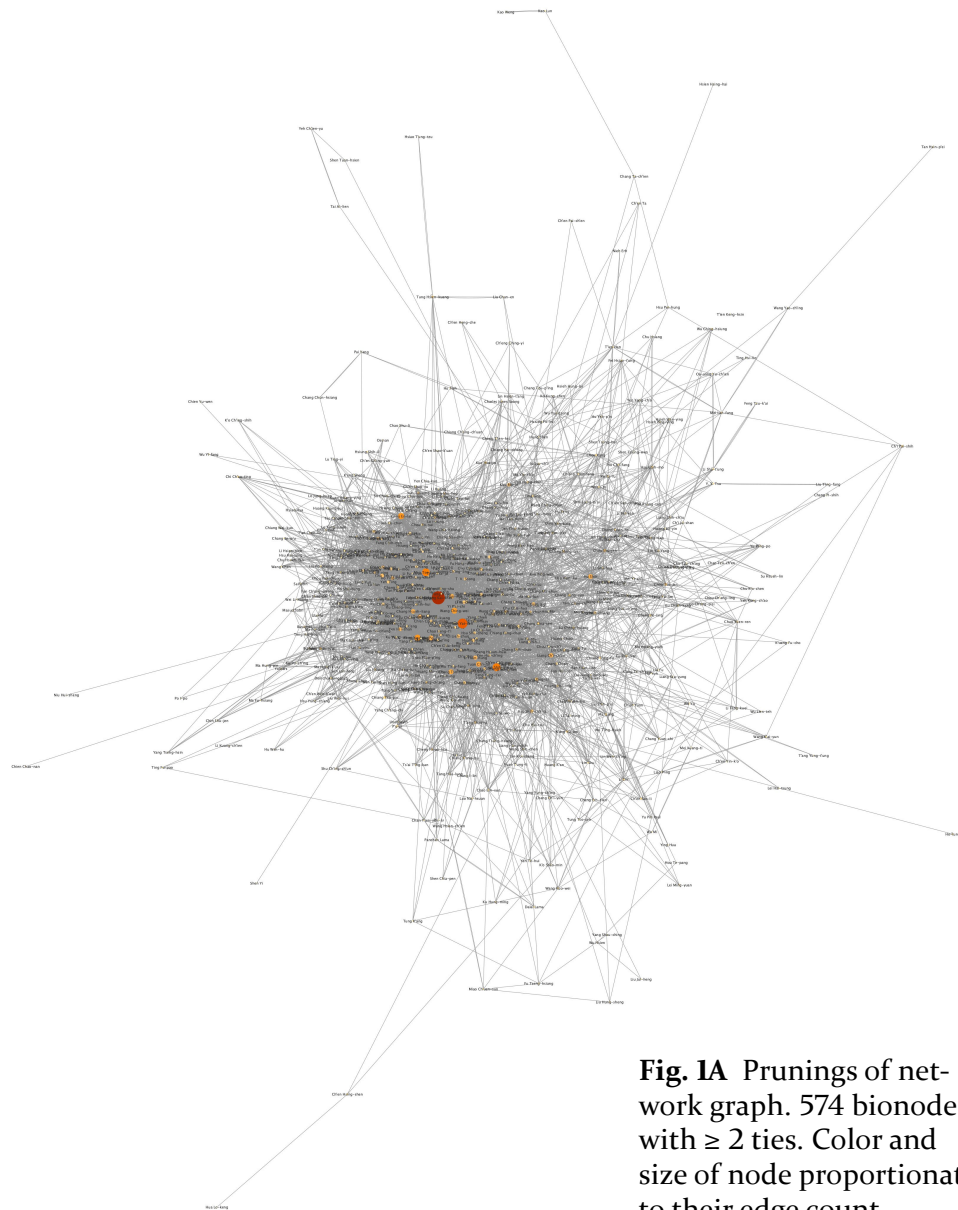
---

13 Marilyn Levine, "Post WWI Chinese Revolutionary Leaders in Europe," *Journal of Historical Network Research* 5 (2021): 187–232.

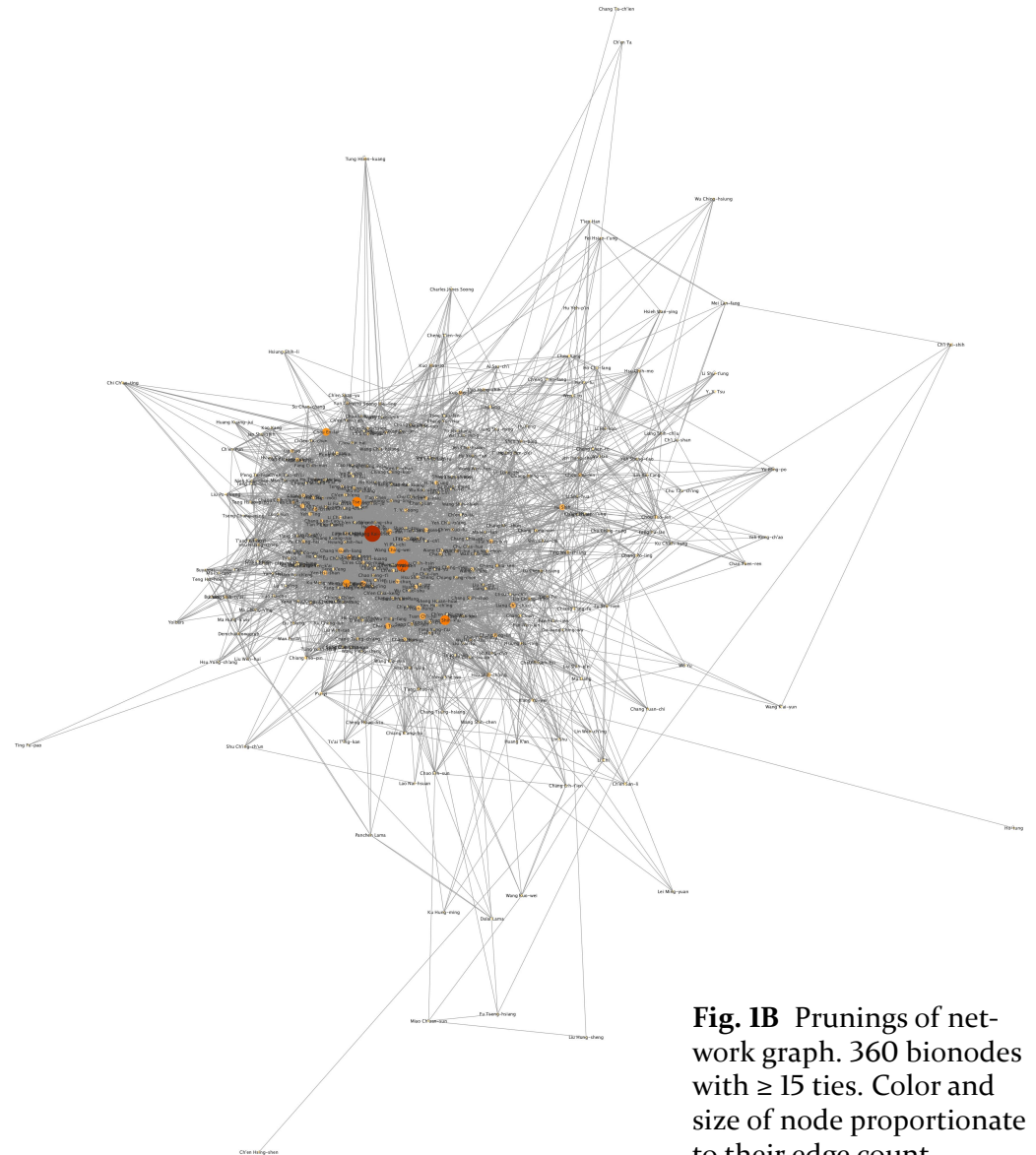
Ties Pruned	Network ties	Pruned individuals	Ratio ties/nodes	Remaining individuals	Representative individuals
0_1	5610	15	10	574	Li Zhi
2_5	5528	33	10	541	Niu Huisheng, Bo Yibo, Dai Ailian
6_9	5185	80	11	461	Luo Ruiqing, Hu Die, Gao Weng
10_14	4470	101	12	360	Qian Yongming, Ma Yinchu, Pan Gongzhan
15_19	3637	84	13	276	Chu Minyi, Tang Enbo, He Xiangning
20_24	2840	71	14	205	Chen Qimei, Gao Gang, Wu Guozhen
25_35	1620	87	14	118	Buyantai, Yan Fu, Zhang Boling
36_49	648	64	12	54	Zhou Fohai, Song Jiaoren, Pu Yi
50_74	175	34	9	20	Ding Wenjiang, Cao Kun, Kong Xiangxi
75_99	102	6	7	14	Zhu De, Wang Jingwei
100_199	10	10	3	4	Liang Qichao, Yan Xishan, Li Zongren

**Tab. 2** Pruning table with the range of edge counts, the ratio between remaining ties and nodes, and the remaining bionodes in the BDRC network at each threshold.

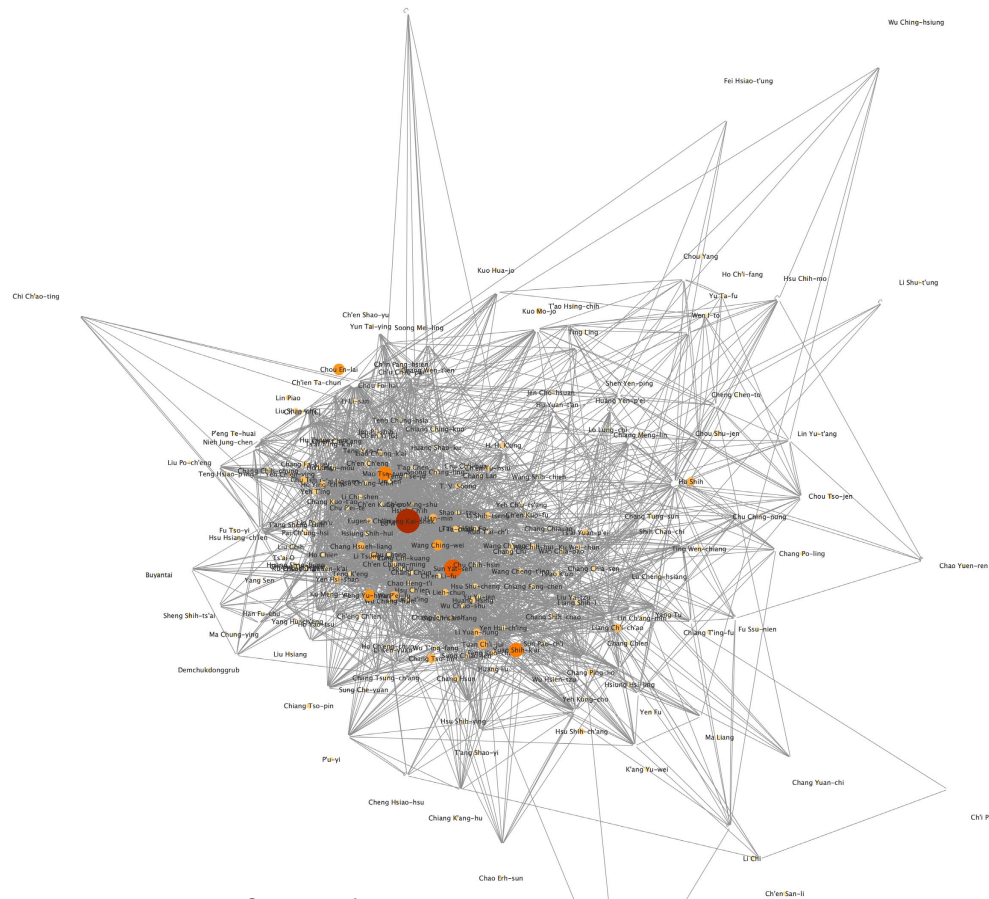




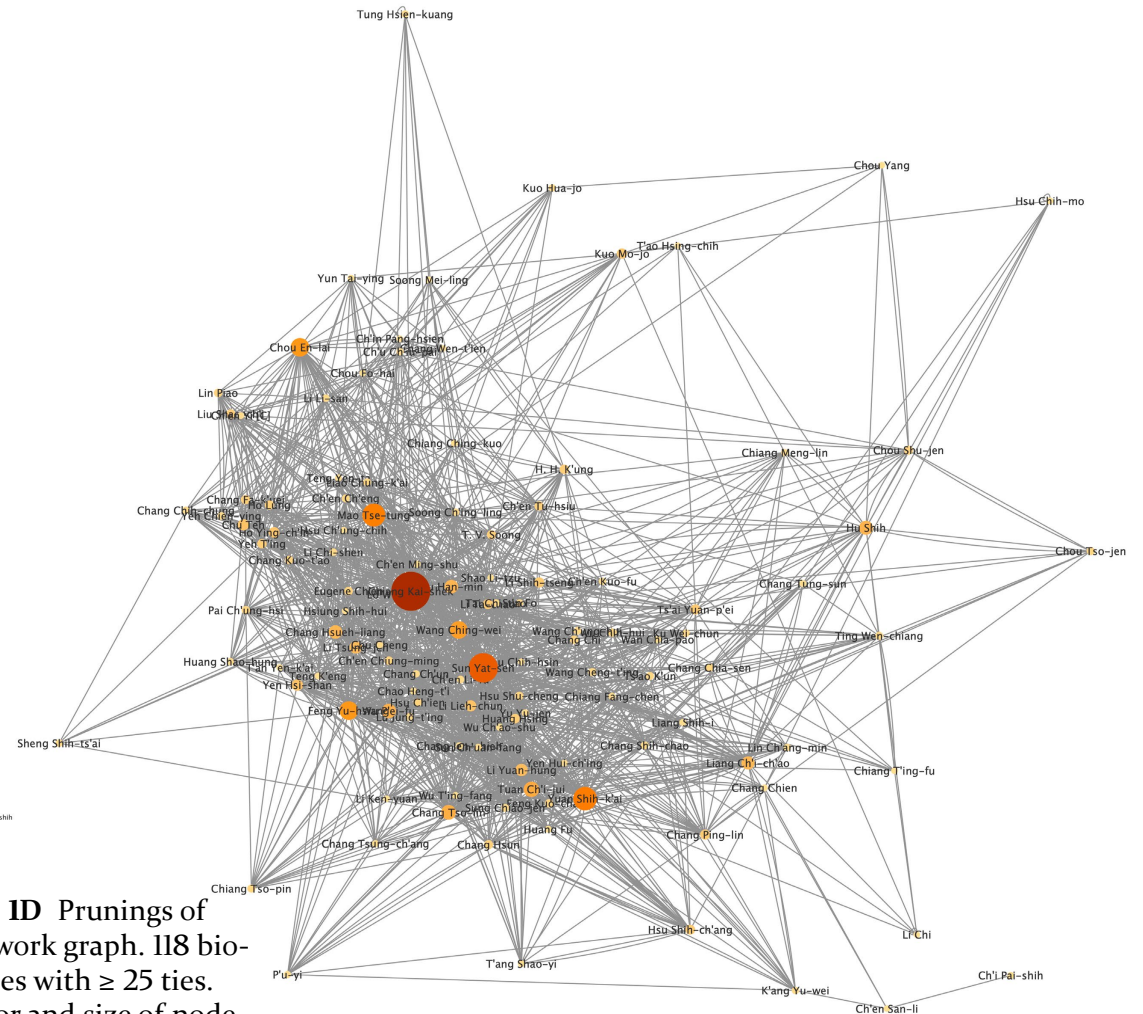
**Fig. 1A** Prunings of network graph. 574 bionodes with  $\geq 2$  ties. Color and size of node proportionate to their edge count.



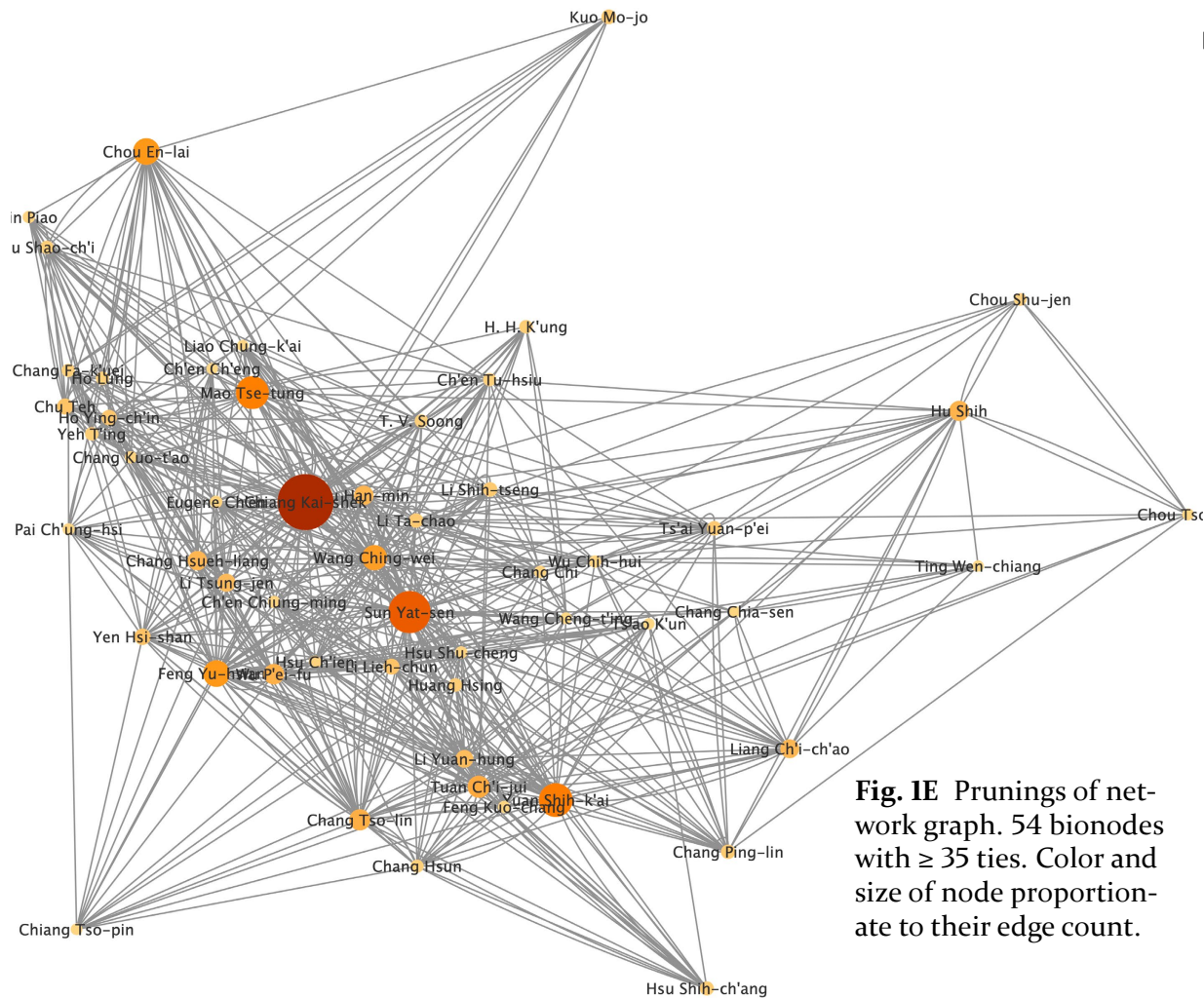
**Fig. 1B** Prunings of network graph. 360 bionodes with  $\geq 15$  ties. Color and size of node proportionate to their edge count.



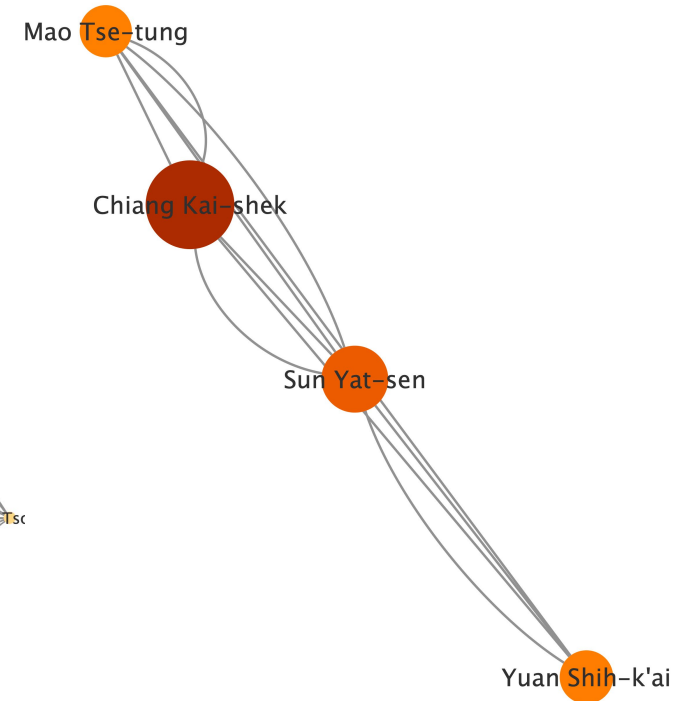
**Fig. 1C** Prunings of network graph. 205 bionodes with  $\geq 20$  ties. Color and size of node proportionate to their edge count.



**Fig. 1D** Prunings of network graph. 118 bio-nodes with  $\geq 25$  ties. Color and size of node proportionate to their edge count.



**Fig. 1E** Prunings of network graph. 54 bionodes with  $\geq 35$  ties. Color and size of node proportionate to their edge count.



**Fig. 1F** Prunings of network graph. 4 bionodes with  $\geq 200$  ties (Mao Zedong [Mao Tse-tung], Yuan Shih-kai [Yuan Shih-k'ai], Sun Zhongshan [Sun Yat-sen], and Jiang Jieshi [Chiang Kai-shek]). Color and size of node proportionate to their edge count.

important role in the biographies of a large number of other individuals. Based on this ratio, we identified three major profiles: (1) individuals with a relative balance between incoming and outgoing edges (ratio  $\approx 50\%$ ); (2) “source” figures with a greater number of outgoing edges (ratio  $< 50\%$ ), predominantly political or military leaders who controlled the chain of command at the top of institutions and were often the source of action or decision; and (3) “referential” figures with a greater number of incoming edges (ratio  $> 50\%$ ).

The referential figures are those who are mentioned far more frequently in other people’s biographies than other people are mentioned in theirs. For instance, Duan Qirui (400%), Wang Jingwei (371%), Yan Xishan (321%), Li Yuanhong (300%), and Liu Bochong (300%) were each mentioned three to four times more than they mentioned other individuals. In terms of an absolute number, the most frequently mentioned individuals were, in descending order, Jiang Jieshi, Sun Zhongshan, Yuan Shikai, Wang Jingwei, Feng Yuxiang, Mao Zedong, and Duan Qirui. As we elaborate later on, these individuals were either solicited for advice or mentioned as contextual references (that is, they are not mentioned as part of an actual relationship, but as an element of historical context). This is a central question that we discuss in the third section. It points to a major shortcoming of previous studies, which often assume that cooccurrences are the expression of social relationships. In the last section, we challenge this assumption through a close analysis of the nature of relationships, based on the computer-assisted annotations of biographies.

Edge count, however, fails to convey the importance of certain individuals, whose significant position in the network does not rely solely on the number of ties. Betweenness centrality offers an alternative way of measuring the importance of object-nodes, rather than just bionodes, who hold a central position in the network.<sup>14</sup> Although they have a relatively low number of neighbors, certain individuals play a structuring role as mediators between different parts of the global network. For example, Paul Pelliot, the French archaeologist, with only 4 ties, holds a connecting position that joins two peripheral branches to the main component. If we remove him, the main component loses 95 nodes and 1,488 links. Similarly, the Qing official Zeng Guofan presents another intriguing case of an object-node with just 14 edges, who nevertheless connects 544 nodes and 7,795 links. This is due to his connection to major bionodes such as Jiang Jieshi and Mao Zedong.

---

14 Betweenness centrality scores are particularly informative because they highlight the individuals who served as essential bridges (“brokers”) between individuals and communities. Technically, betweenness centrality measures the number of shortest paths that travel through a node.



Rank	Name	Name (Chinese)	Edge Count	Outdegree	Indegree	Ratio InD/ EdgC
1	Jiang Jieshi	蔣介石	367	125	242	66%
2	Sun Zhongshan	孫中山	269	83	186	69%
3	Yuan Shikai	袁世凱	208	69	139	67%
4	Mao Zedong	毛澤東	203	98	105	52%
5	Zhou Enlai	周恩來	158	92	66	42%
6	Feng Yuxiang	馮玉祥	158	50	108	68%
7	Wang Jingwei	汪精衛	148	32	116	78%
8	Duan Qirui	段祺瑞	126	24	102	81%
9	Zhang Zuolin	張作霖	119	46	73	61%
10	Wu Peifu	吳佩孚	114	37	77	68%
11	Hu Shi	胡適	112	38	74	66%
12	Hu Hanmin	胡漢民	107	48	59	55%
13	Zhang Xueliang	張學良	106	38	68	64%
14	Liang Qichao	梁啟超	103	28	75	73%
15	Li Zongren	李宗仁	97	37	60	62%
16	Li Yuanhong	黎元洪	95	24	71	75%
17	Yan Xishan	閻錫山	85	20	65	76%
18	Zhu De	朱德	83	41	42	51%
19	Li Liejun	李烈鈞	82	30	52	63%
20	He Yingqin	何應欽	78	36	42	54%
21	Guo Moruo	郭沫若	73	43	30	41%
22	Li Shiceng	李石曾	72	28	44	61%
23	Cai Yuanpei	蔡元培	72	26	46	64%
24	Zhang Binglin	章炳麟	71	39	32	45%
25	Li Dazhao	李大釗	71	25	46	65%

Note: Bionodes are ranked by edge count in descending order, with their respective indegree (number of incoming edges), outdegree (outgoing edges), and ratio indegree/edge count.

**Tab. 3** The 25 bionodes with an edge count above 70.

The list and range of bionodes with the highest scores of betweenness centrality remain very much aligned with the hierarchy defined by edge count. This group comprises 19 political, military, and intellectual figures who appear highly connected among themselves (123 edges). When their direct neighbors are included, these 19 individuals form a network of 870 nodes (including both the bionodes and the object-nodes) (26.7 percent of all nodes) and 6,226 edges (65.4% of all edges). The distinctive feature in the ranking by betweenness centrality is the emergence of a few prominent intellectuals such as Hu Shi, Liang Qichao, and Guo Moruo, who are placed 4<sup>th</sup>, 6<sup>th</sup>, and 8<sup>th</sup> respectively, well ahead of the major political and military figures who, by edge count, rank far above other nodes. It can be argued that the more versatile profiles of these three intellectuals, who had a foot in various circles, explain their position as eminent “brokers” in the BDRC.

The last method we apply to the bionode-only network is clustering. We seek to identify sub-communities of more densely connected bionodes. The algorithm (GLay) detected 23 communities with great variations in their size.<sup>15</sup> The largest cluster (cluster 4) comprises 208 bionodes and 2,279 ties, but the 14 smallest “clusters” each consist of just one bionode. The eight largest clusters are listed in Table 4. For each cluster, we report the number of nodes and ties it contains, list the most representative individuals, and give it a label that best describes its composition based on the names of its members. Some of these are clearly centered on major historical figures or clearly identified social groups. Cluster 4, for instance, revolves around famous military and political leaders (Jiang Jieshe, Sun Zhongshan, Yuan Shikai). Cluster 1 includes major intellectuals such as Hu Shi, Cai Yuanpei, and Liang Qichao. Cluster 3 groups together major Communist leaders (Mao Zedong, Zhou Enlai). Cluster 6 represents the business circle, whereas cluster 9 connects several prominent scientists (physician Wu Liande, biochemist Wu Xian). Other clusters, however, exhibit more complex patterns with less obvious rationale for their grouping (such as clusters 7 and 16). Little can be said about these communities based solely upon the names of their members. In the next section, we address the question of whether they coalesce on the basis of specific attributes.

To conclude the first section, our analysis points to the dual structure of the BDRC as a network of cooccurrences that features several small ego-networks isolated from the main component, but also a polycentric network (the main component) that represents a relatively well-connected group of biographies polarized

---

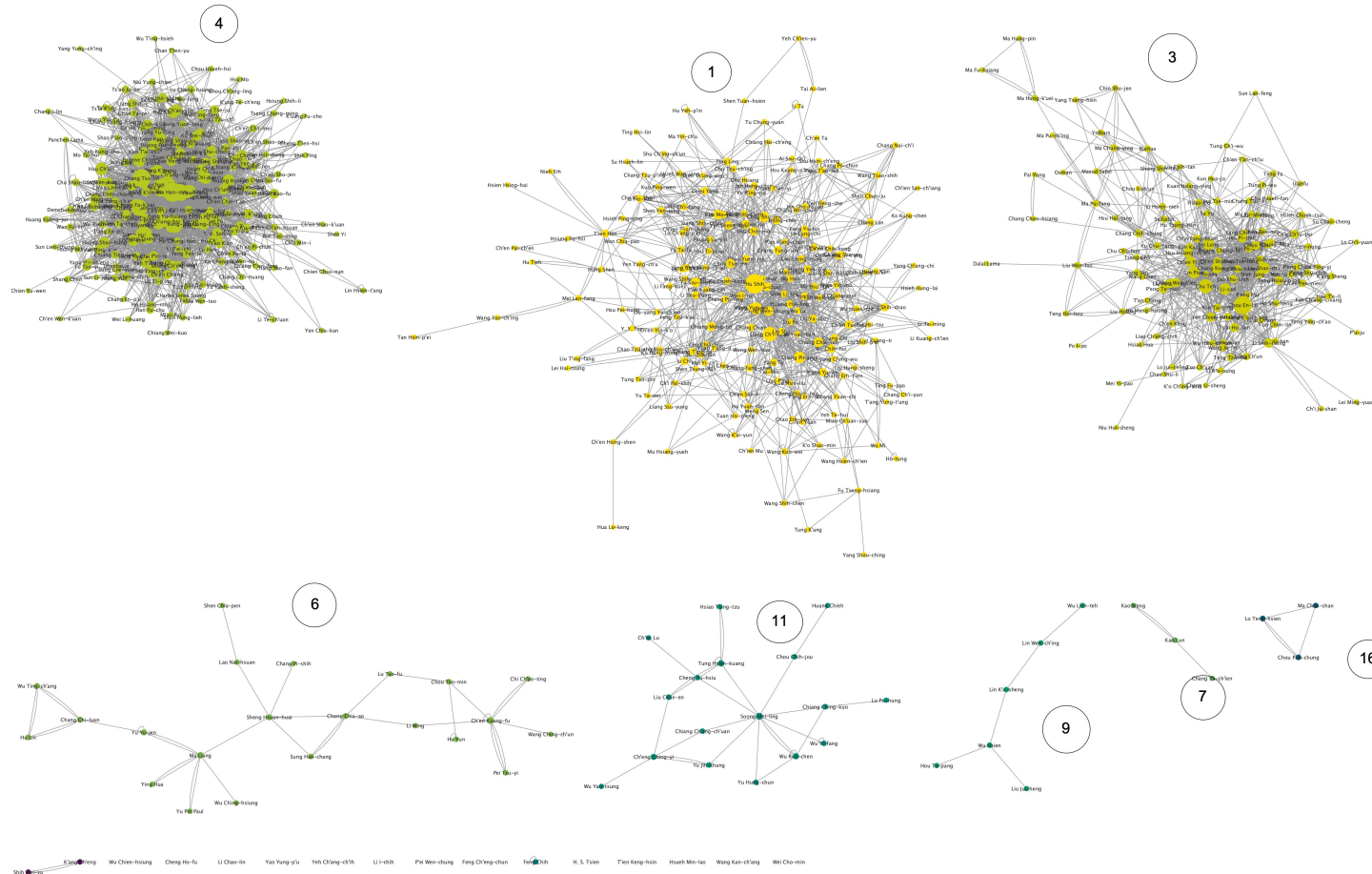
15 We used Cytoscape GLay algorithm (Fast-greedy), which relies on the greedy optimization of modularity score, with different corrections on edge density and cluster size. Previous studies have demonstrated its dramatic performance advantage in handling large networks. Gang Su et al., “GLay: Community Structure Analysis of Biological Networks,” *Bioinformatics* 26, no. 24 (December 15, 2010): 3135–37.

Cluster N°	Number of nodes	Number of ties	Representative individuals	Label
4	208	2279	Jiang Jieshi, Sun Zhongshan, Yuan Shikai, Wang Jingwei, Feng Yuxiang, Duan Qirui, Wu Peifu, Zhang Zuolin, Hu Hanmin, Zhang Xueliang, Li Zongren	Military
1	191	903	Hu Shi, Liang Qichao, Cai Yuanpei, Zhang Binglin, Zhou Shuren	Intellectuals
3	123	786	Mao Zedong, Zhou Enlai, Zhu De	CCP
6	22	37	Yu Youren, Zhang Jiaao, Sheng Xuanhuai, Chen Guangfu	Business
11	17	27	Song Meiling, Jiang Jingguo, Wu Guozhen	Song Meiling
9	6	5	Wu Xian, Wu Liande	Scientists
7	3	3	Gao Weng, Gao Lun	Artists
16	3	3	Zhou Baozhong, Ma Zhanshan, Luo Dengxian	Clique

*Note: For each cluster, the table reports the number of nodes and ties, the most representative individuals, and the label that best describes its composition, based on the names of its members.*

**Tab. 4** The eight largest clusters of bionodes detected by the GLay algorithm.

by a limited number of prominent figures. In brief, because of the considerable number of individuals mentioned in the BDRC, the cooccurrences of names eventually constitute a relatively well-connected network, with a giant component of interconnected individuals. The pruning method based on the edge count and an analysis of betweenness centrality have helped define distinct groups of individuals who are important to varying degrees and play different roles in maintaining the interconnectedness of the global structure. Moreover, beneath this massive “hair ball,” clustering analysis has also revealed subgroups of more densely connected individuals. Do the various subgroups that contribute to the global structure of the BDRC coalesce sociocentric networks based on specific attributes? This is the core question we address in the next section.



Note: Size of nodes is proportionate to degree centrality.

**Fig. 2** Clustered network of bionodes.



### 3. Networks of attributes

After we identified the persons in the BDRC, we used NLP techniques to retrieve a wide range of information related to these persons, such as institutions, positions, locations, events, etc. We propose to use this data as attributes to enrich the network of cooccurrences. Our analysis focuses on bionodes only, because while the BDRC provides information on the provincial origin, education, and other details of all the biographed individuals, such information is entirely absent for those mentioned in their biographies. Retrieving such information at this stage would require a huge amount of time, especially because for many Chinese people mentioned in the biographies of others, the BDRC gives only the initials of their given names.

In this section, we examine whether individuals who share common attributes tend to group together so as to form “sociocentric” networks in the BDRC.<sup>16</sup> We focus on five major attributes: provincial origin (well-studied by historians, and often recognized as being central to Chinese society<sup>17</sup>); military background (well represented in the BDRC<sup>18</sup>); education abroad (also a frequent feature in the BDRC population<sup>19</sup>); CCP affiliation (a self-contained, easily identifiable, and well-studied group<sup>20</sup>); and gender (women<sup>21</sup>). For each attribute, we built the network in two steps, first identifying all the bionodes with the selected attribute, and then building an extended network that includes these bionodes and their first neighbors. It is these extended networks that we study below.

- 
- 16 A “sociocentric network” is a network based on shared social attributes. This notion is borrowed from Tamper, Hyvönen, and Leskinen, “Visualizing and Analyzing Networks of Named Entities in Biographical Dictionaries for Digital Humanities Research.”
  - 17 William T Rowe, *Hankow: Commerce and Society in a Chinese City, 1796–1889* (Stanford, Calif.: Stanford University Press, 1984); Bryna Goodman, *Native Place, City, and Nation: Regional Networks and Identities in Shanghai, 1853–1937* (Berkeley: University of California Press, 1995); Richard Belsky, *Localities at the Center: Native Place, Space, and Power in Late Imperial Beijing* (Cambridge, Mass.: Harvard University Asia Center: Distributed by Harvard University Press, 2005).
  - 18 Diana Lary, *Region and Nation: The Kwangsi Clique in Chinese Politics, 1925–1937* (London; New York: Cambridge University Press, 1974); Jerome Ch'ên, *The Military-Gentry Coalition: China under the Warlords* (Toronto: University of Toronto-York University Joint Centre on Modern East Asia, 1979); Edward Allen McCord, *The Power of the Gun: The Emergence of Modern Chinese Warlordism* (Taipei: SMC Pub., 1997).
  - 19 Y. C. Wang, *Chinese Intellectuals and the West, 1872–1949* (Chapel Hill: University of North Carolina Press, 1966).
  - 20 Marilyn Levine, *The Found Generation: Chinese Communists in Europe during the Twenties*. (Seattle, Wash.: University of Washington, 1993); Steve Smith, *A Road Is Made: Communism in Shanghai 1920–1927* (Honolulu: University of Hawaii Press, 2018).
  - 21 Gail Hershat, *Women in China's Long Twentieth Century* (Berkeley: Global, Area, and International Archive: University of California Press, 2007); Barbara Mittler, Michael Hockx, and Joan Judge, eds., *A Space of Their Own: Women and the Periodical Press in China's Long Twentieth Century* (Cambridge: Cambridge University Press, 2018).

The hypothesis that provincial origin could provide the basis for specific networks did not generally pan out. For example, natives of Zhejiang are well represented in the BDRC, with 79 biographed individuals (13.9 percent of all the bionodes) who are connected to 434 other individuals in the extended network. Two isolated ego-networks surrounding two scientists (Qian Xuesen and Wang Ganchang) from Zhejiang have no direct or indirect connection with the other Zhejiang natives. Furthermore, almost all the provinces of China are represented in the main component (1,040 nodes and 6,084 edges) of the Zhejiang network, with Jiangsu (57) and Hunan (56) being the most represented provinces, followed by Guangdong (48). Due to their importance in the Guomindang elites, Guangdong natives apparently offered the prospect of a tighter-knit group. Their network includes 376 bionodes and 4,050 edges, with 67 Cantonese nodes that possess a total of 239 edges. As in the case of the Zhejiang natives, however, a wide range of 23 provinces are represented, and the Cantonese account for only 18 percent of the total. We also observed that the network was in fact composed of seven separate components, with six ego-networks built around specific personalities who had no link with each other.<sup>22</sup> In other words, there is little evidence of homophily by provincial origin among the Zhejiang and Guangdong natives in the BDRC.

Only one group based on the same provincial origin stands out in the BDRC: the natives of Hunan province, who form a large network of six components with 722 nodes and 4,908 edges. The presence of such prominent figures as Mao Zedong may have introduced a bias in terms of both provincial origin and political affiliation (CCP). Yet, even after removing Mao Zedong, the main component of the Hunan network still reveals a group of densely connected Hunanese whose pillars are also CCP members (Cai Hesen, Liu Shaoqi, Peng Dehuai, Peng Shuzhi, etc.). In this network, 72 of the bionodes are Hunanese, who have links to Zhejiang (47), Jiangsu (40), and Cantonese (37) natives.<sup>23</sup> Altogether, 21 provinces are represented. Yet the main feature is the considerable number of CCP members in this network (112 bionodes), of which the Hunan natives claim a substantial share (37). One could argue that the Hunanese clearly form a more homophilic network, which intersects with political affiliation.

The BDRC includes a large number of military figures. Individuals with any kind of military positions in their careers form a population of 466 bionodes with 4,035 edges. The size of this network is an indication of the high level of co-occurrences in the biographies of these individuals. Of these individuals, however,

---

22 The six individuals were Wei Zhuomin (religious leader), Hu Die (actress), Xu Guangping (female writer), Li Fanggui (linguist), Dai Ailian (female writer), and Luo Dengxian (labor activist).

23 On the place of Hunanese in CCP networks, see Levine, "Post WWI Chinese Revolutionary Leaders in Europe."

only a much smaller number (43 bionodes with 94 edges) received a military education or held continuous military positions. Taking only these forty-three individuals into account, this smaller network displays a high density (0.322 vs. 0.074 in the larger network), although five individuals constitute isolated components with few links each (Xie Bingying, Huang Kecheng, Sun Lanfeng, Sun Liren, He Zhonghan). The major broker in this group is Li Zongren, who has the highest degree and betweenness centrality. One level down (although almost at par) are three lesser-known figures: Liu Zhi, Xue Yue, and Yang Sen, who each form their own clusters. Only Xue Yue and Yang Sen are directly connected. In terms of indegree, Li Zongren still ranks first, followed by Duan Qirui and Bai Chongxi.

There is no clear evidence that graduation from the same academic institution was a strong connecting factor, except for the military officers who graduated from Japanese military academies. Graduates from Shinbun Gakko and Shikan Gakko, for instance, show a strong propensity to group together, but it needs to be established whether this is a sign of actual social relationships or an artifact of contextual mentions. The generational factor may reinforce the impact of cooccurrences. There is a clear overrepresentation of individuals born during the decade 1886–1896 who graduated mostly from the new military schools and academies established by the Qing, or from Japanese military academies between 1906 and 1920. This generational group includes 25 bionodes (58%) and 60 edges (67%) and accounts for all the main brokers in the military-only network.

The Communists form a very identifiable and self-contained set of individuals who are included in the BDRC exclusively because of their affiliation with the party (CCP). The network of CCP members includes 865 nodes and 5,391 edges. It is a very large network that reflects the share of CCP members in the BDRC (121 bionodes) and the large number of individuals tied to them (760 edges). The network of CCP members alone is made up of nine components, with eight isolated individuals, and has only four women. The outdegree distribution highlights 20 individuals with more than 30 neighbors. We can delineate four groups. The first, with an outdegree above 35, includes six distinct individuals (Mao Zedong, Zhou Enlai, Lin Biao, Zhu De, He Long, Li Dazhao). These six individuals actually connect almost every CCP member (93/99). At the next level (outdegree between 25 and 34), we find a second group of seven tightly knit individuals (Liu Shaoqi, Ye Ting, Li Lisan, Zhang Guotao, Li Jishen, Chen Duxiu, Guo Moruo). Taken together, these thirteen individuals (Guo Moruo is an outlier) form the backbone of the CCP network in the BDRC. Betweenness centrality reveals a limited number of mediators (10), yet with large discrepancies between them. The entire network is clearly centered around Mao Zedong, who serves as the main broker (0.3), followed by Zhou Enlai (0.12), Zhou Yang (0.06), Ye Ting (0.05), and Lin Biao (0.04). Within the CCP network, as discussed above, the Hunanese lead the pack with 65 individuals, followed by natives of Zhejiang (52), Guangdong (44), Jiangsu (44), Hebei (27), Hubei (21), and Jiangxi (18). The CCP net-

work includes 13 military officers, but these CCP military figures do not form a specific community.

Most of the biographed characters in the BDRC received a high level of education. 519 (88 percent) received a college degree. Of these, many had the opportunity to study abroad, which places them in the specific category of “returned students.” These returned students are commonly grouped according to where they studied (country, university). In the BDRC, they form a population of 200 individuals who attended and graduated from a total of 343 different academic programs. Since the returned students established alumni associations or held events in China that brought together those who had studied in the same country or region, one can hypothesize that networks may have been built on this basis.<sup>24</sup>

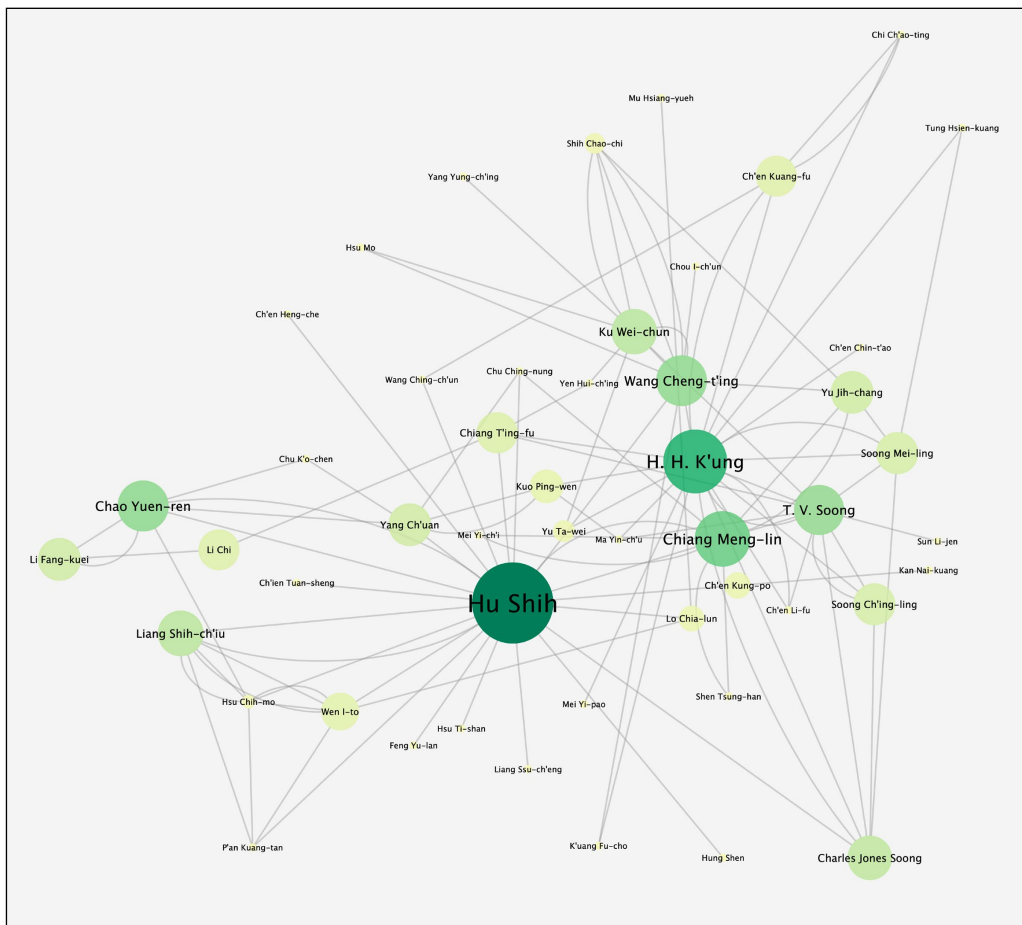
The United States ranks first in the BDRC with 70 returned students, followed closely by Japan with 67 individuals. Europe received the next largest batch, although the returned students from Europe were distributed across several countries: United Kingdom (24), Germany (14), France (15), the Soviet Union (8), and a number of other countries (6). We built networks based on the country of study to examine to what extent the returned students from a given country had a propensity to connect with each other. Each such network includes bionodes who were not educated in the country of reference. For example, for the American-trained Chinese, 43 percent of their neighbors were not trained in the United States. Previous works have demonstrated their propensity to mingle with diverse communities, which corroborates our observation.<sup>25</sup> We found the same ratio among Europe returnees. It is only among the Japan-returned students that the ratio was slightly lower, possibly indicating a greater degree of homophily. Yet, as we discuss below, other factors may explain this higher level of homogeneity.

The American-trained students form one of the most interesting networks. In fact, it can be read as a microcosm of the entire BDRC global network. A fair number of individuals (14) are not mentioned in any of the biographies of their peers. Their networks do not intersect with the individuals in the main component, nor with the dyads formed by another five individuals. The 51 bionodes in the main component, however, tend to exhibit a higher level of connectedness.

---

24 Stacey Bieler, *“Patriots” or “Traitors”? A History of American-Educated Chinese Students* (New York: Routledge, 2003); Liu Xiaoqin 劉曉琴, “Minguo liu Mei shetuan yu liu Mei sheng de shehui wangluo – yi Chengzhihui yu Zhang Boling de fenxi wei zhongxin” 民國留美社團與留美生的社會網絡 – 以成志會與張伯苓的分析為中心 [Chinese Students’ Clubs and Social Networks in the United States: A Study of Zhang Boling and the ‘Cross and Sword’ Society], *Huaqiao Huaren Lishi Yanjiu*, no. 4 (2019): 88–95.

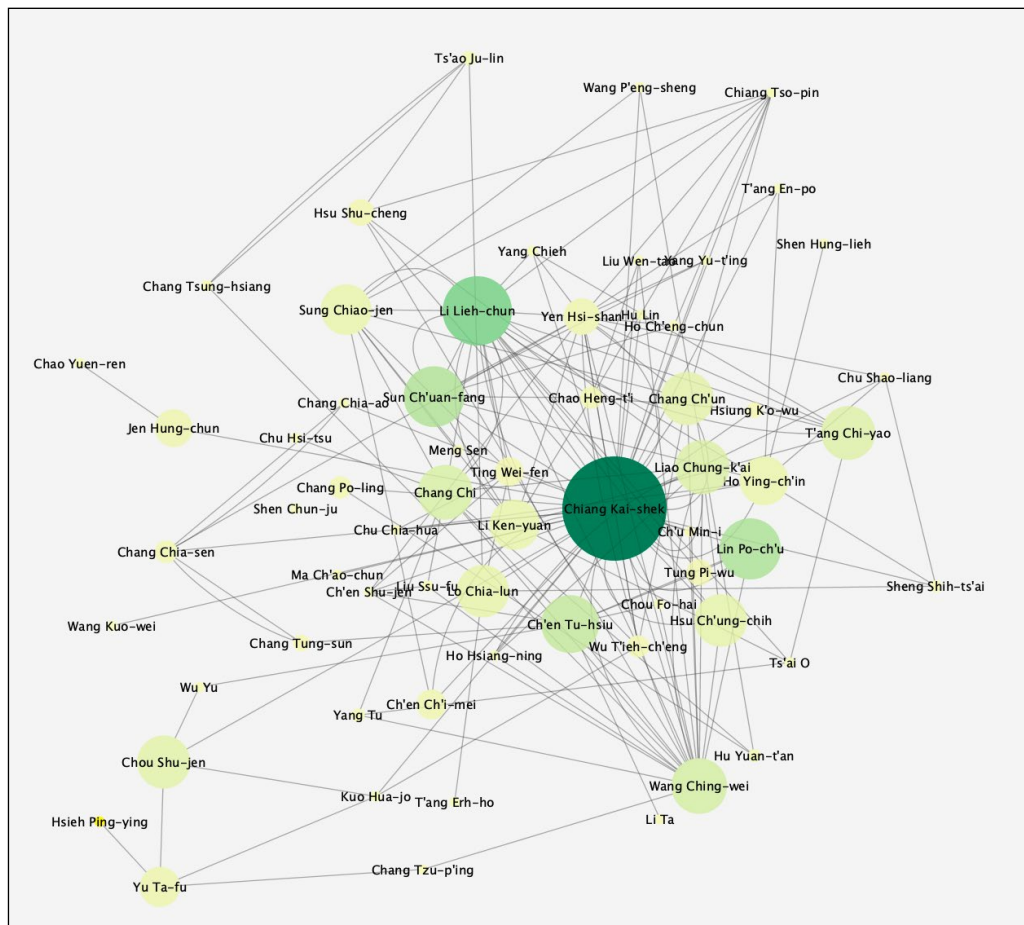
25 Cécile Armand, “Foreign Clubs with Chinese Flavor: The Rotary Club of Shanghai and the Politics of Language,” in *Knowledge, Power, and Networks: Elites in Transition in Modern China*, ed. Cécile Armand, Christian Henriot, and Huei-min Sun (Leiden: Brill, forthcoming).



Note: Size of nodes is proportionate to betweenness centrality.

**Fig. 3A** Network of American-returned students (main component).

Hu Shi, one of China's leading intellectuals, serves as the main broker in this network (betweenness centrality = 0.15). He is connected to 21 peers through a total of 37 edges. His peers include mostly intellectuals, but very few political figures. He is not linked to the second most important broker, Kong Xiangxi, an eminent figure in both the business and political worlds. Kong is essentially connected to political figures, including his family relations (Song Ziwen, the Song sisters, and their father). The only intellectuals in his network are scholars with an administrative role (Guo Bingwen, Ma Yinchu). Two other figures also play an important role in the network of American returnees, each in a different register: Song Ziwen, with a profile quite similar to his brother-in-law Kong Xiangxi; and Jiang Menglin, a multi-faceted intellectual bridging the worlds of education, culture, and politics. While an exclusive pattern of homophily based on the country of education cannot be established among the American-trained students, their



Note: Size of nodes is proportionate to betweenness centrality.

**Fig. 3B** Network of Japan-trained students (main component).

network suggests that they shared a common cultural, educational and linguistic background that may have served to establish professional and political relationships during the course of their lives.

The network of Japan-trained students presents a very different structure. A striking but obvious feature is the centrality of Jiang Jieshi (by degree and betweenness measures). He connects almost all the individuals in the Japan network, and more importantly, within this network, those who received military training. If we remove Jiang from the network, two other figures emerge: Li Liejun and Wang Jingwei, each at the center of a substantially different network. Wang is mostly connected to political figures, with very few military leaders. Li, by contrast, reaches out to all the military leaders. If we enlarge the scope of observation to include all the bionodes connected to the Japan-trained students, we

find a network largely made up of military figures trained in China or elsewhere. In other words, the single most important factor in the Japan-trained individuals is less the country where they studied than the military education they received there, which set them on a career path that connected them to a wider circle of military figures. If we compare this with the American returnees, one could say that the degree of heterophily based on the country of education is higher among the Japan returnees than their American counterparts.

Very few women were selected for a biography in the BDRC. Altogether, they account for only 25 of the 589 biographies. What compelled the editors to select these women? Was it for their own profile and intrinsic importance, or because they were related to prominent men?<sup>26</sup> How do these women fit into the network structure of the BDRC? How do they contribute to male-dominated specific communities? In other words, can we identify a women's network in the BDRC?

The network of women includes 189 nodes and 990 edges, with four components: one main component and three ego or bi-ego networks. The bi-ego networks are those previously identified in the global analysis of the BDRC network of cooccurrences, namely the first two women physicians trained in the United States who graduated in 1896. Their network is composed exclusively of foreigners – which reflects the fact that Boorman focused on their period of education and training before they returned to China. The other small component revolves around Wu Jianxiong, a female physicist whose career was mostly outside China, who also has a large number of foreigners in her network.

Women thus do not form any cohesive network. Their limited number may be part of the explanation for the lack of more obvious networking. Five of them stand alone, and four women form two distinct pairs. The main component is made up of a core of a very small number of highly connected women – Song Qingling, Ding Ling (writer), and Song Meiling. Except for Ding Ling, who stands apart, Song Qingling and Song Meiling – two sisters from an influential family and among the most prominent and politically active women of Republican China – are interconnected at the same level. Their marriage to men of prominence (Sun Zhongshan and Jiang Jieshi) placed them within a larger network that included many prominent figures of the Republican period. Yet, even after removing the three main male figures (Mao, Jiang, and Sun), the centrality of the three women remains the same. However, they fail to connect directly with any significant number of women, and do not even connect with each other. In the case of Song Meiling, it is through He Xiangning, the wife/widow of Liao Zhongkai (d. 1925), that she connects with Chen Bijun (wife of Wang Jingwei)

---

26 Henrike Rudolph, "Structures of Empowerment: A Network Exploration of Women Activists' Collective Biographies in Twentieth-Century China."

and Deng Yingchao (wife of Zhou Enlai). The remaining group of five women are even more tenuously connected.

Ding Ling's network branches out in two main directions: CCP members, including the men of letters in the party (Zhou Yang, Hu Feng) and literary figures (two other women writers, [Xie Wanying and Su Xuelin] and three male writers [Ye Shengcao, Lu Xun, Cao Yu]). Song Meiling's network is made up of powerful men that include all her direct and indirect next-of-kin (father, husband, brother-in-law, etc.), as well as military and political figures who served her husband or her more directly (Yu Hongjun, Wu Guozhen). There is no CCP figure in her network. Song Qingling, the older sister, presents a similar profile in terms of next-of-kin relations, although her network branches out to both Guomindang and CCP figures, which quite accurately reflects her positioning in Republican politics, and in the People's Republic of China when she became the willing pawn of the CCP. In brief, the analysis of the main component reveals four main profiles of women that may have served as a guide for including them in the BDRC: scientists, artists, writers, and political women. Political women appear in the BDRC for their own merit, but also due to their marriage with important political figures in Republican China, whereas intellectuals such as Ding Ling appear only due to their own merit.

To conclude this section, the study of attribute-based networks reveals that neither provincial origin, education, nor any single attribute alone is sufficient to constitute significant sub-communities in the BDRC. It is only the combination of multiple attributes that can account for the most densely connected clusters of biographies in the BDRC. For instance, this paper has reasserted, in line with previous scholarship, the effect of Hunanese origin combined with CCP affiliation, that of study-abroad experience in Japan combined with military training, and the combination of marriage, political influence, but also professional skills in the case of women. The patterns delineated in the study of attribute-based networks tend to support the hypothesis that there is more to the cooccurrence of names than the aggregated mentions of individuals in the various biographies. The repeated mentions suggest the existence of actual relationships. The network of cooccurrences, however, does not permit us to fully ascertain this. A more in-depth examination of the nature of the relationships is needed. This is the purpose of the next section.

#### 4. Cooccurrences or relations?

Our analysis in the previous sections suggests the existence of two major categories of mentions associated with two dominant types of links: textual/contextual references, and actual social contacts. However, it has proved difficult, if not impossible, to firmly establish the difference so long as we remain focused on the network of cooccurrences. In the first section, we made the hypothesis that some



individuals may be mentioned as elements of historical context, or were a source of inspiration for the biographed characters. We also pointed to historical characters from the past, such as Adam Smith or Zhuangzi, who could not in reality have met the individuals biographed in the BDRC. In the particular case of “referential” figures with high indegree centrality, we highlighted that these individuals were most often sought for advice or served as contextual references. A close reading of their biographies further reinforces this impression. These observations led us to question the more general assumption that the cooccurrence of names in a biography could be systematically considered as the expression of a genuine social relationship.

In this section, we move a step further and explore in greater depth the nature of the links between individuals in a selected sample of 36 biographies that we annotated manually (see Appendix). Our approach is to examine the actual grounds for the mention of a name in a given biography and qualify the relation between the named individuals. Our ultimate goal is to build networks from these annotations that better reflect historical relationships, and not just textual cooccurrences. This section follows three steps. First, we present the method for annotating the relations in the biographies. Second, we analyze the results statistically. Third, we build and analyze various networks based on the (most significant) extracted annotations, which we compare with the attribute-based networks discussed in section 2.

The first challenge was to create a “representative” sample to annotate, for which we relied on two criteria. First, we selected the individuals on the basis of the edge count. A high edge count was a sign that these individuals were involved in the widest and richest range of possible relations. Second, we refined the sample to include the greatest possible variety of individual profiles in order to correct the bias produced by relying solely on the edge count. The selected biographies represent only 6 percent of the total number of biographies, but 18 percent of the total number of words in the BDRC.

In each biography, we focused only on the relations involving the biographed individual. For instance, in the biography of A, we annotated the relations between A and B and between A and C, but discarded any potential relation between B and C. The selection of qualifying terms was based on the terms identified through close reading. We ensured that the manual annotations reflect only the language and the terms used in the text without adding any layer of interpretation or external knowledge, because the model for automatic annotations would ultimately rely only on the text itself and the particular combinations of words through which different relations are expressed. As shown in Table 5, we classified the relationships into twelve categories: acquaintance, protégé, friendship, liaison, kinship, *tongxiang* (native to the same place), education (master/disciple, co-disciple), professional, political, military (military conflict), indirect (no direct relation: context, third-party reference, etc.), and neutral (uncatego-

rized mentions). We were aware that our categories represented a wide range of choices, but previous experiments had convinced us that a narrow range of terms might produce results that were too broad to analyse.

For instance, a previous study based on a similar corpus – a biographical dictionary of Dutch socialists – had also attempted to qualify the relations between individuals through manual and automatic annotations. All the cases of cooccurrences, however, were considered as meaningful relations. Relying on sentiment analysis, the authors chose to classify the relations into three main types: positive, antagonistic, and neutral. They eventually found that the trilogy was too reductive, especially the neutral one that regrouped too many cases to make it a significant marker.<sup>27</sup> This categorization may have been relevant for relationships within a coherent and like-minded group, but the BDRC presented us with a wider array of highly distinct profiles. In our case, we were interested in qualifying the relationships among a richer set of terms based on the words used in the text itself to describe the relationships. Some categories did not really pan out, due to the limited number of such relationships (*liaison*, *protégé*, *tongxiang*). Yet the process provides a preliminary schema in the form of pre-determined categories that would eventually be used in the model for automatic annotations.

For the annotation workflow, we relied on InCeption, a machine-assisted interactive annotation platform.<sup>28</sup> Each biography was annotated manually by a pair of annotators who worked independently, and we then curated together their respective biographies.<sup>29</sup> There were significant variations in the annotations, mostly due to the inevitable propensity to “interpret” based on prior knowledge of the individuals mentioned in the biography and the difficulty to disentangle multifaceted relationships. Eventually, all the biographies passed to a single curator, who homogenized the annotations. These manual annotations were used to train a model for expanding the annotations automatically to the entire corpus in the future.

The annotation workflow produced a total of 3,227 annotated relations. As shown in Table 6, the three most frequent types of relations represent 77 percent of the total. These are political relations (34 percent), indirect relations (26 per-

---

27 Matje van de Camp and Antal van den Bosch, “A Link to the Past: Constructing Historical Social Networks,” in *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, WASSA '11 (Stroudsburg, PA, USA: Association for Computational Linguistics, 2011), 61–69.

28 Jan-Christoph Klie et al., “The INCEPTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation,” n.d., 5. See also: <https://inception-project.github.io/>

29 The annotators included Cécile Armand (ENP-China, Aix-Marseille University), Guo Wei-ting (ENP-China, Aix-Marseille University), Christian Henriot (Aix-Marseille University), Jiang Jie (Shanghai Normal University), David Serfass (Inalco), and Sun Hueimin (IMH, Academia Sinica).

Term	Definition
Acquaintance	Initial meeting, casual encounter
Education	Relations of master/disciple, co-disciple
Friendship	Genuine friendship relation
Indirect	Absence of a relation: contextual mention, third-party reference
Kinship	Relations based on ties of blood, marriage, or adoption
Liaison	Pre-marital and extra-marital relations
Military	Relations in a military conflict: allies, enemies, battles, etc.
Neutral	Undefined mentions: trip together, getting together, etc.
Political	Relations in a political context/organization
Professional	Relations between colleagues or superior/subordinates in a professional context
Protégé	Relation of protection by mentors in education, politics, work
Tongxiang	Same native place

**Tab. 5** Types and definitions of annotations in the BDRC.

cent), and professional relations (17 percent). Although military and kinship relations garner a good number of annotations, they represent only 7 percent and 6 percent of the total, respectively. All the other categories, including that of *tongxiang* (same native place), failed to produce a significant number of annotations. They were thus not used to train our model. The high percentage of the top three types of relations indicates that the contributors who wrote the biographies centered primarily on the work and public life of the individuals, especially their professional activity. More importantly, although a significant share (one quarter) of annotations points to textual cooccurrences (indirect relations), these do not prevail in the BDRC. The remaining 75% are the expressions of actual social relationships between historical actors.

We ran a Principal Component Analysis (PCA) on the data extracted from the annotated biographies. Through PCA, we sought to construct *relational profiles* based on the most frequent correlations of specific relationships.<sup>30</sup> For this PCA, we retained as active variables only the most frequent categories of relations: political (1094, 34%), indirect (823, 26%), professional (563, 17%), and military (239, 7%) and considered the minor relations as supplementary variables. The

30 To conduct our PCA, we relied on the package “Factominer” in R Studio: <http://factominer.free.fr/factomethods/hierarchical-clustering-on-principal-components.html>

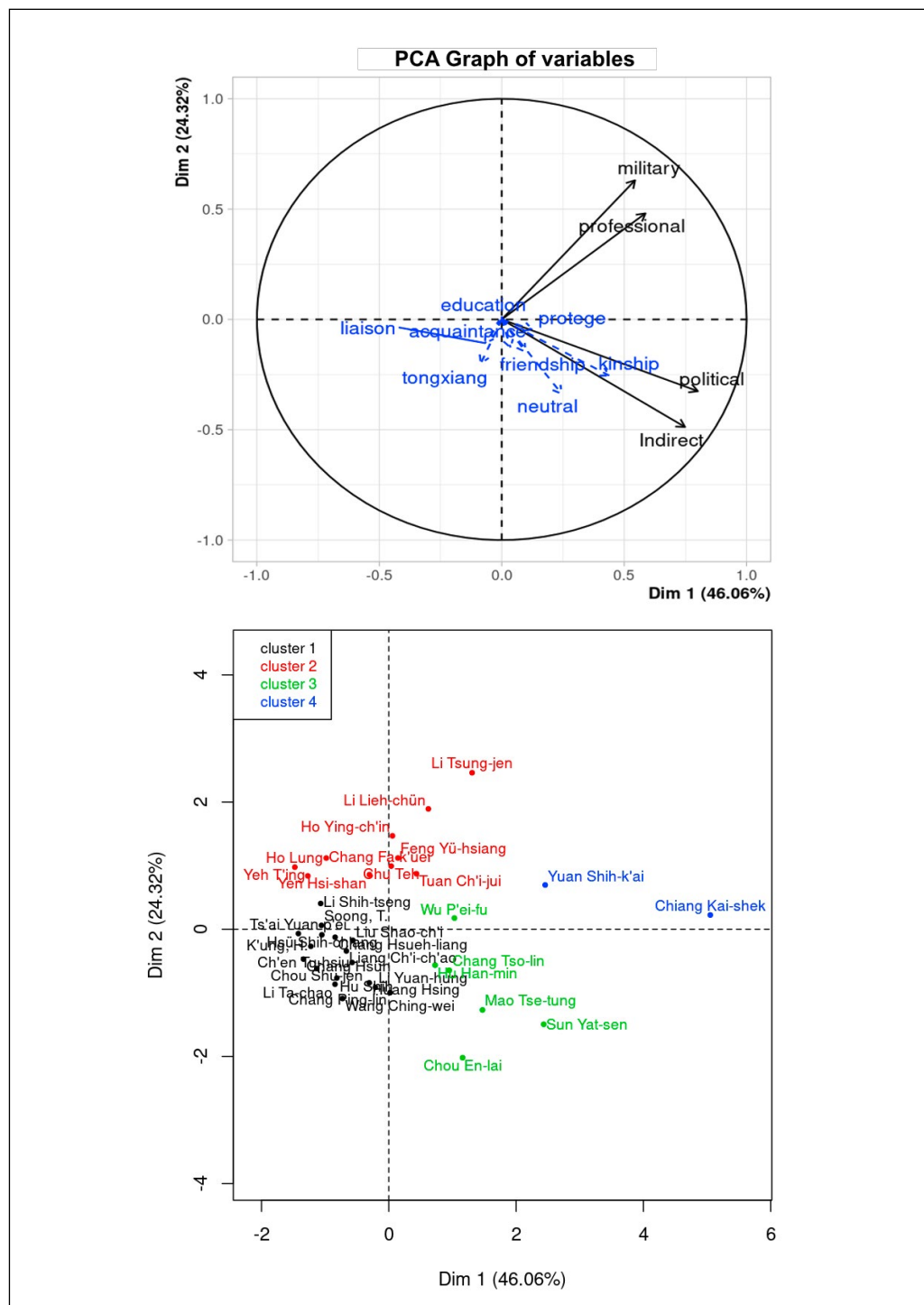
Type	Number	Percentage
Political	1094	33,9%
Indirect	823	25,5%
Professional	563	17,5%
Military	239	7,4%
Kinship	201	6,2%
Friendship	71	2,2%
Education	67	2,1%
Neutral	67	2,1%
Acquaintance	49	1,5%
Protégé	36	1,1%
Liaison	11	0,3%
Tongxiang	5	0,2%
Total	3226	100%

**Tab. 6** Distribution of annotated relations in a sample of 36 biographies from the BDRC.

PCA graph plots all 35 biographed individuals and their relations in two-dimensional space.<sup>31</sup> We selected the first two dimensions, as they best explain the variance among individuals and capture more than 70% of the information (46% and 24% on each dimension). The graph clearly contrasts individuals with many indirect and political relations on the right and those with few such relations on the left. The second dimension, which is primarily determined by military and professional relations, separates individuals with strong military and professional relations above the x axis from those with few such relations below it.

Based on the PCA, we performed hierarchical clustering on all four dimensions. We observed that professional relations contribute the most to the partition at large (0.693), followed by indirect (0.664), military (0.633), and political (0.506) relations. The algorithm grouped individuals into four clusters based on their relational characteristics. In figure 3 below, each individual is color-coded by cluster. Cluster 4, on the right of the graph, isolates two “big names” – Jiang Jieshi and Yuan Shikai – from all the others. What sets these two apart is the

31 We chose to remove Guo Moruo, whose profile was too specific and too different from the rest of the sample.



**Fig. 4** PCA analysis of the 35 annotated biographies: A. Graph of variables. Note: Black arrows: active variables; blue arrows: supplementary variables. B. Hierarchical clustering of individuals.

wealth of professional relations (4.52), the weight of indirect relations (2.88), and finally, the range of political relations (2.41). What dominates cluster 2 is the conjunction of strong professional relations (individuals who held a variety of positions in the army, through which they came into contact with each other) and military confrontations, as allies or as enemies. This cluster brings together almost all the military leaders, both nationalists and communists, except for three warlords (Zhang Xueliang, Wu Peifu, and Zhang Zuolin). Individuals in cluster 3 are characterized by few military and professional relations and a stronger weight of political and indirect relations. This cluster includes Mao Zedong, Hu Hanmin, Zhou Enlai, and Sun Zhongshan. Wu Peifu and Zhang Zuolin, two major warlords of the post-Yuan Shikai era, also belong to this cluster, which suggests their respective biographies defined them much more by their political relations than by their professional or military ones.

In the last step, we built a series of networks based on annotated relations. We constructed one network for each of the most significant types of relationships. We expected annotated relations to determine with greater confidence the reality of a relationship, and to more precisely delineate the nature of the links between the biographies and between the individuals. The annotated relations may not fundamentally alter the overall structure of the co-occurrence networks, nor the dominant position of high-profile individuals such as Jiang Jieshi or Yuan Shikai – we could only compare them with the co-occurrence network of 36 annotated biographies – but they do substantiate the hypothesis that a relationship could, in fact, reveal different configurations of proximities and interactions.

The category of “indirect relations” lumps together different types of relations (contextual mention, source of inspiration, connection through a third party, etc.). In future analyses, we will refine this category to separate the purely contextual relations from the other types of relations, and will adjust the annotation workflow accordingly. Nonetheless, the network of indirect relations confirms the weight of Sun Zhongshan, Mao Zedong, and Jiang Jieshi (in descending order of betweenness centrality) in purely contextual mentions. In this function, they have no direct relationship with the person in whose biography they appear. Most of these mentions come under formulas such as “When A came to power,” “At the time of A’s death,” “Under A’s regime,” etc. Moreover, the clustering of this network produced very interesting subgroups centered on specific individuals who shared common characteristics. One of these clusters is dominated by Jiang Jieshi and Sun Zhongshan, a second by the northern military leaders *cum* warlords, and another by CCP leaders, whereas two less densely connected groups revolved around intellectual figures (Hu Shi, Li Shizeng, Li Dazhao) and late Qing revolutionary activists (Huang Xing, Liang Qichao, Zhang Binlin). These indirect relations, even if they do not denote an actual social relationship, provide a sort of “index” of relevance in terms of contextual mentions.

Political relations are the most prominent feature in the connection between individuals. They neatly delineate two separate worlds: CCP leaders on one side and all the other main historical figures on the other. They highlight the centrality of Jiang Jieshi and Sun Zhongshan (betweenness centrality) in the whole network, while Mao Zedong and Zhou Enlai are significant only within the CCP sub-network. None of them reach out very much into non-communist circles, except Zhou Enlai. This is mainly due to Zhou's later career as premier of the People's Republic, which considerably extended his contacts internationally. In contrast, Li Dazhao, more of a secondary figure, had a broad network of relations across political lines even though he was executed in 1927 at age 38. Yuan Shikai also built an extensive and highly diverse network of political relations with political and mostly non-military figures, including a good number of Qing officials, but also major opponents such as Huang Xing, Song Jiaoren, or allies-turned opponents like Liang Qichao and, of course, Sun Zhongshan.

The exploration of professional relations redraws the previous configurations and leads to a very different network structure. Jiang Jieshi and Yuan Shikai are the two most central figures in this network. By placing emphasis on their professional relations, the resulting network gives more weight to their careers in government and the army. Their respective networks, however, diverge significantly. Yuan Shikai is connected to all the military leaders of the early republican period, except Wu Peifu, and his protégés were numerous. His professional relations include only a few political or intellectual figures such as Sun Zhongshan, Hu Hanmin, Wang Jingwei, or Cai Yuanpei. Yuan shares these figures with Jiang Jieshi's professional network. They are the connecting points, albeit indirectly, between Jiang and Yuan. Jiang's network includes two major military figures on the nationalist side – He Yingqin and Zhang Fakui – and a host of less central individuals. What distinguishes the network based on professional relations is the greater diversity one can see in the multiple sub-networks built around individuals (Hu Shi, Wu Peifu, Zhang Xueliang) who are only remotely connected to Jiang and Yuan. This observation also holds true for CCP leaders. Mao Zedong's professional relations place him as a secondary figure in the network, and he connects mostly to the triad formed by Zhou Enlai, Liu Shaoqi, and Zhu De. Yet again, this reflects the nature of post-1949 relations.

Military relations provide a good case to compare the networks of cooccurrences and annotations. We compared the network based on annotated military relations with the corresponding network of cooccurrences (based on military attributes) that we built in section 2, which we then filtered down to the 36 annotated biographies. In the network based on annotations, the number of nodes and edges decreases greatly, from 227 nodes and 811 edges to 101 nodes and 225 edges (Table 7). This is mostly due to the fact that we have focused only on the relations involving the bionodes. More significantly, the lower number of edges and the lower clustering coefficient suggest that military operations played only a limited part in the biographies of military elites. The BDRC instead emphasizes their in-

	Co-occurrences	Annotations
Nodes	227	101
Edges	811	225
Clustering Coefficient	0,339	0,074
Average Number of Neighbors	6,211	2,911

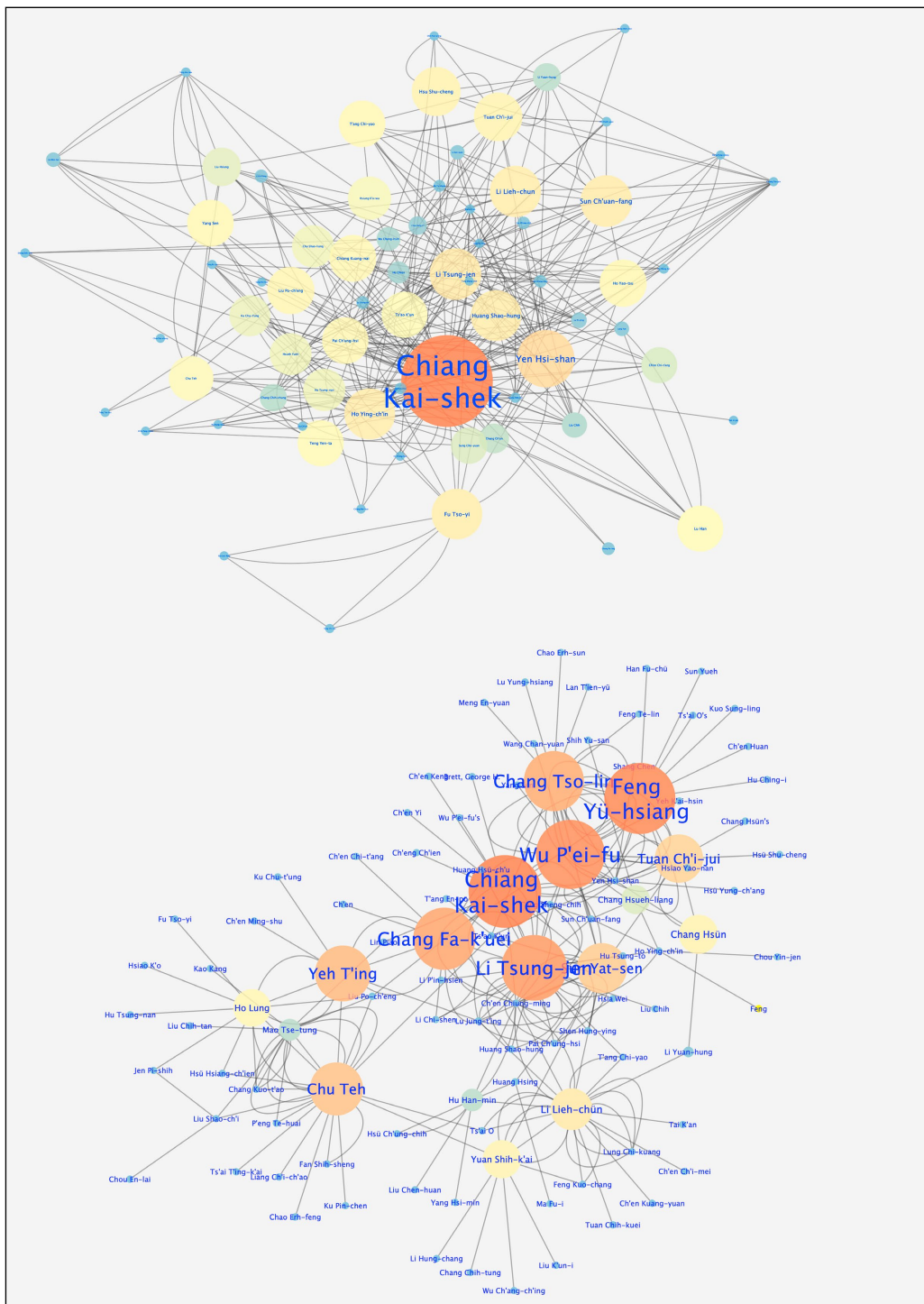
**Tab. 7** Comparative analysis, military networks (global metrics).

volvement in a broader range of relations. The power and prestige resulted less from their military deeds – victories or defeats on the battlefield – than from other sources of influence (political negotiations, professional contacts, recommendations).

Who are the most important players in these two networks? In the network of cooccurrences based on military attributes, betweenness centrality places Jiang Jieshi as the dominant broker, connecting a host of second-rank military and non-military actors (including Sun Zhongshan). In the network of annotations, Jiang remains central, but at the same level with other military leaders. Moreover, non-military actors are relegated to a secondary position. Rather interestingly, Yuan Shikai becomes a minor, more marginal broker with a very limited range of military relations.

To conclude this section, the analysis based on annotated relations in biographies reveals that individual mentions in the BDRC are not just cooccurrences – i.e., names connected through textual links – but also refer to historical actors who actually came into contact in the course of their life. There is, however, a substantial number of cases of indirect relations (25% of all annotations) and even of dropped names that argue for the need to exercise greater caution when considering any cooccurrence as the expression of an actual historical relation. On the other hand, the BDRC features a great variety of actual links within and between the biographies. The relational patterns we delineated through PCA and SNA also demonstrated the intermingling of multiple relationships among individuals, which complicates the previous typologies based on sentiment analysis. Clearly, annotations provide a necessary and efficient way to add historical substance to the analysis of networks simply based on cooccurrences.





**Fig. 5** Military networks: A. Based on attributes; B. Based on annotated relations (sample of 36 biographies). Note: Size of nodes is proportionate to betweenness centrality.

## 5. Concluding Remarks

Reading the BDRC through the lens of social network analysis may seem like putting old wine in a new bottle. The major biases of this work have been established in previous academic reviews.<sup>32</sup> Our contribution is not to reassert these biases in terms of content (due to problems of sampling and representativeness), but rather to uncover the underlying structure of the book, namely the hidden relations between individual biographies and between the individuals therein. This allows us to: (1) extract a “collective portrait” of the entire population based on their individual characteristics (attributes); (2) assess whether and to what extent cooccurrences were constitutive of networks; and (3) propose an approach that defines and qualifies relationships much more accurately.

We demonstrated that in the BDRC, political and professional relations far outweigh the “three sames” (native place, education, trade/business) commonly accepted in the historical literature. This goes against the grain, but we argue that this is not due solely to the nature of the elites selected in the BDRC.<sup>33</sup> The persistence of these types of relations, especially native place ties, was undeniable in Republican China. Chinese society, however, increasingly departed from the patterns studied for late imperial China. There was a flurry of new types of social organization that offered the possibility for individuals to become involved in multiple groups and networks. Political parties are a prime example of a completely novel type of organization, but professional or cultural associations also provided numerous arenas based on non-partisan grounds. While it is difficult to escape the conventional categorization (politician, merchant, military, etc.) that historians use to define elites – some individuals do fit in such categories – the relational profiles we have revealed challenge the relevance of such narrow categorizations for the Republican elites. The complex web of relations in which the individuals in the BDRC were enmeshed cut across such categories, and their careers often followed more than one path, sometimes in parallel.

From a methodological perspective, we have also demonstrated that much original knowledge can be gained from biographies through this approach. Network analysis reconnects the individualized and self-contained biographies and open pathways through the BDRC mosaic. While it falls short of creating a global narrative, it does revisit the nature and function of a biographical dic-

- 
- 32 J. K. Fairbank, review of *Biographical Dictionary of Republican China*, Volume I, by Howard L. Boorman and Richard C. Howard, *The American Historical Review* 73, no. 2 (December 1, 1967): 565–66; Lucien Bianco, review of *Biographical Dictionary of Republican China*, Volume I, by Howard L. Boorman and Richard C. Howard, *Annales* 23, no. 5 (1968): 1133–35; D. C. Twitchett, review of *Biographical Dictionary of Republican China*, by Howard L. Boorman and Richard C. Howard, *Political Science Quarterly* 84, no. 4 (1969): 650–52.
- 33 Boorman and Howard, *Biographical Dictionary of Republican China*, viii.

tionary. The relations we have unveiled can open a new means of navigating<sup>34</sup> through the BDRC in a digital version, such as we have released (X-Boorman). However, network analysis has also allowed us to put the BDRC to a truth test. It reveals the tensions in this work between a largely densely connected “main component” – featuring a group of leading elites in the political and military fields – and different subsets of individuals, and even various disconnected individuals.

The BDRC contains a large but finite volume of biographical data. We processed only what was in the text, with no addition of external information. Yet the implementation of data mining and annotation methods based on NLP, followed by exploration with network analysis and PCA, allowed us to identify and trace patterns of relationships, to question certain assumptions about the types of relations among this composite elite population, and to breathe life into the stock of knowledge contained in the BDRC. The set of manual annotations of just a small sample of the biographies proved highly instructive, as a learning experience for historians to “define” the nature of relations in a text. People are multifaceted and it proved very challenging to reduce the nature of their relations to a single word. This also demonstrates the need for human intervention at every step, from close reading to defining terms and expressions, and to annotating the text.

This is an on-going experiment, but we believe that the models that we trained on our set of manual annotations have enormous potential to move toward automatic annotations of large biographical corpora. It paves the way for the exploration of similar corpora, such as the main English-language dictionaries and the numerous Chinese-language works published both before and after 1949.

---

34 <https://xboorman.enpchina.eu/> (accessed 18 December 2021).

## 6. Appendix: List of 36 Biographies

Name	Name (Chinese)	Name (Wade-Giles)
Chen Duxiu	陳獨秀	Ch'en Tu-hsiu
Zhang Fakui	張發奎	Chang Fa-k'uei
Zhang Xueliang	張學良	Chang Hsueh-liang
Zhang Xun	張勳	Chang Hsün
Zhang Binglin	章炳麟	Chang Ping-lin
Zhang Zuolin	張作霖	Chang Tso-lin
Jiang Jieshi	蔣介石	Chiang Kai-shek
Zhou Enlai	周恩來	Chou En-lai
Zhou Shuren	周樹人	Chou Shu-jen
Zhu De	朱德	Chu Teh
Feng Yuxiang	馮玉祥	Feng Yü-hsiang
Kong Xiangxi	孔祥熙	H. H. K'ung
He Long	賀龍	Ho Lung
He Yingqin	何應欽	Ho Ying-ch'in
Xu Shichang	徐世昌	Hsü Shih-ch'ang
Hu Hanmin	胡漢民	Hu Han-min
Hu Shi	胡適	Hu Shih
Huang Xing	黃興	Huang Hsing
Guo Moruo	郭沫若	Kuo Mo-jo
Li Liejun	李烈鈞	Li Lieh-chün
Li Shizeng	李石曾	Li Shih-tseng
Li Dazhao	李大釗	Li Ta-chao
Li Zongren	李宗仁	Li Tsung-jen
Li Yuanhong	黎元洪	Li Yuan-hung
Liang Qichao	梁啟超	Liang Ch'i-ch'ao
Liu Shaoqi	劉少奇	Liu Shao-ch'i
Mao Zedong	毛澤東	Mao Tse-tung
Sun Zhongshan	孫中山	Sun Yat-sen
Song Ziwen	宋子文	T. V. Soong
Cai Yuanpei	蔡元培	Ts'ai Yuan-p'ei
Duan Qirui	段祺瑞	Tuan Ch'i-jui
Wang Jingwei	汪精衛	Wang Ching-wei
Wu Peifu	吳佩孚	Wu P'ei-fu
Ye Ting	葉挺	Yeh T'ing
Yan Xishan	閻錫山	Yen Hsi-shan
Yuan Shikai	袁世凱	Yuan Shih-k'ai

## 7. References

- Aragon, Pablo, David Laniado, Andreas Kaltenbrunner, and Yana Volkovich. "Biographical Social Networks on Wikipedia: A Cross-Cultural Study of Links That Made History." In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration – WikiSym '12*, 1. Linz, Austria: ACM Press, 2012.
- Armand, Cécile. "Foreign Clubs with Chinese Flavor: The Rotary Club of Shanghai and the Politics of Language." In *Knowledge, Power, and Networks: Elites in Transition in Modern China*, edited by Cécile Armand, Christian Henriot, and Huei-min Sun. Leiden: Brill, forthcoming.
- Belsky, Richard. *Localities at the Center: Native Place, Space, and Power in Late Imperial Beijing*. Cambridge, Mass.: Harvard University Asia Center, 2005.
- Bianco, Lucien. Review of *Biographical Dictionary of Republican China*, Volume I, by Howard L. Boorman and Richard C. Howard. *Annales* 23, no. 5 (1968): 1133–35.
- Bieler, Stacey. *"Patriots" or "Traitors"?: A History of American-Educated Chinese Students*. New York: Routledge, 2003.
- Blouin, Baptiste, Pierre Magistry, and Nora Van den Bosch. "Creating Biographical Networks from Chinese and English Wikipedia." *Journal of Historical Network Research* 5 (2021): 303–317.
- Boorman, Howard L., and Richard C. Howard. *Biographical Dictionary of Republican China*. New York: Columbia University Press, 1967.
- Camp, Matje van de, and Antal van den Bosch. "A Link to the Past: Constructing Historical Social Networks." In *Proceedings of the 2Nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, 61–69. WASSA '11. Stroudsburg, PA, USA: Association for Computational Linguistics, 2011.
- "The Socialist Network." *Decision Support Systems* 53, no. 4 (November 2012): 761–69.
- Ch'ên, Jerome. *The Military-Gentry Coalition: China under the Warlords*. Toronto: University of Toronto-York University Joint Centre on Modern East Asia, 1979.
- Fairbank, J. K. Review of *Biographical Dictionary of Republican China*, Volume I, edited by Howard L. Boorman and Richard C. Howard. *The American Historical Review* 73, no. 2 (December 1, 1967): 565–66.
- Goodman, Bryna. *Native Place, City, and Nation: Regional Networks and Identities in Shanghai, 1853–1937*. Berkeley: University of California Press, 1995.
- Hershatter, Gail, Berkeley University of California, and Area Global and International Archive. *Women in China's Long Twentieth Century*. Berkeley: Global, Area, and International Archive: University of California Press, 2007.

- Klie, Jan-Christoph, Michael Bugert, Beto Boullosa, Richard Eckart de Castilho, and Iryna Gurevych. "The INCEpTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation," n.d., 5.
- Lary, Diana. *Region and Nation: The Kwangsi Clique in Chinese Politics, 1925–1937*. London; New York: Cambridge University Press, 1974.
- Levine, Marilyn. "Post WWI Chinese Revolutionary Leaders in Europe." *Journal of Historical Network Research* xx, no. xx (n.d.): xx–xx.
- Levine, Marilyn Avra. *The Found Generation: Chinese Communists in Europe during the Twenties*. Seattle, Wash.: University of Washington, 1993.
- Liu, Xiaoqin 劉曉琴. "Minguo liu Mei shetuan yu liu Mei sheng de shehui wangluo – yi Chengzhizhihui yu Zhang Boling de fenxi wei zhongxin" 民國留美社團與留美生的社會網絡 – 以成志會與張伯苓的分析為中心 [Chinese Students' Clubs and Social Networks in the United States: A Study of Zhang Boling and the 'Cross and Sword' Society]. *Huaqiao Huaren Lishi Yanjiu*, no. 4 (2019): 88–95.
- McCord, Edward Allen. *The Power of the Gun: The Emergence of Modern Chinese Warlordism*. Taipei: SMC Pub., 1997.
- Mittler, Barbara, Michael Hockx, and Joan Judge, eds. *A Space of Their Own: Women and the Periodical Press in China's Long Twentieth Century*. Cambridge: Cambridge University Press, forthcoming.
- Rowe, William T. *Hankow: Commerce and Society in a Chinese City, 1796–1889*. Stanford, Calif.: Stanford University Press, 1984.
- Rudolph, Henrike. "Structures of Empowerment: A Network Exploration of Women Activists' Collective Biographies in Twentieth-Century China." In *Knowledge, Power, and Networks: Elites in Transition in Modern China*, edited by Cécile Armand, Christian Henriot, and Huei-min Sun. Leiden: Brill, forthcoming.
- Shemo, Connie Anne. *The Chinese Medical Ministries of Kang Cheng and Shi Meiyu, 1872–1937: On a Cross-Cultural Frontier of Gender, Race, and Nation*. Bethlehem: Lehigh University Press, 2011.
- Smith, Steve and Taylor & Francis. *A Road Is Made: Communism in Shanghai 1920–1927*. Honolulu: University of Hawaii Press, 2018.
- Su, Gang, Allan Kuchinsky, John H. Morris, David J. States, and Fan Meng. "GLay: Community Structure Analysis of Biological Networks." *Bioinformatics* 26, no. 24 (December 15, 2010): 3135–37.
- Tamper, Minna, Eero Hyvönen, and Petri Leskinen. "Visualizing and Analyzing Networks of Named Entities in Biographical Dictionaries for Digital Humanities Research." EasyChair, April 8, 2019.
- Twitchett, D. C. Review of *Biographical Dictionary of Republican China*, by Howard L. Boorman and Richard C. Howard. *Political Science Quarterly* 84, no. 4 (1969): 650–52.
- Wang, Y. C. *Chinese Intellectuals and the West, 1872–1949*. Chapel Hill: University of North Carolina Press, 1966.

Warren, Christopher N., Daniel Shore, Jessica Otis, Lawrence Wang, Mike Finegold, and Cosma Shalizi. "Six Degrees of Francis Bacon: A Statistical Method for Reconstructing Large Historical Social Networks." *Digital Humanities Quarterly* 010, no. 3 (July 12, 2016).



ANNE S. CHAO/JIE YANG/  
ZHANDONG LIU/QIWEI LI

# Network of Words: A Co-Occurrence Analysis of Nation-Building Terms in the Writings of Liang Qichao and Chen Duxiu

*Journal of Historical Network Research* 5 (2021) 154–186

**Keywords** Liang Qichao, Chen Duxiu, Network Modules

**Abstract** Liang Qichao (1873–1929) and Chen Duxiu (1879–1942) were two of the most brilliant writers and influential public intellectuals in the late nineteenth, early twentieth century China. Born six years apart, both men electrified the country with their publications of *New Citizen's Journal* and *New Youth*, respectively, and heralded the character of a new and modern citizen, befitting a new century and a new China. Central to both men's concerns is the relationship between the citizen and the state. Liang, at the end of his checkered political career, concluded that the uninformed Chinese population would best be governed by enlightened autocracy. Chen, after his expulsion from the Chinese Communist Party and alienation from the Chinese Trotskyists, wavered between the dictatorship of the proletariat and democratic socialism. Did both men seemingly opt for authoritarian rule for China, and reject their ideal of liberty and democracy from their younger days? Our paper aims to test this hypothesis, that Chen and Liang both saw the need for a centralized power, albeit in different political frameworks, by using quantitative literature analysis. We examined the similarities and differences between their writings on the pairwise co-occurrence of thirty terms related to the topic of nation-building. We created a network with these thirty terms, where an edge between a pair of terms indicates significant relationship. The relationship is defined as the proportion of writings where both terms co-occurred. The visualization yields information on some preliminary differences in the writings of both men, to be further examined.



## 1. Introduction\*

In the waning years of the Qing dynasty (1644–1912), and during the first decade of the Republic of China, two men used their well-honed pens to excoriate, cajole, and teach their readers on how to act as members of a “modern” nation. Liang Qichao 梁啟超 (1873–1929) and Chen Duxiu 陳獨秀 (1879–1942) were arguably the two most influential public intellectuals of their times. Liang’s *Xinmin Congbao* 新民叢報 [New People’s Miscellany], published from 1902 to 1907,<sup>1</sup> enjoyed a circulation of 10,000 copies at the height of its popularity,<sup>2</sup> and had a huge impact on generations of readers, including the younger, so-called May Fourth generation. He wrote at length on how to construct a new China with a new citizenry, and introduced to his readers new political theories and new terminology from the West. While six years younger, Chen Duxiu (1879–1942) stands in equal stature to Liang in terms of influence and writing productivity. Publisher of the influential journal, *Xin Qingnian* 新青年 [New Youth] (1915 to 1922),<sup>3</sup> and widely considered leader of the New Culture movement (ca. 1915–1923), Chen admitted that he became a “Kang-Liang” [Kang Youwei and Liang Qichao] convert when he first traveled to Nanking for the civil service exam at the age of 19: “[I switched] from the study of witches and monsters to the Kang-Liang camp.”<sup>4</sup> Hu Shi, a renowned scholar and diplomat, hailed the *New People’s Miscellany* and *New Youth* as two of the three most influential journals of the century. *New Youth* featured progressive journalism of the sort pioneered so brilliantly by Liang Qichao a decade or so before.<sup>5</sup> The success of their publications propelled them to political

---

\* Corresponding author: Anne Chao (Rice University), annechao@rice.edu

1 This was published in Japan during Liang’s exile, but made its way to China without resistance.

2 Philip C. Huang, *Liang Ch’i-ch’ao and Modern Chinese Liberalism* (Seattle: University of Washington Press, 1972), 5.

3 *Xin Qingnian*’s last issue as a monthly ended with vol. 9, no. 6, published on July 1, 1922. Thereafter it became a quarterly publication. See Baoming Zhang 張寶明, *Duowei shiye xia de 《Xin Qingnian》 yanjiu* 多維視野下的《新青年》研究 [A multi-perspective study of *Xin Qingnian*] (Beijing: Shangwu yinshu guan, 2007), 319.

4 “Wo you xuanxue yaonie bian dao Kang, Liang pai” 我由選學妖孽變到康梁派。See Shen Ji, “Xinhai geming shiqi de Chen Duxiu” 辛亥革命時期的陳獨秀 [Chen Duxiu during the 1911 Revolution], *Jianghuai luntan* 江淮論壇 [Yantze and Huai River Forum], no.2 (Dec. 1979): 56–63, and “Shi’an zizhuan” 實庵自傳 [The autobiography of Shi’an], in *Chen Duxiu zhuzuo xuanbian* 陳獨秀著作選編 [A selected edition of the works of Chen Duxiu], ed. Ren Jianshu 任建樹 (Shanghai: Shanghai renmin chubanshe, 2008), vol. 5, 211. Henceforth ZZXB.

5 Hu Shi 胡適 enumerated three journals: Liang Qichao’s *The China Progress* [時務報], *New People’s Miscellany* [新民叢報], and Chen’s *New Youth* [新青年]. Chen Pingyuan called it the most influential journal of the century. See Chen Pingyuan 陳平原, *Chumo lishi: wusi renwu yu xiandai Zhongguo* 觸摸歷史: 五四人物與現代中國 [Touching history: The people of the May Fourth and modern China] (Guangzhou: Guangzhou chubanshe, 1999), 39, 61.

prominence. Both men engaged in political activities with the idealistic goal of saving China, but both came away disillusioned by their experiences. Toward the end of their lives, they both changed their minds about the pursuit of Western-style democracy, opting for a far more authoritarian system of political rule, albeit of very different political persuasions.

With the passage of time, Liang's iconoclasm, which initially inspired the May Fourth generation, was superseded and ignored by the latter. As Joseph Levenson observed: "The 'new youth' of *Xin Qingnian*, when they seemed to speak the same language as Liang, spoke so little of it with him. For the language was not really the same, 'identical' ideas were not identical."<sup>6</sup>

To explore what lies behind the differences in the seemingly identical language that represents dissimilar ideas, we turn to network analysis. We aim to understand the contextual change surrounding the "identical ideas," or key terms, by the use of word co-occurrence networks. We selected thirty key terms related to the theme of nation-building in the writing of both men and tracked them over three different time periods.

## 2. Historical Background

Several striking parallels appeared in the lives of the two intellectuals, and a brief biography of each is useful in order to contextualize their writing. Liang Qichao, born into an educated family near Guangzhou, passed the first level of the civil service exam for the *xiuca*i degree in 1883. He studied in the famous *Xuehai Tang* 學海堂 [Sea of Learning Hall] in Guangzhou, and successfully passed the second level, earning the distinction of being the youngest holder of the *juren* degree in 1889.<sup>7</sup> A year later, he traveled to Beijing for the next level of examinations and came across a copy of Xu Jiyu's (1795–1873) *Yinghuan zhilüe* 瀛寰志略 [A short account of the maritime circuit, 1840]. The book introduced him to knowledge about other countries in the world.<sup>8</sup> Having failed the exam, Liang became Kang's student and taught in the latter's academy, the Thousand Thatched Hall 萬木草堂 (*Wanmucao tang*).<sup>9</sup> Kang, who succeeded in obtaining a *jinshi* degree in 1895, but did not receive an official appointment due to the discrimination of an examiner, developed an eclectic philosophy in the early 1890s that was grounded in New

---

6 Joseph R. Levenson, *Liang Ch'i-ch'ao and the Mind of Modern China*. Cambridge, MA: Harvard University Press, 1965, 10–11.

7 Julia C. Schneider, *Nation and Ethnicity: Chinese Discourses on History, Historiography, and Nationalism (1900s–1920s)*, (Leiden: Brill, 2017), 68.

8 Schneider, 69.

9 Schneider, 69.

Text scholarship but which also drew upon Buddhist and Western ideas.<sup>10</sup> Kang borrowed ideas from the ancient texts to argue that reforms and changes were inevitable, and that Confucius was a reformer.<sup>11</sup>

Under Kang's tutelage, Liang engaged in reform advocacy on multiple fronts. In 1895, together with Kang, they collected almost a thousand examinees' signatures in a ten-thousand-word petition to the throne. The famous "Memorial of the Candidates" (*gongche shangshu*) 公車上書 protested against the Qing government's acceptance of the humiliating terms of the Treaty of Shimonoseki after China's defeat in the Sino-Japanese war of 1894–95, and urged the emperor to implement reforms. During this period, Liang failed the *jinshi* examination three more times. He became the secretary of Kang's study-society, the *Qiangxue hui* 強學會 (China Strengthening Society), and chief editor of the paper affiliated with the society, the *Qiangxue bao* 強學報. This paper was later renamed *Shiwu bao* 時務報 (China Progress) and was extremely influential. The visibility afforded by the paper led to Liang's appointment as director of a translation bureau in the Qing bureaucracy.<sup>12</sup> By 1898, Kang and Liang's reform efforts had reached the ear of the emperor Guangxu (1871–1908), who acted upon their recommendations over a period of a hundred days during the summer, from June 11 to September 20 (called the Hundred Days Reform). The drastic reforms included the abolition of the eight-legged essay, the dismissal of sinecure offices, the appointment of progressive-minded officials, and the creation of administrative bureaus, among other radical practices.<sup>13</sup> Threatened by the Emperor's new direction, Empress Dowager Cixi placed the Emperor under house arrest, ordered the death warrants for Kang and Liang, and executed their colleagues.

Kang and Liang fled to Japan, their passage and living expenses initially funded by the office of the Japanese prime minister Matsukata Masayoshi.<sup>14</sup> In Japan, Liang learned to read Japanese and became acquainted with Japanese scholars and politicians.<sup>15</sup> Liang greatly admired the reforms initiated by the Meiji emperor, which he understood as the reason for Japan's strength and power. Living in Japan during the years from 1898 to 1912, Liang first founded the newspaper, *Qingyi bao* 清議報 [*Pure Discussion Paper*], and later the *New People's Miscellany*. During this time he traveled to Hong Kong, Singapore, Ceylon, Australia, Taiwan,

---

10 Anne S. Chao, "Chen Duxiu's Early Years: The Importance of Personal Connections in the Social and Intellectual Transformation of China 1895–1920" (Ph.D. diss., Rice University, 2009), 43.

11 Immanuel C. Y. Hsu, *The Rise of Modern China* (New York: Oxford University Press, 2000), 365.

12 Schneider, 70.

13 Hsu, 365–379.

14 Schneider, 71.

15 Peter Zarrow, *After Empire: The Conceptual Transformation of the Chinese State, 1885–1924* (Stanford: Stanford University Press, 2012), 69.

Canada, the United States and the Philippines, as well as penning his most influential tracts, such as *Xin Shixue* 新史學 [New historiography] and *Xinmin shuo* 新民說 [Discourse on the new citizen]. Liang's vision of a new citizenry took the imagination of his readers by storm. Education, Liang argued, would produce a new people that was informed, empowered, responsible and capable of disinterested action."<sup>16</sup>

By 1912 the Qing dynasty has been overthrown, and Liang was able to return to the newly formed Republic of China. His renown was such that both Sun Yatsen and Yuan Shikai courted him aggressively. He sided with Yuan by merging his political party, *Minzhu dang* 民主黨 (Democratic party) with Yuan's *Gonghe dang* 共和黨 (Republican party), and the *Tongyi dang* 統一黨 (United party). The new entity was named *Jinbu dang* 進步黨 (Progressive party).<sup>17</sup> Liang became Minister of Justice and later the first Director of the Monetary Bureau in Yuan's government. After Yuan's death in 1916, he assumed the post of Minister of Finance in the new administration led by Li Yuanhong and Duan Qirui. He also formed a new study group, the *Xianfa yanjiu hui* 憲法研究會 (Association for constitutional research), also known as *Yanjiu xi* 研究系 (Research clique). Liang's group was embroiled in bitter political struggles between the Anhui and the Communications cliques. When his party lost the majority in the national assembly, Liang decided to leave politics. He raised enough funds to travel to the Paris Peace Conference in May 1919 as an unofficial adviser. The disheartening outcome of China losing the previously German-held territory in Shandong to Japan, and the lack of Western support in this dispute, thoroughly alienated Liang, leading to his complete renunciation of politics.<sup>18</sup> From 1919 until his death, Liang took teaching positions at Dongnan and Tsinghua universities, and wrote little about politics. He died in 1929 at the age of 55.

Chen Duxiu, six years younger than Liang, was born in Anqing, Anhui province, to a family with a long line of tutors. Educated at home by his grandfather and his older brother, he gained first place in the *xiuca* exam at the relatively young age of seventeen. By his own account, he intentionally failed at the next level, the *juren* exam, which took place in the provincial capital of Nanjing. As mentioned earlier, he was instantly captivated by the writings of Kang and Liang. In fact, his first publication, "An Account of the Topography of the Yangzi River" 揚子江形勢論略 [Yangzi jiang xingshi lunlüe] was an essay adapted from two articles in Liang's *China Progress*, which were written by Germans and translated into Chinese: "A Discussion of the Defense of the Yangzi River"

---

16 Peter Zarrow, "Liang Qichao and the Notion of Civil Society in Republican China," in *Imagining the People: Chinese Intellectuals and the Concept of Citizenship, 1890–1920*, ed. Joshua A. Fogel and Peter G. Zarrow (Armonk, N.Y., M. E. Sharpe: 1997), 232–257.

17 Schneider, 73.

18 Schneider, 75.

揚子江籌防芻議 [Yangzi jiang choufang chuyi] and “An Examination of the Cannon Bulwarks Along the [Yangzi] River” 查閱沿江砲臺稟 [Chayue yanjiang paotai bing]. He supplemented this twelve-page tract with details about his hometown and references to passages in the historical novel, *Romance of the Three Kingdoms*.<sup>19</sup> After spending a few years in Northeast China working with his uncle, and having witnessed first-hand the Russian mistreatment of the Chinese, Chen left China to study in Japan in 1901. In Japan, Chen, like Liang, encountered a world of new Western political philosophy, history, and social theories via Japanese translations. Although no mention was made of Chen and Liang being in touch, we know that Liang’s *New People Miscellany* was widely read by Chinese readers in Japan and in China. Ironically, while just a few years before Chen had respected Liang as a progressive political thinker, now Chen joined radical student societies, and soon regarded Liang as reactionary.

While Liang occupied cabinet positions in Yuan’s government, Chen served as secretary of the provisional government of Anhui province, and plotted to overthrow Yuan. When the attempt, as part of the “Second revolution,” failed, Chen’s attention turned to writing, predicting that his pen would have a “great impact” on Chinese society within a decade.<sup>20</sup> The publication that Chen created in 1915, *New Youth*, was wildly successful, and introduced iconoclastic and revolutionary ideas to young intellectuals in China in much the same way that Liang’s publication did a decade ago. Chen was invited to become the Dean of the School of Letters at Peking University by the venerable scholar/revolutionary, Cai Yuanpei. With a new cast of powerful intellectuals as contributors, Chen and the magazine brought about the New Culture movement, forever changing the cultural and social landscape of China.

With the debacle of the Paris Peace Conference in 1919, and undoubtedly influenced by the pessimistic tone of Liang’s reporting from Paris, Chen’s disillusionment with Western democracy deepened. Within two years, he formed the Chinese Communist Party (CCP) with the help of Russian Comintern agents and turned his attention to running a political organization. Politics proved to be too treacherous, however, and Chen was ousted from the CCP in 1929. Jailed by Chiang Kaishek from 1932 to 1937, Chen spent his time in prison writing and researching philology, a scholarly interest that he maintained throughout his life. He was released from jail at the onset of the Resist Japan war, but became a polit-

- 
- 19 Lü Xiaopo 閻小波, *Zhongguo zaoqi xiandaihua zhong de chuanbo meijie* 中國早期現代化中的傳播媒介 [The communication media in China’s early modernization] (Shanghai: Shenghuo, dushu, xinzhi, sanlian shudian, 1995), 227. The two articles are from *Shiwu bao* 時務報 no. 21–26, and no. 28–30, respectively.
- 20 Shen Ji 沈寂, “Wang Mengzou yu Chen Duxiu” 汪孟鄒與陳獨秀 [Wang Mengzou and Chen Duxiu], in *Chen Duxiu yanjiu* 陳獨秀研究 [The study of Chen Duxiu] (Beijing: Dongfang chubanshe, 1999), vol. 1, 380.

ical pariah whose writings few dared to publish. However, he continued to comment on the state of Chinese politics and died in penurious circumstances in 1942.

There were many similarities in the lives of Liang and Chen: both men attained national fame initially from their publications, both were deeply concerned with the problems facing China and explored ways to save the country, both engaged in politics by founding political parties, and both were rebuffed by the political machinations that robbed them of their offices. The difference lies in the fact that Liang came to national prominence before the fall of the Qing dynasty, and wrote his most influential essays in exile in Japan. Returning to China to assume political office in the various warlord governments, Liang concluded that China was not ready for democracy, but that constitutional monarchism may not be the best course for the country either. Chen began writing when the dynasty was moving toward a constitutional monarchy, but concluded that the only way to save China was to overthrow the system. He gained prominence on the national stage after the 1911 revolution, and his tenure as Dean of the School of Letters at Peking University elevated the stature and quality of his journal, *New Youth*, and of his own reputation. Interestingly, his subsequent role as secretary-general of the CCP, and later his expulsion from the party, led him to conclude that until China was ready for the democratic rule of the proletariat, a bourgeois democracy was better than Soviet and fascist autocracy.<sup>21</sup> The differences in the meaning of their “identical ideas” are partly a result of their divergent political convictions, and partly the changing political climate in which each man operated.

We applied a combination of text analysis and network analysis on the writings of the two men, looking for differences and similarities in their use of 30 terms related to their ideas on how to save China, a central concern in both men’s writing. We selected 391 pieces of Liang’s writing (essays, op-eds and letters) from the complete set of *Liang Qichao quanji* 梁启超全集 [The complete collection of Liang Qichao], and 491 pieces of Chen’s writing, from the six volumes of *Chen Duxiu zhuzuo xuanbian* 陳獨秀著作選編 [A selected edition of Chen Duxiu’s writings]; the selection criteria was based on the articles’ relevance to the theme of nation-building.<sup>22</sup>

The 30 terms are words that are judged to be seminal in the writings of both men, and that have the potential to generate different interpretations under each man’s pen. It is not a comprehensive list of keywords in the concept of nation-building, but rather a representative list that highlights the building blocks of both intellectuals’ approach to nation-building. The differences in the length of each piece of writing are normalized by the use of relative word frequency, as

---

21 Chen, “Gei Xiliu de xin” 給西流的信 [Letter to Xiliu], *ZZXB* vol. 5, 352–357

22 This selected edition is the most comprehensive collection of Chen’s writing to date.

we do not rely on the absolute number of times a word has appeared. Capturing synonyms for a key term is a challenge; in this paper we include “woman” and “women,” as well as “democracy” and “democratic,” as examples of how to capture as many expressions for these two keywords as possible.<sup>23</sup> The use of terms in a negative context is detectable because once the bag of words is generated, the author reads the 150 characters on either side of the keyword, thereby ensuring that the true meaning of the passage is detected.

“資產階級” – capitalist	“國民” – citizen
“憲政” – constitutional rule	“立憲” – constitutionalism
“民主” – democratic	“民主主義” – democracy
“生計” – economy	“進化” – evolution
“財政” – finance	“國政” – governance
“法律” – law	“自由” – liberty
“權限” – limit to power	“君政” – monarchy
“道德” – morality	“國會” – national assembly
“民族主義” – nationalism	“人民” – people
“國民會議” – people’s assembly	“民政” – people’s rule
“民權” – people’s power	“無產階級” – proletariat
“共和” – republic	“革命” – revolution
“社會主義” – socialism	“主制” – sovereignty
“主權” – sovereign rule	“蘇維埃” – soviet
“婦人” – woman	“婦女” – women

Chronologically, we have divided the writing of both men into three separate time periods, each corresponding approximately to changes in their careers. For Liang Qichao, the three periods are: 1892–1898, 1899–1912, and 1913–1920. The analysis begins with Liang’s earliest writing on nation-building up to 1898, when he escaped to Japan after the failure of the Hundred Day Reforms. The next period was the most prolific of Liang’s writing career, and he authored some of his most influential tracts. The third period begins with Liang’s active engagement in politics and ends with his leaving office for academia.<sup>24</sup>

- 
- 23 The use of embedding methods can partially address the use of synonyms in text analysis. The embedding method represents each distinct word with a list of numerical numbers. The value of the number is chosen carefully so that words with semantic similarity are closer to each other when simple distance functions are applied, such as cosine similarity. However, due to the rapid change in Chinese modern vocabulary, the embedding method is likely to require a large corpus of text to learn word associations. Thus, we prefer not to use the embedding method in this paper.
- 24 After 1920, Liang devoted his energy to academic issues and wrote few articles on nation-building. We therefore ended the third period in 1920, and not in 1929 when he passed away.

For Chen Duxiu, the three periods are: 1897–1919, 1920–1929, and 1930–1942. We begin our analysis in the first period with the publication of “An Account of the Topography of the Yangzi River,” and end with Chen’s departure from Peking University and taking a turn toward socialism and Marxism. The second period begins with Chen’s establishing the first cell of the Chinese Communist Party (CCP) and ends with his expulsion from the CCP. The third period begins with Chen’s advocacy of Trotskyism and ends with his death.

### 3. Methodology

To build a network that is related to the topic of nation-building (i.e. 30 terms) for each man, we propose a novel two-stage analytical framework based on word occurrences. First, we generate an  $n$ -by- $m$  bag-of-word count matrix  $C$  based on one man’s selected writings (i.e. 391 pieces for Liang and 491 pieces for Chen), where  $n$  is the number of writings and  $m$  is the number of terms. Each row of  $C$  corresponds to a piece of writing, each column corresponds to a term, and each entry  $c_{ij} \in N$  records the absolute frequency of term  $j$  in writing  $i$ , where we use  $i$  and  $j$  to index writings and terms. The count matrix usually suffers from sparsity, heterogeneity, and unknown mean-variance relationship, leading to a challenging task. To alleviate these problems, we dichotomize this matrix  $C$  to a binary matrix  $B$  with the same dimensionality, named the keyword indicator matrix. Each entry in  $B$ , i.e.  $b_{ij} \in \{0,1\}$ , indicates if term  $j$  is a keyword in writing  $i$ . This could be done via specifying an empirical threshold  $t \in Z^+$  and let  $b_{ij} = I(c_{ij} \geq t)$ , where  $I(\cdot)$  is an indicator function. For our analysis, we chose  $t = 3$ , which means if a selected term occurs three or more times in a piece of writing, then we consider it as a keyword within the text. To adjust for the length of each piece of writing, we could also consider setting  $b_{ij} = I(c_{ij}/l_i \geq \tilde{t})$ , where  $l_i$  is the length of writing  $i$  and  $\tilde{t} \in (0,1)$  is the normalized cutoff. We choose to use the absolute number of appearances, for the sake of simplicity.

The second step is to construct a network that reflects the dependency between each pair of terms based on their co-occurrences as keywords. Specifically, we first summarize a 2-by-2 contingency table  $O_{jj'}$  for each pair of terms  $j$  and  $j'$ . Here  $o_{jj'}^{00} = \sum_{i=1}^n I(b_{ij} = 0)I(b_{ij'} = 0)$ ,  $o_{jj'}^{01} = \sum_{i=1}^n I(b_{ij} = 0)I(b_{ij'} = 1)$ ,  $o_{jj'}^{10} = \sum_{i=1}^n I(b_{ij} = 1)I(b_{ij'} = 0)$ , and  $o_{jj'}^{11} = \sum_{i=1}^n I(b_{ij} = 1)I(b_{ij'} = 1)$ , corresponding to the number of writings where neither term  $j$  nor  $j'$  are the keywords, only term  $j'$  is a keyword but not  $j$ , only term  $j$  is a keyword but not  $j'$ , and where both of them are the keywords, respectively. Based on table  $o_{jj'}$ , we conduct a Pearson’s chi-squared test for independence, which is used to determine whether there is a statistically significant difference between the expected and the observed frequencies in one or more entries in a contingency table. The null hypothesis is that these two terms are independent in terms of being keyword to any writings. If the resulting  $p$ -value is smaller than a pre-specified significance level, then we reject the null hypothesis. For instance, if  $o_{jj'}^{00} + o_{jj'}^{11} \gg o_{jj'}^{01} + o_{jj'}^{10}$ , indicating their co-occurrences/



co-occurrences as keywords are more frequently observed than the opposite scenarios, then we expect an extremely small  $p$ -value, revealing they are less likely to be independent. We repeat the above statistical inferences simultaneously for all  $m(m - 1)/2$  pairs of terms. To prevent the inflation of false positive rates that occur with multiple statistical tests, we use the Benjamini-Hochberg procedure to control for the false discovery rate.<sup>25</sup>

A graph summary  $G = (V, E)$  of the set of terms  $V$  used by each author can be constructed by linking any two terms for which their adjusted  $p$ -value is smaller than a significance level  $\alpha$ . The value of  $\alpha$  is empirically set at a value of 0.001, or any value that results in an interpretable and meaningful network. An edge in  $E$  connects a pair of terms only if both terms are significantly interdependent, as quantified by the Pearson's chi-squared test. Note that the edge is undirected and thus the graph  $G$  does not have directionality. We further highlight their common edges and unique edges to compare the two men's networks.

Taking the keyword co-occurrence graph, we could extend the pairwise interaction defined by the co-occurrence of any two terms into a higher-order interaction among all the  $m$  terms. To understand these higher-order interactions, we adopted several commonly used graph metrics, such as node degree and betweenness centrality from graph theory.

The node degree is defined as the number of edges extending from a given node. In our context, it measures the number of direct terms that showed up in the same article as keywords. A higher node degree indicates that the term is often used with other terms together as keywords. The betweenness centrality measures the influence of a node over the flow of information between all possible pairs of nodes in a network, with the assumption that information flows through the shortest path connecting any two nodes. Mathematically, it is computed using the following three steps: 1) compute the shortest paths between all pairs of nodes in our graph  $G$ ; 2) for a pair of nodes  $(s, t)$  compute the fraction of shortest paths that pass through the vertex  $v$ ; 3) sum this fraction over all pairs of vertices  $(s, t)$ . In our context, it measures to what degree a key term links all the other terms together to form a semantic concept. Note that a higher node degree does not necessarily mean a higher betweenness centrality. To understand how a graph is organized, we applied a community detection algorithm, the Louvain algorithm, to search for modules.<sup>26</sup> Resolution is a tuning parameter used to con-

- 
- 25 Benjamini, Y. and Hochberg, Y. "Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: series B (Methodological)* 57, no. 1 (1995): 289–300.
  - 26 Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre, "Fast Unfolding of Communities in Large Networks." *Journal of Statistical Mechanics: Theory and Experiment* 10 (2008): 155–68.

trol the size of each module. We empirically set the tuning parameter to 1.0 which is also the default value based on visual inspection of the detected modules in each graph. Modularity was designed to measure the strength of a network's division into groups, clusters, or communities. Networks with high modularity have dense connections between the nodes within modules, but sparse connections between nodes in different modules. A modularity score can be computed based on the differences in the density of links within modules compared to links between modules. The modularity score ranges from  $-0.5$  to  $1.0$  with a higher value indicating nodes in a graph are densely connected within the same module and sparsely connected between modules. In our context, each module is composed of multiple terms that significantly co-occurred in each author's writings. Terms in a module often represent a cohesive topic or concept. We used node degree and betweenness centrality to understand the structure of the term network. Specifically, a term with a high node degree indicates that this term is frequently used with other linked terms in the author's writing. While a term with high betweenness centrality indicates that different concepts within a single network are connected through this node.

Once these calculations have revealed community modules of keywords, as well as identified keywords with the highest degree and betweenness centrality, we then conduct a close reading of the relevant passages containing these keywords. Our interpretation of the meaning of these keywords is derived from a close reading of the text based on the above calculations. For a visual representation of the workflow, please see the chart in Appendix B.<sup>27</sup>

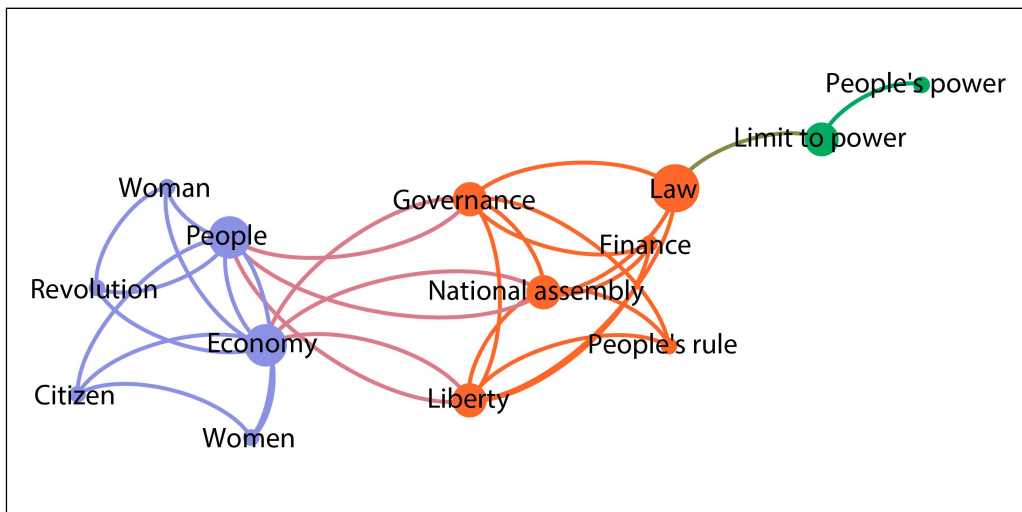
#### 4. Results

We chose over sixty-five articles written by Liang in the six-year period from 1892 to 1898, a time when he became a disciple and fellow political activist of Kang Youwei. Liang considered these to be the foundational years of his intellectual awakening.<sup>28</sup> Authoring these in China before his exile to Japan, Liang laid the groundwork for his discourse on nation-building. The network visualization shows three modules of words that are closely tied together. The blue module contains words primarily focused around personhood (woman, citizen, people, etc.). The orange module is anchored by words such as “national assembly” and “governance,” and the green module is concerned with “power,” both in terms of the people's power and the limits to power. At the same time, keywords “economy” and “people” ranked highest in degree and betweenness centrality [see Appendix A], followed by “governance,” “national assembly,” “law,” “people's rule,”

---

27 The data can be found on this website: <https://jyl-2.shinyapps.io/shiny/>

28 Chang, Hao. *Liang Ch'i-ch'ao and Intellectual Transition in China, 1890–1907* (Cambridge, MA: Harvard University Press, 1971), 60.



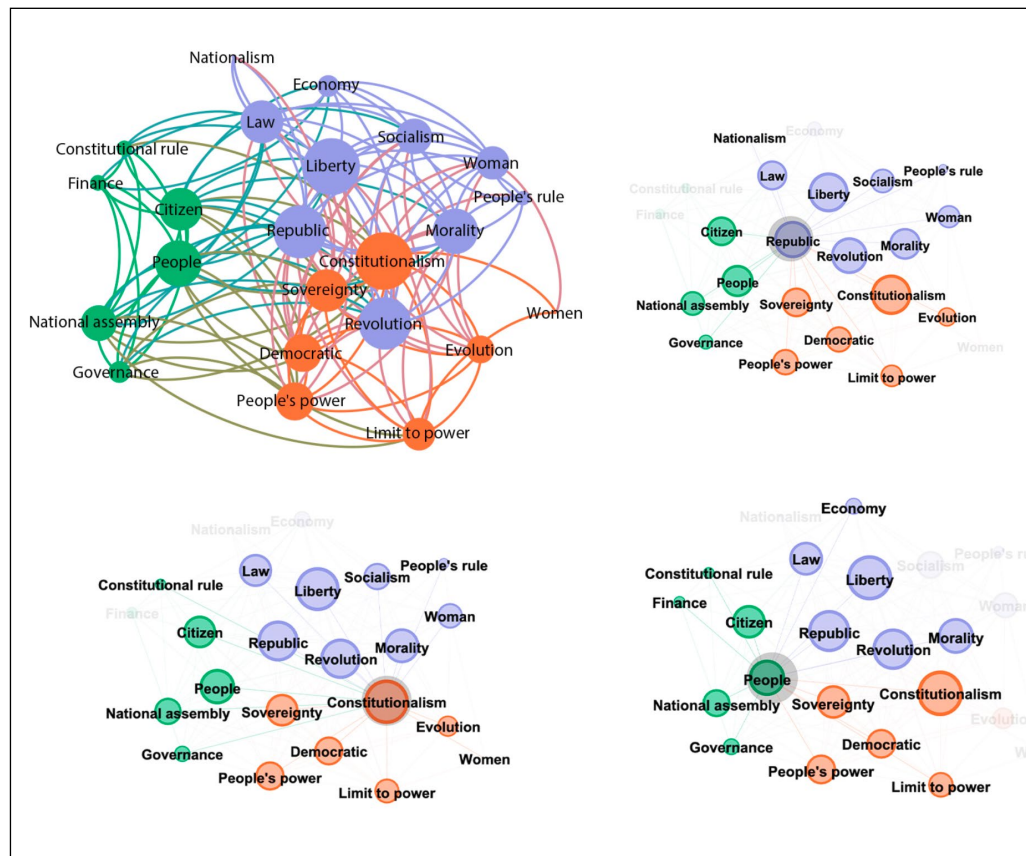
**Fig. 1** Liang Qichao 1st period 1892–1898. Three modules are detected in this period. Each node is colored by its module membership and the node size is proportional to its betweenness centrality. Modules were detected using the Louvain algorithm with resolution set to 1.0, which controls the number of modules. Our network has a modularity score of 0.3.

and “limit to power.” Translated into lay terms, it means that the words “economy” and “people” were most often paired with eight different keywords, but that all of the keywords above served as a popular route through which other keywords would pass. Liang introduced the concept of democracy as first practiced in ancient Greece and Rome, and explained that democracy could only succeed if the people had high education attainment and strong capabilities.<sup>29</sup> The people must learn how to self-rule, to practice governance in local settings, the better to prepare for participating in provincial and national assemblies.<sup>30</sup> Like all civilized nations, China needed to create a national assembly to tend to national affairs, but at the same time he warned that limits must be imposed on people’s rights.<sup>31</sup> Liang observed that the more complex the legal system of a country, the more

29 Liang Qichao 梁啟超, “Lun junzheng minzheng xiangshan zhi li” 論君政民政相嬗之理 [On the reason for the change from monarchy to people’s rule], *Liang Qichao quanji* 梁啟超全集 [The complete collection of Liang Qichao], (Beijing: Beijing chubanshe, 1999), vol. 1, 96–98. Henceforth *LQCQJ*.

30 Liang, “Wuxu zhengbian ji” 戊戌政變記 [An account of the Hundred Days reform], *LQCQJ*, vol. 1, 181–256.

31 Liang, “Wuxu zhengbian ji” 戊戌政變記 [An account of the Hundred Days reform], *LQCQJ*, vol. 1, 181–256; and “Lun Hunan ying ban zhishi” 論湖南應辦之事 [On tasks Hunan should accomplish], *LQCQJ*, vol. 1, 177–180.



**Fig. 2** Liang Qichao 2<sup>nd</sup> period, 1898–1912. Three modules are detected for this period. Nodes are heavily connected to each other both within each module and across different modules. Resolution is set to 1.0 and the modularity score is 0.1.

civilized that country would be.<sup>32</sup> He also highlighted the importance of a strong economy, of involving women in producing a livelihood, and pointed to the drain on the Qing coffer of the outlay for the Eight Banner troops.<sup>33</sup> The general impression of this period of writing is that Liang was preparing his readers to engage in governance, and to understand the strength and limits of democracy.

In the second period of our analysis (1899–1912), we selected 233 articles that dealt with the theme of nation-building. Liang was in Japan during these years, and visited the United States in 1906. This was a time when Liang absorbed a

32 Liang, “Wuxu zhengbian ji” 戊戌政變記 [An account of the Hundred Days reform], *LQCQJ*, vol. 1, 181–256; and “Lun Zhongguo yi jiangqiu falü zhixue” 論中國宜講求法律之學 [On China’s need to study law], *LQCQJ*, vol. 1, 60.

33 Liang, “Bianfa tongyi” 變法通議 [General discussion on reform], *LQCQJ*, vol. 1, 10–59.

torrent of ideas on Western political philosophy, Western history, Japanese governance and all sorts of concepts on political governance, from his readings and his travels. The network visualization reflects the intricate inter-connectedness of his words. As we see in Figure 3, the three modules are heavily connected, both within each module and with other modules. It is by finding the betweenness centrality and degree centrality of these modules that we can determine the hierarchy among the keywords. There are three words with a high degree of betweenness and centrality: “constitutionalism,” “liberty,” and “people,” and each keyword happens to be the central node in each module. Diving into the text, we learn that Liang favored constitutional monarchy with a division of power into three branches: the legislative, the executive, and the judiciary.<sup>34</sup> He observed the American presidential election process during his tour of the United States and concluded that the constitutional monarchy countenanced less abuse and was more flexible in implementation than a republican system.<sup>35</sup> China must begin with enlightened despotism for a short duration, then transition to constitutional monarchy, and definitely not become a republic at the time of his writing (1905). If China regressed from enlightened despotism, it would become a barbaric state, and regress to revolution. Then the cycle would begin again with enlightened despotism. By 1907, Liang was advocating for constitutional monarchy and the convening of the national assembly, at a time when the Qing court finally agreed to start preparing for a constitutional government, and as provincial assemblies began to appear around the country. With the success of the 1911 revolution, Liang worried that the Chinese people, after thousands of years of despotic rule, would not understand the obligations and responsibilities of a republican state.<sup>36</sup>

While text-mining the word “liberty” during this period, the most striking statement emerged when Liang toured the United States. He considered the San Francisco Chinatown to be a microcosm of China with a higher literacy rate and better developed governance system, but nevertheless one that was rife with violence and lawlessness. If the more politically advanced Chinese in America still abused their liberty, Liang reasoned, then the more backward Chinese in their home country could not possibly behave as responsible citizens in a republic.<sup>37</sup>

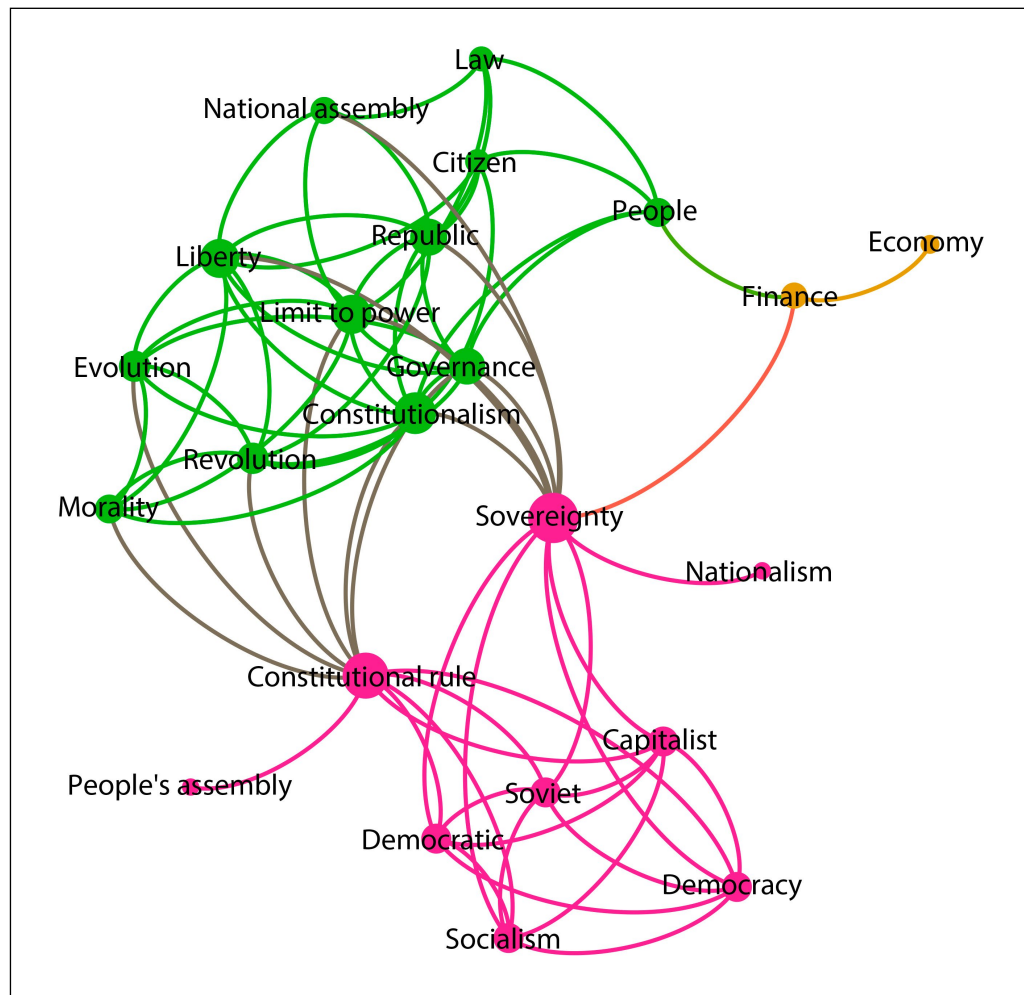
---

34 Liang, “Lixian fayi” 立憲法議 [Discussion on establishing the constitution], *LQCQJ*, vol. 2, 405–408; and “Geguo xianfa yitong lun” 各國憲法異同論 [Theory on the differences and similarities of other nations’ constitutions], *LQCQJ*, vol. 2, 318–322.

35 Liang, “Xin dalu youji jielu” 新大陸遊記節錄 [Excerpts from a tour of the New World], *LQCQJ*, vol. 4, 1125–1228.

36 Liang, “Zhongguo liguo da fangzhen” 中國立國大方針 [Major policy for the establishment of China], *LQCQJ*, vol. 8, 2488–2507.

37 Liang, “Xin dalu youji jielu” 新大陸遊記節錄 [Excerpts from a tour of the New World], *LQCQJ*, vol. 4, 1125–1228.



**Fig. 3** Liang Qichao 3<sup>rd</sup> Period, 1913–1920. The modularity score is 0.3. Two major modules are detected for this period. The red and green modules are clearly separated by “Constitutional rule” and “Sovereignty”.

Limits to liberty must be imposed, Liang argued: the people must agree to a social compact,<sup>38</sup> to obey the law and follow majority rule.<sup>39</sup>

The network ties for “people” during this period demonstrate that Liang was educating his readers on the duties and responsibilities of a new citizen: how to participate in parliament and how to supervise the government.<sup>40</sup> This was the time of his most influential publication, *Xinmin Shuo* [Discourse on the new citizen]. Warning that without limiting the respective powers of the state and the people, the nation could descend into despotism or anarchy, Liang called for clearly defined limits to the power of the state over the people, as well as imposing limits on individual liberty.<sup>41</sup> After the 1911 revolution, Liang worried that the people’s bonds to the state were weak, because the state never prepared its people to learn the duties and rights of a modern citizenry.<sup>42</sup> In these years of exile, Liang wrote hundreds of articles on the role of the state, the duties of citizens, and the preparations necessary to transition China into a republic.

In the third period of his writing, 1913–1920, Liang had returned to China from exile. He dove headlong into politics, hoping to use his knowledge and expertise to fortify the nascent republic. Liang served first as the Minister of Justice and director of the Monetary Bureau under Yuan Shikai’s government, and later as Minister of Finance in the Beiyang government and mastermind behind the Research Clique. We chose 88 articles from this period, and found that two keywords, “sovereignty” and “constitutional rule,” occupy gatekeeper positions bridging two community modules of keywords. In general, the module in fuchsia indicates broad concepts of political systems while the green module involves concepts dealing with the mechanics of governance. “Sovereignty” and “constitutional rule” both have the highest degree centrality (19) and betweenness centrality [See Appendix A]. If we look at the context surrounding the keyword “sovereignty,” we find that over time, Liang shifted his position on the locus of political power. In 1913, at the beginning of Yuan Shikai’s rule, Liang argued that sovereignty should reside in the hands of the government. By 1920, deeply hurt by the political infighting between warlord cliques, Liang concluded that sovereignty should only be in the hands of the citizens, because “if the people did not take it up themselves, the constitution would never be realized.”

---

38 Liang, “Zhongguo liguo da fangzhen” 中國立國大方針 [Major policy for the establishment of China], *LQCQJ*, vol. 8, 2488–2507.

39 Liang, “Shizhong dexing xiangcheng xiangfan yi” 十種德性相成相反議 [Discourse on the ten instances where virtue complement or contradict], *LQCQJ*, vol. 2, 428–432.

40 Liang, “Xinmin shuo” 新民說 [Discourse on the new citizen], *LQCQJ*, vol. 3, 655–735.

41 Liang, “Lun zhengfu yu renmin zhi quanxian” 論政府與人民之權限 [On the limit of power of the government and the people], *LQCQJ*, vol. 4, 881–883.

42 Liang, “Zhongguo liguo da fangzhen” 中國立國大方針 [Major policy for the establishment of China], *LQCQJ*, vol. 8, 2488–2507.

Analyzing the context of the keyword “constitutional rule,” we find that Liang was at his most pessimistic in assessing the future of the republic. Liang at first explained the need for constitutional rule, as it brings transparency to governance and imposes discipline on the behavior of the officials. But soon he determined that no good would come from citizens whose understanding of democracy was infantile. He lamented that the people behaved irresponsibly and were unfit to deliberate abstract political issues. By 1916, with the debacle of Yuan’s bid to monarchism, Liang was increasingly doubtful and critical of China’s attempt to establish a constitutional republic. He observed, “from the head of state to the officials, both in China and overseas, high and low, all abhorred the restrictions imposed by the law. They often paid lip service as they pleased. This is one big obstacle to constitutional rule... The people do not have an interest in politics nor an understanding of politics; their morality and capability cannot organize a real political party... In our country the majority of people lack the basic understanding of politics and their weakness in political ability makes it impossible to establish the foundations of constitutional rule.” Writing in 1916, Liang reverted back to his position in 1903, when he invoked German statist philosophers Johann Kaspar Bluntschli and Gustav Bornhak’s argument that a republic is an unstable form of government, subject to the whims and contentiousness of the different social groups. Deploring the immaturity of the Chinese people, and despairing of the corruption of the officials, Liang believed that China still needed an authoritarian mode of governance.

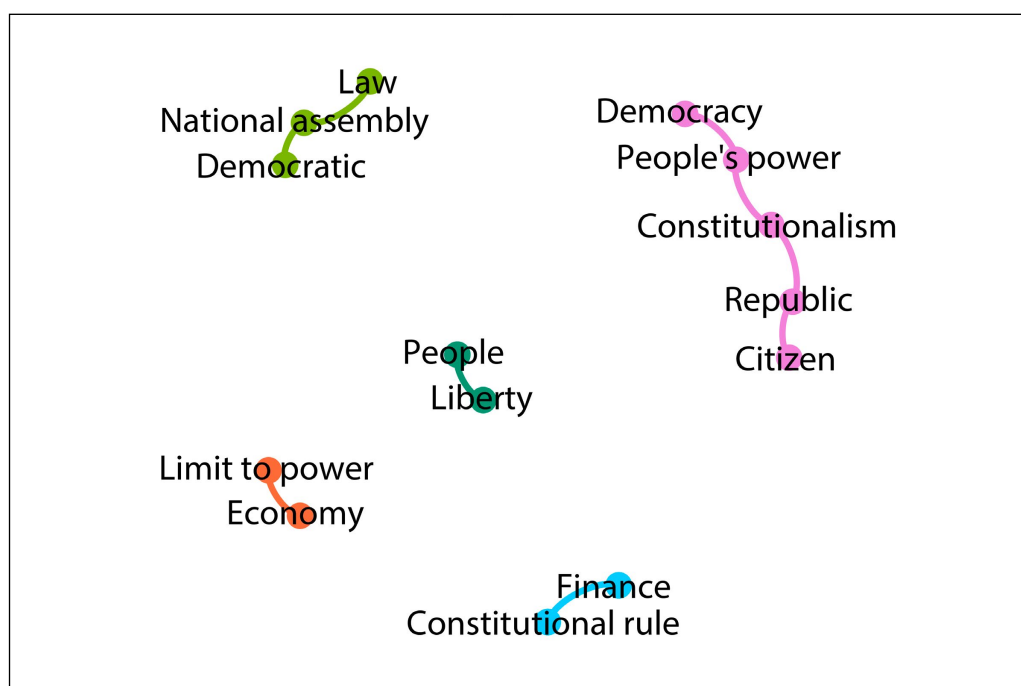
Chen Duxiu began his illustrious publishing career with the 1904–5 *Anhui Suhua bao* 安徽俗話報 [Anhui vernacular paper] and achieved national fame with the publication of *Xin Qingnian* [New Youth]. The keywords fall into five small community modules linked to the theme of nation-building. As with Liang’s first period, Chen is educating his readers on the building blocks of a republican system. But his position is more radical. In reading the passages around the keyword “constitutionalism,” we find that in 1914 Chen refuted Liang by saying “if the people did not have the intelligence to establish a republic, it does not follow that they could function better in a constitutional monarchy!”<sup>43</sup> In 1916, Chen shared Liang’s frustration with Yuan’s pretense to the throne, and agreed with Liang that the initiative to establish constitutionalism must come from the people, and not from the government. But ever the iconoclast, Chen blamed the root of China’s weakness on the Confucian tradition and called for the eradication of Confucian tenets such as the “three bonds and five relationships.”<sup>44</sup> Only then, he

---

43 Chen Duxiu 陳獨秀, “Aiguo xin yu zijue xin” 愛國心與自覺心 [Patriotism and Self-Consciousness]. In *ZZXB* vol. 1, 146–150; and “Da Wang Yonggong (guoti)” 答王庸工(國體) [Response to Wang Yonggong (national system)]. In *ZZXB* vol. 1, 167.

44 For an explanation of the Confucian three bonds and five relationships, see Richard J. Smith, *The Qing Dynasty and Traditional Chinese Culture* (Lanham, MD: Rowman & Littlefield, 2015), 222–223.





**Fig. 4** Chen Duxiu, 1<sup>st</sup> period, 1897–1919. The modularity score is 0.7. Five modules are detected in this period, where each module is a sparse chain graph.

argued, could a constitutional republic be based on the principles of independence, equality and freedom.<sup>45</sup> Writing a month after the May Fourth demonstration, in which thousands of students protested against the Beiyang government's betrayal over the lease of the Shandong peninsula, Chen was shifting away from republicanism. Chen expressed his rejection of constitutional rule and party politics: "Whether it's a constitutional monarchy or a republic, neither can guarantee our people the three big freedoms of religion, assembly and speech... We must not put blind faith in them."<sup>46</sup>

On the topic of national assembly Chen, like Liang, agreed that its most important function is to check the activities of the executive branch of the government, for instance to stop the Beiyang government from illegally pledging national resources to Japan in exchange for funding their private wars.<sup>47</sup> Another point of convergence occurred in Liang's and Chen's view on self-rule. Liang wrote in 1898 that the people must practice self-rule in preparation for participa-

45 Chen, "Wuren zhuihou zhi juewu" 吾人最後之覺悟 [My last awakening]. In ZZXB vol. 1, 201–204.

46 Chen, "Suigan lu" 隨感錄 [Collection of random thoughts]. In ZZXB vol. 2, III–II3.

47 Chen, "Suigan lu" 隨感錄 [Collection of random thoughts]. In ZZXB vol. 1, 407–409.

tion in provincial and national assemblies; in 1919, Chen called on the people to create their own popular assemblies, to take up issues dealing with public health, education, transportation, and then to bring these deliberations to the national assembly.<sup>48</sup> Both men pinned their hopes on the national assembly following a setback in their political activities. Liang's essay appeared after the failure of the Hundred Days Reform, while Chen penned his article upon his release from jail by the Beiyang government for distributing subversive pamphlets. On the topic of people's power, while both men agreed that a constitutional republic must be based on guaranteeing people's freedom,<sup>49</sup> Chen went one step further and suggested to his readers that refusing to pay tax was the most potent weapon Europeans had to rein in their government.<sup>50</sup> At this time, Chen's idea of the national assembly was very similar to Liang's: class division had not entered his vocabulary, but we will see that in the next period, Chen has a very definite understanding of what constitutes a proper national assembly.

The years 1920 to 1929 were a tumultuous time in Chen's life. He co-founded the CCP in 1920–21, only to be expelled nine years later. He transitioned from an intellectual whose chief occupation was teaching and writing, to a political activist who took orders from the Communist International and who tried to subvert the Nationalist Party (also known as the Guomindang, henceforth GMD) all the while claiming to be engaged in a United Front with them. Ultimately he took the blame for the CCP's failed revolution, resigned from the leadership, and was expelled from the party he founded. Three of the four modules are inter-linked, and we will examine the context around two keywords with the highest degree and betweenness centrality: "proletariat," and "revolution," as well as two words that anchor the two inter-linked modules: "evolution," and "people."

A rich context surrounds the keyword "proletariat." In his early days as leader of the CCP, Chen believed in a two-stage class struggle for China: the bourgeoisie versus the feudal warlords, and later the proletariat versus the bourgeoisie.<sup>51</sup> He repeatedly pointed out that because the Chinese industrial base is small, there is not a large enough proletariat class to realize a dictatorship of the proletariat.<sup>52</sup> The Communist International (Comintern) ordered the CCP to join the GMD, overruling Chen's opposition that the powerful GMD would overwhelm the nascent CCP. In order to justify this move to his cadres, Chen classified the Nation-

---

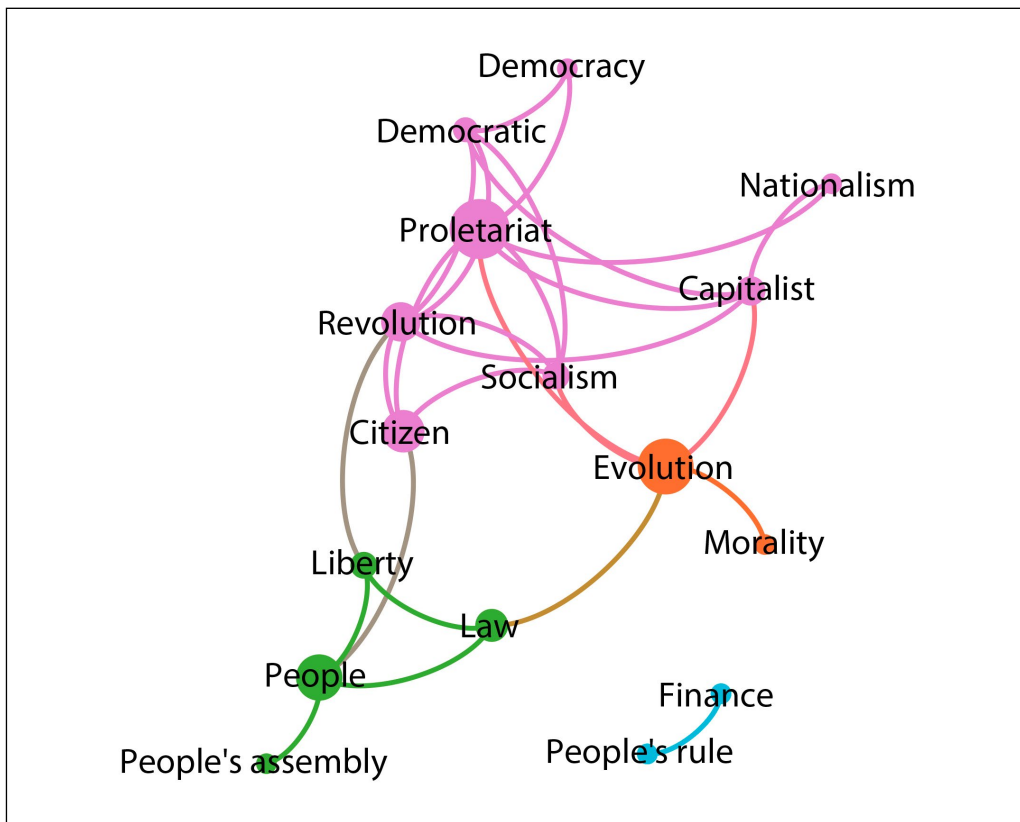
48 Chen, "Shixing minzhi de jichu" 實行民治的基礎 [Implementing the basis of people's rule]. In *ZZXB* vol. 2, 118–126.

49 Chen, "Minquan zhiyu ziyou" 民權之與自由 [People's rights and freedom]. In *ZZXB* vol. 1, 431–436.

50 Chen, "Suigan lu" 隨感錄 [Collection of random thoughts]. In *ZZXB* vol. 2, 66–68.

51 Chen, "Duiyu xianzai zhongguo zhengzhi wenti de wojian" 對於現在中國政治問題的我見 [About my views on today's Chinese political problem], *ZZXB* vol. 2, 467–470.

52 Chen, "Zao guolun" 造國論 [Discourse on nation-building], *ZZXB* vol. 2, 480–482.



**Fig. 5** Chen Duxiu 2nd period 1920–1929. Four modules are detected with a modularity score of 0.3.

alist Party as a party for all classes, and not just for the bourgeoisie.<sup>53</sup> Written at a time when the CCP was financially and technically supported by the Comintern, Chen reasoned that the peasants were half-proletariat, that China as a semi-colonial state could not become a revolutionary force, and that the people's immediate task was to support the GMD's national revolution, not to conduct a proletarian revolution.<sup>54</sup> He did not classify peasants as proletariat, and in order to appease the GMD, went so far as to suggest that China might not need the dictatorship of the proletariat to attain a socialist state.<sup>55</sup>

53 Chen, "Guomindang shi shenmo?" 國民黨是什麼 [What is the Guomindang?], *ZZXB* vol. 2, 483–484.

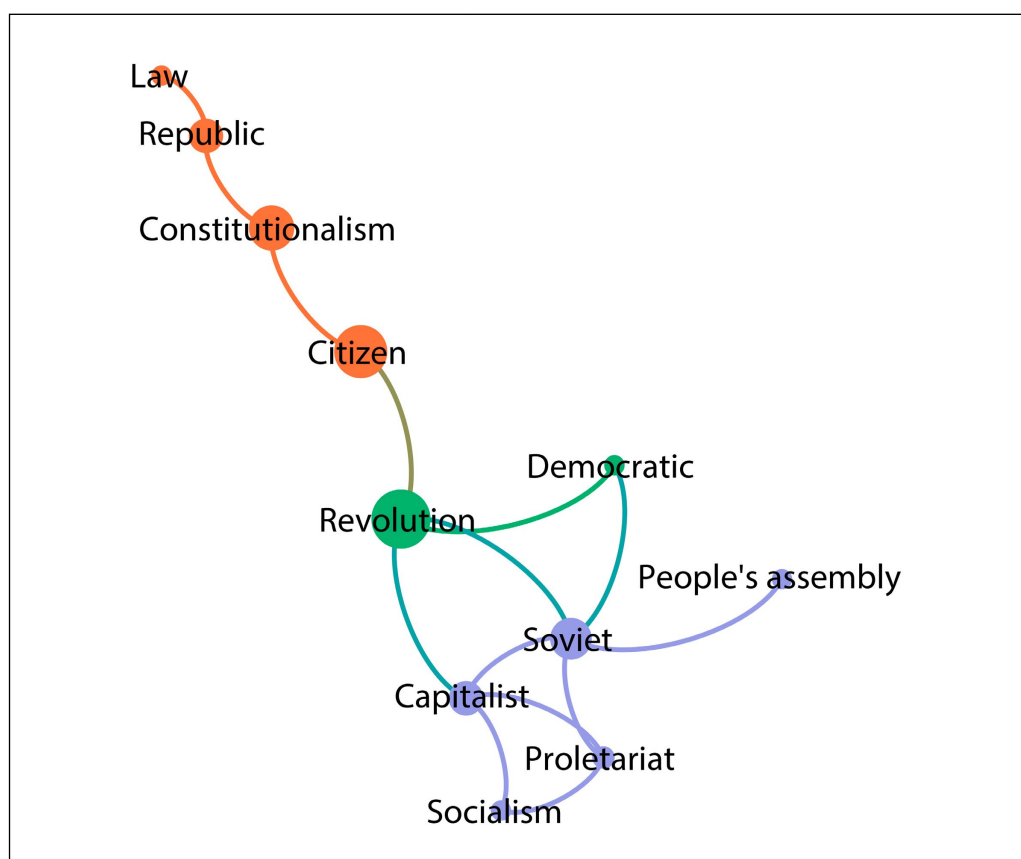
54 Chen, "Zhongguo geming yu shehui ge jieji" 中國革命與社會各階級 [Chinese revolution and all classes in society], *ZZXB* vol. 3, 153–161.

55 Chen, "Da Shen Binqi, Zhu Jinchi (Minguo geming zhi gui qu)" 答沈濱祈, 朱進赤 (民國革命之歸趨) [In response to Shen Binqi, Zhu Jinchi (Direction of the national revolution)], *ZZXB* vol. 4, 268–270.

Chen believed that the only way for the proletariat class to gain power was to conduct a violent revolution, but one guided by organized laborers, not by thugs.<sup>56</sup> Since the current regime did not guarantee the people freedom of speech, assembly, association, publication or religion, it was up to the people to rise up and seize power.<sup>57</sup> Forced into a United Front alliance with GMD by Comintern, Chen declared that now was the time of the GMD-led national revolution, and that the revolutionary proletariat class should ally with the bourgeoisie.<sup>58</sup> The peasants, Chen explained, were a massive force, but uneducated and therefore difficult to mobilize. It was the task of the proletariat to lead the peasants in the revolution.<sup>59</sup> By 1927, with the breakdown of the United Front and the massacre of thousands of CCP members by the right-wing GMD, Chen dropped the pretense of agreeing with the Comintern orders, and accused Chiang Kaishek of being an anti-revolutionary.<sup>60</sup> Declaring the failure of the “anti-imperialist” and “anti-feudal” revolution under the GMD, Chen called for the alliance of the proletariat and for the peasants to lead the next revolution.<sup>61</sup>

In the module with the keyword “evolution” as a central node, Chen considered the change of political systems to be evolutionary and natural, “From feudalism to republicanism, from republicanism to socialism,[it] is a natural evolutionary trajectory for society, and China is no exception.”<sup>62</sup> In addition, Chen explained that Marx views revolution as the natural outcome of socio-economic evolutionary forces.<sup>63</sup> Chen further elaborated that all races of people in the world would demand freedoms of speech, assembly and religion once they had achieved economic development and urban growth. And at that point, the political evolutionary force would generate a movement toward democratic constitutional rule.<sup>64</sup> In the module anchored by the keyword “people,” Chen spoke mostly of the powerless masses. He called for the government to convene a national assembly by the

- 
- 56 Chen, “Tan zhengzhi” 談政治 [Speaking of politics], *ZZXB*, vol. 2, 249–257; and “Suigan lu” 隨感錄 [Collection of ruminations], *ZZSB*, vol. 2, 311–316.
- 57 Chen, “Benbao xuanyan- *Xiangdao* fakanci” 本報宣言-《嚮導》發刊詞 [Our paper’s manifesto-Preamble to *Xiangdao*’s publication], *ZZXB*, vol. 2, 477–478.
- 58 Chen, “Zichan jieji de geming yu geming de zichan jieji” 資產階級的革命與革命的資產階級 [The revolution of the bourgeoisie and the revolutionary bourgeoisie], *ZZXB*, vol. 3, 33–37.
- 59 Chen, “Letter to Safarov” 給薩法羅夫的信 [Letter to Safarov], *ZZXB*, vol. 3, 108–109.
- 60 Chen, “Jiang Jieshi fandong yu Zhongguo geming” 蔣介石反動與中國革命 [Reactionary Chiang Kaishek and the Chinese revolution], *ZZXB*, vol. 4, 303–308.
- 61 Chen, “Cuntie” 寸鐵 [Inch of iron], *ZZXB*, vol. 4, 370–372.
- 62 Chen, “Guqing jinian de jiazhi” 國慶紀念的價值 [The value of national independence celebration], *ZZXB* vol. 2, 277–80.
- 63 Chen, “Makesi xueshuo” 馬克思學說 [The theory of Marxism], *ZZXB* vol. 2, 441–449.
- 64 Chen, “Benbao xuanyan- <*Xiangdao*> fakanci” 本報宣言-《嚮導》發刊詞 [Our paper’s manifesto-Preamble to *Xiangdao*’s publication], *ZZXB*, vol. 2, 477–478.



**Fig. 6** Chen Duxiu 3rd period 1930–1942. The modularity score is 0.36.

people, and not one led by factions of warlords.<sup>65</sup> The writings of this period reflect Chen’s focus on leading the CCP and accommodating the erroneous directives of Stalin through the agents of the Comintern. In the third period of our analysis, we will see a reversal in Chen’s perspective on politics.

In this period, three modules were detected. The red module contains words such as “law,” “republic,” “constitutionalism” and “citizen,” forming a theme around state-building. The purple module contains words such as “socialism,” “capitalist,” “people’s assembly,” “proletariat,” and “soviet”, reflecting his late conversion to Trotskyism and his debate with the Stalinists of the CCP. The green module contains two words, “revolution” and “democratic,” with “revolution” having a high betweenness centrality value. “Revolution” is the keyword linking

65 Chen, “Guomin huiyi cucheng hui yu Zhongguo zhengju” 國民會議促成會與中國政局 [Preparatory meeting for convening the national assembly and Chinese political situation], *ZZXB* vol. 3, 399.

the topics represented by the red and purple modules. We will examine the context surrounding “soviet” and “revolution” to understand Chen’s last views.

The 1930s saw Chen convert to Trotskyism, then jailed for almost five years by the GMD, and finally released when the Resist Japan war began in 1937. He was a political pariah and was unable to publish many of his writings. He lived in penurious circumstances until his death in 1942. Even though he declared that his political views were no longer partisan and reflected only his personal opinion, his loyal Trotskyist followers still claimed that he remained a Trotskyist till the end.

“Soviet” is a trope for both the Stalinists the Trotskyists; both sides agree that at some point during the proletariat revolution, it should be used as a “slogan” to rally the base and a strategy to conduct mass movement. But the two camps disagree on when to use this instrument, and who belongs in the soviet.<sup>66</sup> Chen believed that the constituent assembly and the soviet were two complementary vehicles for fighting the Japanese imperialists, as well as to wrestle power away from the GMD. Chen maintained that a successful soviet should be led by a democratic coalition of the proletariat, the peasants, and the soldiers, which would rally the masses to bring about the revolution. The Stalinists called for a soviet that was composed of a dictatorship of the proletariat and the peasants, which Chen viewed as a mistake.<sup>67</sup> At the time of writing, the peasants had already created soviets in the countryside, but Chen did not think they had the power to control the urban centers.<sup>68</sup> Chen regarded the soviet as the highest form of democratic rule and the most powerful and flexible stage in the evolution to a people’s democracy.<sup>69</sup>

Chen evoked the keyword “revolution” when he recast the Resist Japan war as a revolutionary act against an imperialist power. He also used the word to describe the masses, as in “the revolutionary masses must arm themselves and join the war against imperialism.”<sup>70</sup> The revolutionary national assembly that Chen championed was to be organized by representatives of all anti-Japanese patriots,

66 Gregor Benton ed. *Prophets Unarmed: Chinese Trotskyists in Revolution, War, Jail, and the Return from Limbo* (Leiden: Brill, 2015), 24.

67 Chen, “Gao quandang tongzhi shu” 告全黨同志書 [Letter to all Party comrades], *ZZXB*, vol. 4, 414–429; and “Yige jinji de zhengzhi wenti” 一個緊急的政治問題 [An urgent political problem], *ZZXB*, vol. 5, 10–19.

68 Chen, “Yige jinji de zhengzhi wenti” 一個緊急的政治問題 [An urgent political problem], *ZZXB*, vol. 5, 10–19.

69 Chen, “Women ya zeyang de minzhuzhengzhi?” 我們要怎樣的民主政治 [What kind of democratic rule do we want?], *ZZXB* vol. 5, 22–27.

70 Chen, “Cici kangri jiuguo yundongde kangzhuang dadao” 此次抗日救國運動的康莊大道 [The healthy and proper way to the Resist Japan National Salvation movement], *ZZXB* vol. 4, 518–525.

and not by bourgeois elements appointed by the GMD. Under those conditions, the revolutionary national assembly would rise to lead the anti-Japanese war effort for the whole country.<sup>71</sup> Indeed, Chen declared, the Resist Japan war was the beginning of a new revolution, a culmination of the revolutions that began with the Self-Strengthening movement of the 1860s, the Hundred Days Reform, the 1911 Revolution, and the Northern Expedition. Chen called for the revolutionary masses around the world to come to the aid of the Chinese people and help them overthrow the Japanese imperialists.<sup>72</sup> Chen also cautioned that not all historic times were conducive to stage revolutions. He condemned the Comintern's deceitful depiction of a revolutionary high tide after the massacre of Communists at the hands of the GMD in 1927, which subsequently mobilized more Communist cadres into deadly encounters with the GMD.<sup>73</sup>

In the final years of his life, Chen opposed any kind of dictatorship, including that of the proletariat. In the five big countries in the world in 1940, Chen singled out three dictatorships: Russia, Germany and Italy. "Any struggle in the world today must be linked to destroying these three bastions... otherwise any euphemism, such as proletariat revolution, people's revolution, would all inadvertently help these three bastions aggrandize their power. If we consider destroying these three powers to be the most important struggle, then we first must admit that the not-so-thoroughly democratic systems of Britain, France, and the United States deserve to be preserved."<sup>74</sup> He cited the positive and negative aspects of these regimes: in Britain, France and the U.S., national assemblies are elected by the people; opposition parties are allowed to exist, and there is debate during meetings. In those countries, people cannot be arrested or killed without court order. The people enjoy extensive freedom of thought, speech and publication; and waging a strike is not a crime. The opposite is true for Russia, Germany and Italy. In the end, Chen called for a "democracy of the proletariat," and pointed out that political democracy and economic socialism are complementary systems for a country.<sup>75</sup>

---

71 Chen, "Zhongguo minzu yinggai zeyang jiuguo ji ziji" 中國民族應該怎樣就國及自救 [How should the Chinese people save the country and themselves], *ZZXB* vol. 5, 1–4.

72 Chen, "Kangri zhanzheng zhi yiyi" 抗日戰爭之意義 [The meaning of the Resist Japan war], *ZZXB* vol. 5, 176–180.

73 Chen, "Wo de genben yijian" 我的根本意見 [My fundamental views], *ZZXB* vol. 5, 358–361.

74 Chen, "Gei Xiliu de xin" 給西流的信 [Letter to Xiliu], *ZZXB* vol. 5, 352–357.

75 Chen, "Wo de genben yijian" 我的根本意見 [My fundamental views], *ZZXB* vol. 5, 358–361.

## 5. Conclusion

In the process of creating the global networks of all keywords, for all the years of both men, we unexpectedly found that Liang had a richer and more varied vocabulary, especially during his second period, whereas Chen used a smaller set of keywords but repeated them often. (See Fig. 7. The red edges in both diagrams indicate commonly paired keywords in both men's writing). Stylistically, this indicates that Chen relied on the rhetorical device of using the same key terms repeatedly to drive home a point. Liang wrote on a broad range of topics and delved deeply into each area, as for example when he discussed finance, where he analyzed topics such as public debt, bonds and currency in great detail. By contrast, Chen's writing reflects his intent to use succinct ideas in order to persuade his readers to pursue a particular set of actions. Further study would undoubtedly yield interesting implications about this difference.

We began our inquiry with the hypothesis that both Liang and Chen, after a lifetime of advocating for a more open political system, reverted back to a more restrictive governance at the end of their respective lives. Our methodology produced keywords with the highest co-occurrence frequency, and thereby identified passages in both men's writings, which were then subject to a close-reading process. This process is supplemented by frequent fact-checking with secondary historical sources and occasional reading of the entire article, as a way to ensure maximum accuracy.

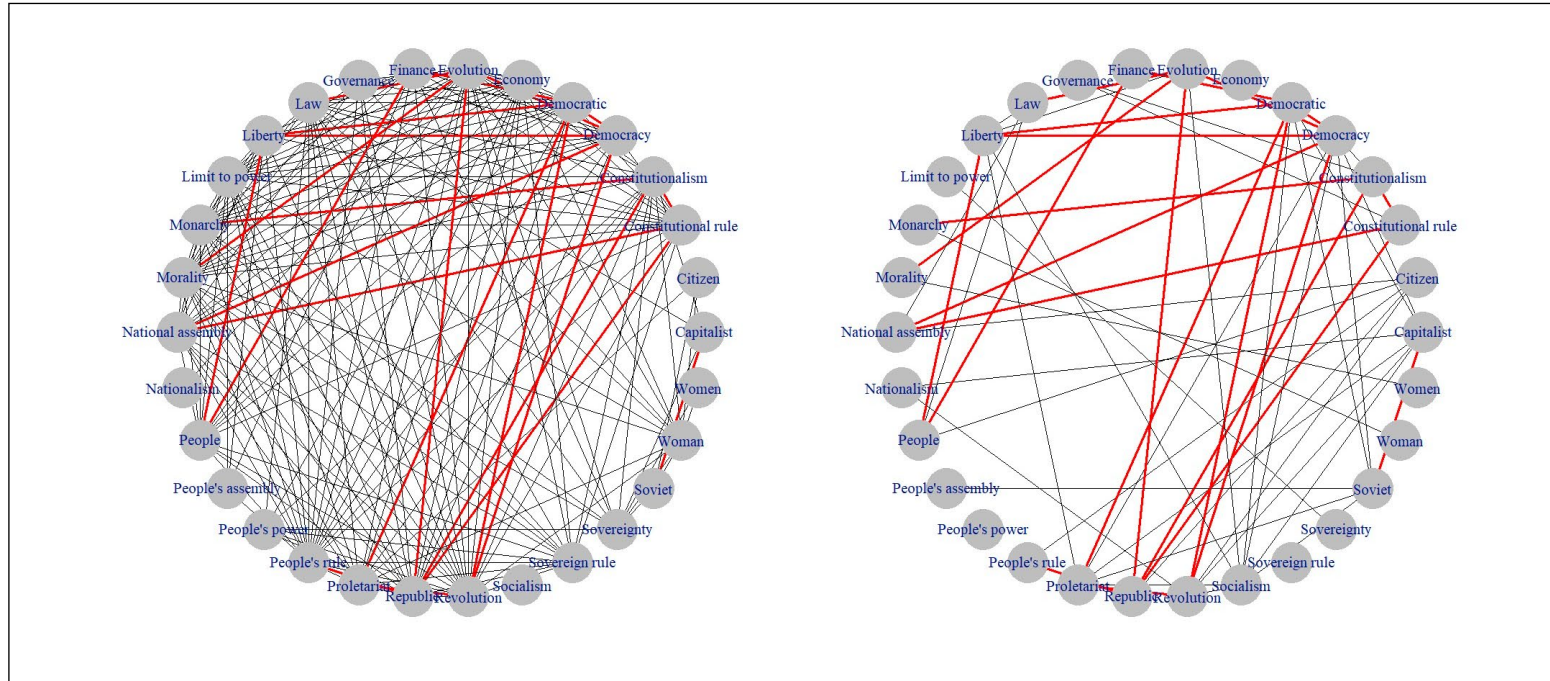
For Liang Qichao, who at first advocated constitutional monarchy, and who later worked in the Republican government, our finding reflects his pessimistic assessment that China was not ready for a constitutional republic. However, rejecting China's return to monarchism, Liang opted for a limited form of "democratic rule" in which the government and the people were both subject to checks and balances. As Peter Zarrow concluded, Liang's idea "... is not democracy conceived in terms of participation and direct decision making... [It] places the practical realm of politics at some distance from the people."<sup>76</sup> In the case of Chen Duxiu, his last few articles reflect a certain ambivalence on the feasibility of the democracy of the proletariat. Historians have debated at length whether Chen, at the end of his life, reverted to his younger idealism for Western democracy. His praise for the "not-so-thoroughly democratic" countries of Britain, France and the United States harkens back to his May Fourth proclamation that the French civilization gifted mankind with three major ideas: "human rights, biological evolution, and socialism."<sup>77</sup> However, his Trotskyist comrades insisted

---

76 Zarrow, "Liang Qichao and the Notion of Civil Society," 252.

77 Chen, "Falanxi ren yu jinshi wenming" 法蘭西人與近世文明 [The French and Modern Civilization], *ZZXB* vol. 1, 164–166.





**Fig. 7** The global networks of Liang Qichao (left) and Chen Duxiu (right). Each node in the network is a curated term. Two nodes are connected by an edge if the  $p$ -value of Pearson's chi-squared test is less than 0.001. This means the probability is less than 0.1% based on the observed writing, if it is assumed that the two terms independently occurred as keywords among all the writing. The edges in red are shared by both men.

that Chen remained a true Trotskyist until the end of his days, and that Chen's ultimate goal was not a proletariat dictatorship nor a capitalist democracy, but a socialist democracy.<sup>78</sup>

## 6. References

- Benjamini, Y. and Y. Hochberg. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: series B (Methodological)* 57, no. 1 (1995): 289–300.
- Benton, Gregor ed. *Prophets Unarmed: Chinese Trotskyists in Revolution, War, Jail, and the Return from Limbo*. Leiden: Brill, 2015.
- Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte and Etienne Lefebvre, "Fast Unfolding of Communities in Large Networks." *Journal of Statistical Mechanics: Theory and Experiment* 10 (2008): 155–68.
- Chang, Hao. *Liang Ch'i-ch'ao and Intellectual Transition in China, 1890–1907*. Cambridge, MA: Harvard University Press, 1971.
- Chao, Anne S. "Chen Duxiu's Early Years: The Importance of Personal Connections in the Social and Intellectual Transformation of China 1895–1920" Ph.D. diss., Rice University, 2009.
- Chen, Pingyuan 陳平原. *Chumo lishi: wusi renwu yu xiandai Zhongguo* 觸摸歷史: 五四人物與現代中國 [Touching history: The people of the May Fourth and modern China]. Guangzhou: Guangzhou chubanshe, 1999.
- Hsu, Immanuel C. Y. *The Rise of Modern China*. New York: Oxford University Press, 2000.
- Huang, Philip C. *Liang Ch'i-ch'ao and Modern Chinese Liberalism*. Seattle: University of Washington Press, 1972.
- Levenson, Joseph R. *Liang Ch'i-ch'ao and the Mind of Modern China*. Cambridge, MA: Harvard University Press, 1965.
- Lü, Xiaobo 閻小波. *Zhongguo zaoqi xiandaihua zhong de chuanbo meijie* 中國早期現代化的傳播媒介 [The communication media in China's early modernization]. Shanghai: Shenghuo, dushu, xinzhi, sanlian shudian, 1995.
- Ren, Jianshu 任建樹 ed. *Chen Duxiu zhuzuo xuanbian* 陳獨秀著作選編 [A selected edition of the works of Chen Duxiu]. Shanghai: Shanghai renmin chubanshe, 2008, vols. 1–6.
- Schneider, Julia C. *Nation and Ethnicity: Chinese Discourses on History, Historiography, and Nationalism (1900s–1920s)*. Leiden: Brill, 2017.
- Shen, Ji 沈寂 ed. *Chen Duxiu yanjiu* 陳獨秀研究 [The study of Chen Duxiu]. Beijing: Dongfang chubanshe, 1999.

---

78 Zheng Chaolin, "Chen Duxiu and the Trotskyists," in Gregor Benton ed., *Prophets Unarmed: Chinese Trotskyists in Revolution, War, and Return from Limbo* (Chicago: Haymarket Books, 2017), 681.

- Shen, Ji 沈寂. "Xinhai geming shiqi de Chen Duxiu" 辛亥革命時期的陳獨秀 [Chen Duxiu during the 1911 Revolution], *Jianghuai luntan* 江淮論壇, no. 2 (1979): 56–63.
- Yang, Gang 楊鋼, and Wang Xiangyi 王相宜, eds. *Liang Qichao quanji* 梁啟超全集 [The complete collection of Liang Qichao]. 21 vols. Beijing: Beijing chubanshe, 1999.
- Zarrow, Peter. *After Empire: The Conceptual Transformation of the Chinese State, 1885–1924*. Stanford: Stanford University Press, 2012.
- Zhang, Baoming 張寶明. *Duowei shiye xia de 《Xin Qingnian》 yanjiu* 多維視野下的《新青年》研究 [A multi-perspective study of *Xin Qingnian*]. Beijing: Shangwu yinshu guan, 2007.

## 7. Appendix

Time	Betweenness Centrality	Degree Centrality
<b>LQC 1st period (1892–1898)</b>		
Economy	0.044334975	8
People	0.044334975	8
Governance	0.030377668	7
National Assembly	0.030377668	7
Law	0.054187192	4
Limit to Power	0.02955665	2
<b>LQC 2nd period (1899–1912)</b>		
Constitutionalism	0.044922927	19
Liberty	0.034782288	19
Republic	0.026061903	18
Revolution	0.018253169	18
Morality	0.010895194	16
People	0.017464405	16
Citizen	0.013883202	15
Sovereignty	0.014930178	15
Law	0.018377388	15
Democratic	0.003775883	14
People's power	0.005309339	14

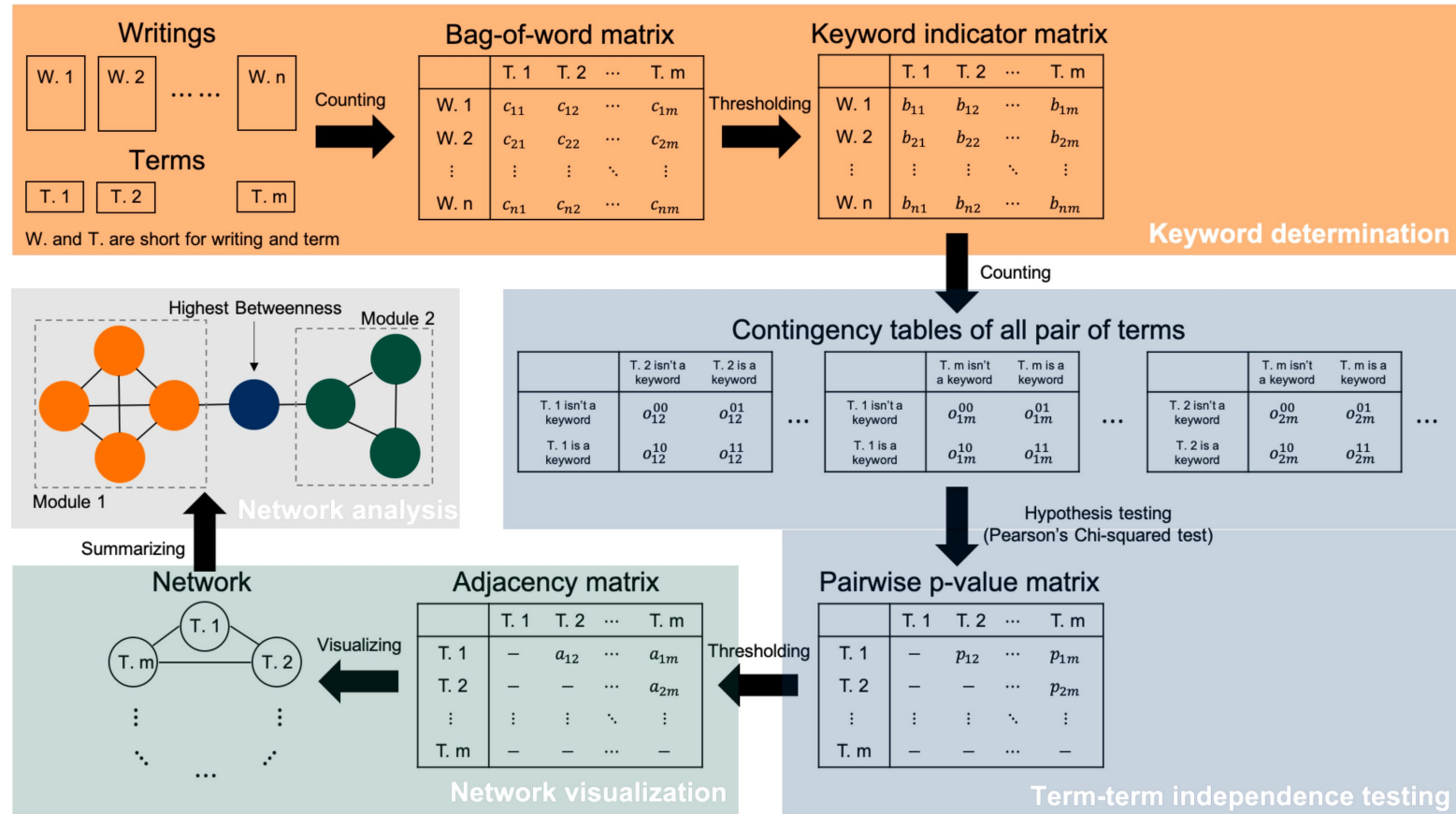
Time	Betweenness Centrality	Degree Centrality
National assembly	0.010232515	13
Socialism	0.015845737	13
Limit to Power	0.012112327	12
Woman	0.016059078	12
Evolution	0.001443623	11
Governance	0.003343783	9
Economy	0.004679803	9
Finance	0.003920361	7
Constitutional rule	0.001724138	7
Women	0.000307882	3
<b>LQC 3rd period (1913–1920)</b>		
Sovereignty	0.201642036	13
Constitutional rule	0.118988975	12
Constitutionalism	0.050639221	11
Limit to Power	0.039449918	10
Liberty	0.024448745	10
Governance	0.02559817	9
Republic	0.023956134	9
Evolution	0.001583392	7
Revolution	0.003571429	7
Capitalist	0.002216749	6
Socialism	0.002216749	6
Morality	0.001172883	6
Democracy	0.002216749	6
Democratic	0.002216749	6
Soviet	0.002216749	6
Finance	0.055829228	3
National assembly	0.005541872	5
People	0.024753695	5

Time	Betweenness Centrality	Degree Centrality
Law	0.003899836	4
Citizen	0.001436782	4
<hr/> <b>LQC Global (1892–1920)</b>		
Morality	0.071705107	19
Revolution	0.084526735	17
Sovereign rule	0.065035601	17
Constitutional rule	0.031406228	16
Constitutionalism	0.071610575	16
Republic	0.018140722	15
Evolution	0.014900996	14
Democratic	0.008729762	14
Monarchy	0.024818548	14
People	0.017100007	14
People's rule	0.021361808	14
Proletariat	0.003849446	12
Law	0.068090905	12
Limit to Power	0.011714273	11
National assembly	0.002256729	10
Liberty	0.064350122	8
Democracy	0.003974884	8
Finance	0.021337945	7
Woman	0.067763018	7
Sovereignty	0.009461256	6
Economy	0.00249814	5
Soviet	0.009537965	3
Governance	0.000410509	2
<hr/> <b>CDX 1st period (1897–1919)</b>		
Constitutionalism	0.009852217	2
National assembly	0.002463054	2

Time	Betweenness Centrality	Degree Centrality
People's power	0.007389163	2
Republic	0.007389163	2
<b>CDX 2nd period (1920–1929)</b>		
Proletariat	0.05090312	8
Revolution	0.024220033	6
Capitalist	0.011288998	5
Democratic	0.005541872	5
Evolution	0.045566502	5
Socialism	0.008415435	5
Citizen	0.028325123	4
People	0.033251232	4
Liberty	0.008004926	3
Law	0.016009852	3
<b>CDX 3rd period (1930–1942)</b>		
Soviet	0.033661741	5
Capitalist	0.022577997	4
Revolution	0.061165846	4
Proletariat	0.003284072	3
Republic	0.022167488	2
Citizen	0.051724138	2
Constitutionalism	0.039408867	2
<b>Global (1897–1942)</b>		
Democracy	0.017423	8
Capitalist	0.017171	7
Democratic	0.0251	7
Liberty	0.053249	7
Proletariat	0.017171	7
Soviet	0.03908	7

Time	Betweenness Centrality	Degree Centrality
Constitutionalism	0.036166	6
Revolution	0.019171	6
Socialism	0.009905	6
Evolution	0.055665	5
People	0.017734	5
Citizen	0.018227	4
Law	0.039702	3
Republic	0.014368	3
People's assembly	0.000821	2
National assembly	0.00821	2

**Appendix A:** Table of high betweenness centrality and degree centrality of LQC and CDX keywords, arranged in order of degree size.



Appendix B: Workflow chart.





The Journal of  
**HISTORICAL  
NETWORK  
RESEARCH**

MARILYN LEVINE

# Post WWI Chinese Revolutionary Leaders in Europe

*Journal of Historical Network Research 5 (2021) 187–232*

**Keywords** Chinese Communist Party, Zhou Enlai, Liu Shaoqi, Deng Xiaoping, Soviet Returned Leaders, network analysis

**Abstract** This article is an exploration of a leadership network of 133 Chinese revolutionary leaders who were trained in Western Europe and the Soviet Union during the 1920s. These leaders are in a dataset called Soviet Returned Leaders (SRL). The article utilizes quantitative, geospatial, and network analytical methods as well as proposing two approaches to expand network visualization utility and understanding. The results demonstrate that there are two groups within the SRL, namely those who went to both Western Europe and the Soviet Union (the Euro-Soviet Group) and those who went solely to the Soviet Union (the Soviet Group). The new network dissection techniques and network graphs reveal the importance of both well-known and relatively unknown individuals who may be potentially influential actors indicating that this initial historical network approach to biographical data affords an additional means of studying these important historical individuals.

## 1. Introduction\*

As he lay ill in 1923, Vladimir Lenin dictated his last published article which concerned his hopes for continuing the revolution. Disappointed in the prospects for revolution in the western countries, Lenin's new faith was in the eastern countries who were already "drawn into the revolutionary movement."<sup>1</sup> Thousands of miles away, writing in February 1945 at the end of another World War, Zheng Chaolin (鄭超麟 1901–1998), recounted his time in the Soviet Union in 1923 while Lenin was alive. Zheng was in Moscow for training in 1923, after three years in France as a worker-student.<sup>2</sup> His written account of twelve Chinese traveling to the land of revolution and the experiences they had in Moscow is compelling reading.<sup>3</sup> It was Zheng Chaolin's generational cohort who adopted communist ideology with a focus on Leninism. They led the Chinese revolution for three decades and ruled a nation for another four decades after the 1949 revolution. The Soviet experience was a dynamic episode in Chinese Communist Party (CCP) political history. Important revolutionaries among the Soviet Returned Leaders who are studied in this article include early CCP martyrs like Cai Hesen (蔡和森 1895–1931), Zhao Shiyan (趙世炎 1901–1927), Xiang Jingyu (向警予 1895–1928), Qu Qiubai (瞿秋白 1899–1935), and Zhang Tailei (張太雷 1898–1927). It also includes longer lived, powerful leaders such as Zhou Enlai (周恩來 1898–1976), Liu Shaoqi (劉少奇 1898–1969), Zhu De (朱德 1886–1976), Deng Xiaoping (鄧小平 1904–1997), Nie Rongzhen (聶榮臻 1899–1992), Cai Chang (蔡暢 1900–1990), Li Fuchun (李富春 1900–1975), Zhang Wentian (張聞天 1900–1976), Wang Ming (王明 1904–1974), Qin Bangxian (秦邦憲 1907–1946), Ye Jianying (葉劍英 1897–1986), Wang Jiaxiang (王稼祥 1906–1974), and Yang Shangkun (楊尚昆 1907–1998).

---

\* **Acknowledgements:** I would like to acknowledge two grants that allowed the initial creation of the Chinese Biographical Database which is the source of the SRL data subset. These grants are from the Chiang Ching-kuo Foundation and the Idaho Higher Education Major Research Grant program. I benefited from collaborations and help from Chen San-ching, Yves Chevrier, Christian Henriot, Jean-Louis Bouilly, Liu Guisheng, and Zhu Yuhe. For original entry in the CBD, the help of Zhou Baodi and Eric Barnes was excellent and in updating the database I am indebted to Howie X. Lan, John Bowen, Tan Hongxing, and Wu Yiwei.

**Corresponding author:** marilyn.levine@cwu.edu

- 1 Vladimir Il'ich Lenin, *The Lenin Anthology*, trans. Robert C Tucker (New York: WW Norton & Company, 1975), 745–55.
- 2 Zheng Chaolin, Interview, Shanghai, interview by Marilyn Levine and Zhu Yuhe, October 29, 1985; Zheng Chaolin, Interview, Shanghai, interview by Marilyn Levine and Zhu Yuhe, June 18, 1990.
- 3 Zheng Chaolin, *An Oppositionist for Life: Memoirs of the Chinese Revolutionary Zheng Chaolin*, trans. Gregor Benton (Atlantic Highlands, N.J.: Humanity Books, 1997), xxii.

This article will first present a historical introduction to the Western European and Soviet Union experiences of these Chinese leaders. Second, there will be an introduction to a Soviet Returned Leaders data subset (hereafter called SRL dataset), with 133 individuals, including its composition, features, attributes, and limitations. Third, the article will share results from quantitative, geospatial, and network analyses, followed by some conclusions. The research questions that underlie this study are twofold. First, what can these kinds of research approaches contribute to a new understanding of Chinese revolutionary leaders? Second, can historians develop methods that offer better clarity and visualization of an overall network and its internal structure-function?

## 2. Historical Context of the Soviet Returned Leaders

### 2.1 From Worker-Students to Communist Party Members, 1919–1923

In the early twentieth century, the Chinese youth were challenged as a generation at several key historical turning points. The civil service examination, the path to an official career, was abolished in 1905, alongside other reforms under the Qing Dynasty. The major challenge for the youth was to save the nation by “finding a new path” as the nascent republic that emerged after the 1911 Revolution was not robust and devolved into warlordism and intrusive foreign imperialism. Debates on eastern and western philosophies and exploration of ideologies were framed in a larger conversation about commitment to social reform or political revolution. To put it simply, the significance of selecting a life purpose during this era affected this generation in new ways, which expanded their options.<sup>4</sup> The post-WWI youth activism culminated in the nationwide May Fourth movement to protest against China signing the Versailles Treaty. Students were galvanized as activists to seek a better future by following new ideas. They could not rely on the older generation to lead them and felt a strong sense of zeal to resolve the fate of the nation.

---

4 Chow Tse-tung, *The May 4th Movement: Intellectual Revolution in Modern China*, 5th Printing edition (Cambridge, Mass.: Harvard University Press, 1960); Wang Y. C., *Chinese Intellectuals and the West, 1872–1949* (The University of North Carolina Press, 1974); Lin Yusheng, *The Crisis of Chinese Consciousness: Radical Antitraditionalism in the May Fourth Era* (Ann Arbor, Mich.: University of Michigan Press, 2000); Arif Dirlik, *The Origins of Chinese Communism* (New York: Oxford University Press, 1989); Alexander Pantsov, *Bolsheviks and the Chinese Revolution 1919–1927* (Honolulu: University of Hawai'i Press, 2000); Peter Gue Zarrow, *Anarchism and Chinese Political Culture* (New York: Columbia University Press, 1990); Hans J Van de Ven, *From Friend to Comrade: The Founding of the Chinese Communist Party, 1920–1927* (Berkeley Calif.: University of California Press, 1991).

Among the many paths of exploration, the idea of study abroad was the vision of four prominent educational leaders, Li Shizeng (李石曾 1881–1973), Cai Yuanpei (蔡元培 1868–1940), Wu Zhihui (吳稚暉 1865–1953), and Wang Jingwei (汪精衛 1883–1944) who promoted the Travel to France Diligent-Work Frugal-Study Movement (赴法勤工儉學運動, work-study movement). At the end of WWI, the work-study movement was advanced due to the decimation of the French male population, which led these educational leaders to believe that there would be multiple openings in French factories. The idea was to encourage the Chinese youth to work with their hands alongside the oppressed laborers, earn money for a college education, and gain technical and scientific degrees to return to China and modernize the nation. Beginning in 1919, over 1,800 youth heeded this call and provincial governments even set up stipends and preparatory schools to encourage this path.

Although these four leaders had solid connections in France, Germany, and Belgium, the success of this scheme was limited by post-war economic realities in France. The worker-students lived in dire poverty and were politicized by several events, as well as their exploration of ideologies, particularly anarchism, communism, nationalism, and social democracy. During 1921, at the nadir of their impoverished living conditions, there were three political struggles that involved the worker-students: in February, led by Cai Hesen, a faction in the city of Montargis mobilized the February 28<sup>th</sup> Movement that advocated Chinese government stipends for the worker-students so that they could pursue studies and live adequately. The Montargis faction were the most progressive Marxists, and were opposed by a group led by Zhao Shiyan who believed in the work-study ethos of mutual-aid and self-sufficiency. Both factions organized others into Marxist study societies and other youth groups, and due to the unification efforts by Zhao Shiyan, the two factions merged to fight the government in the second struggle of 1921 – the Loan Struggle, where they successfully opposed a French loan to the Chinese warlord government of Xu Shichang (徐世昌 1855–1939). The Loan Struggle was notable for the public physical beating of the Chinese Foreign Minister's secretary who was forced into signing a renunciation of a possible loan. The *contretemps* with the Chinese Minister in both of these struggles did not bode well for the third struggle, known as the Lyons Incident in the fall of 1921, where the increasingly desperate worker-students were sidelined by the founders of the work-study movement who brought over 100 students from China to matriculate at the newly created Sino-French Institute at the University of Lyon. The angry worker-students mobilized and occupied a dormitory at Lyon University, whereupon over 100 of them were arrested and shortly thereafter deported back to China. The deported worker-students included Cai Hesen, Chen Yi (陳毅 1901–1972), and Li Lisan (李立三 1899–1967). Among the leaders who escaped from this imprisonment was Zhao Shiyan, who organized the European Branches of the Chinese Communist Organizations (旅歐中國共產主義組織, est. 1922, ECCO) during the summer and winter of 1922. The ECCO had branches in France, Germany, and Belgium. Most of the members belonged to the Youth

Corps, and members had to be formally reviewed for their ideological commitment in order to enter the Communist party.<sup>5</sup> These radicalized worker-students, particularly those who joined the ECCO, had an entire cohort of their membership return to China via study in the Soviet Union. As mentioned above, they contributed some of the most important CCP leaders of the Chinese revolution and post-revolution period.

The development of the ECCO needs to be contextualized by understanding several historical trends: the deterioration of the work-study movement, the Three Struggles of 1921, and the unique political milieu of Chinese youth activism in France and other European countries. Altogether these five political parties formed between 1922 and 1924. The five parties were the Anarchist Party or Surplus Society, (工餘社, est. 1922, GYS), the ECCO, the Chinese Social Democratic Party (中國社會民主黨, est. 1922, SDP), the European Branch of the Chinese Nationalist Party (中國國民黨旅歐支部, est. 1923, EGMD), and the Chinese Youth Party (青年黨, est. 1923, QND).<sup>6</sup>

Chinese political party activists were able to learn about Western ideologies, and to conduct recruitment, propaganda, and agitation activities in an environment with more access to materials and greater freedom. Chinese activists of all five parties were, however, closely watched by the French Sûreté. The extensive French surveillance on the Chinese activist community includes political analyses, reports from Asian and French agents in the field, and captured documents

- 
- 5 Marilyn A. Levine, *The Found Generation: Chinese Communists in Europe during the Twenties* (Seattle: University of Washington Press, 1993); Paul Bailey, "The Chinese Work-Study Movement in France," *The China Quarterly*, no. 115 (1988): 441–61, <http://www.jstor.org/stable/654865>; Thomas Kampen, "Chinese Communists in Austria and Germany and Their Later Activities in China," *Asian and African Studies* XI, no. 1–2 (2007): 21–30; Tsinghua University Faculty Research Unit on the History of the Communist Party, *Fu Fa Qingong jianxue yundong shiliao* 赴法勤工儉學運動史料 [*Documents on the Travel to France Work-study movement*] (Beijing: Beijing chubanshe, 1979); Zhang Yunhou 張允侯, Yin Xuyi 殷敘彝, and Li Junchen 李峻晨, *Liu Fa Qingong jianxue yundong* 留法勤工儉學運動 [*The Travel to France Work-study movement*], 2 vols. (Shanghai: Shanghai renmin chubanshe, 1980); Chen San-ching 陳三井, *Qingong jianxue yundong* 勤工儉學運動 [*The Diligent-Work Frugal-Study Movement*] (Taipei: Zhengzhong shuju, 1981).
  - 6 Marilyn A. Levine and Chen San-ching, *The Guomindang in Europe: A Sourcebook of Documents* (Berkeley, Calif.: Institute of East Asian Studies, University of California, Berkeley, Center for Chinese Studies, 2000); Marilyn A. Levine and Chen San-ching, "Communist-Leftist Control of the European Branch of the Guomindang, 1923–1927," *Modern China* 22, no. 1 (1996): 62–92, <http://www.jstor.org/stable/189290>; Shiu Wentang, "Les Organisations Politiques Des Étudiants Chinois En France Dans l'entre-Deux Guerres." (Paris, Université de Paris VII, 1990).

(letters, political agendas and meeting minutes, as well as other confiscated materials).<sup>7</sup>

In terms of the ECCO members, they saw their activities in France as training for their own revolution in China. Beginning in 1923 the CCP ordered several rounds of ECCO members to return to China through the Soviet Union, where the intention was to gain greater ideological depth and activist training in the cauldron of contemporary revolution. There were also more than a dozen Chinese activists in France who returned directly to China but visited the Soviet Union for conferences and/or study. The Soviet Union phase of their leadership development meant expanding their knowledge and joining with others in forming a broad network of revolutionaries trained abroad, which is discussed in the following sections.

## 2.2 The Soviet Experience and Bolshevik Loyalties, 1923–1926

Many of the Chinese who came from Western Europe pursued studies at the Toilers of the East University in Moscow (KUTV). Their arrival in the early 1920s occurred while Lenin was ill and before Stalin was able to exert his increasing power to affect the international education of foreign revolutionaries inside the country.<sup>8</sup>

During this period the Soviet Union was a poverty-stricken country, that had been devastated by the First World War, the Bolshevik Revolution, and a long Civil War. Discussing the conditions for several institutions, Josephine Fowler has explained that “[A]djusting to life in a city whose shortages of housing, fuel, and food were widespread, and where the students received far better accommodations than ordinary Muscovites, and whose culture was alien to both Chinese and

---

7 Levine and Chen, *The Guomindang in Europe*, 17–22; Marilyn Levine, “Conducting Research in the French Archives on Chinese Radicalism,” *Republican China* 22, no. 2 (April 1997): 93–102.

8 Alexander Pantov [sic] and Daria A. Spichak, “New Light From the Russian Archives: Chinese Stalinists and Trotskyists at the International Lenin School in Moscow, 1926–1938,” *Twentieth-Century China* 33, no. 2 (2008): 29–50, <https://doi.org/10.1353/tcc.0.0001>; Josephine Fowler, *Japanese and Chinese Immigrant Activists: Organizing in American and International Communist Movements, 1919–1933* (Rutgers University Press, 2007), <https://doi.org/10.2307/j.ctt5hj7ms>; Yueh Sheng, *Sun Yat-Sen University in Moscow and the Chinese Revolution; a Personal Account*, International Studies, East Asian Series Research Publication; No. 7 (Lawrence, KS: Center for East Asian Studies, University of Kansas, 1971); Anna Belogurova, “Networks, Parties, and the ‘Oppressed Nations’: The Comintern and Chinese Communists Overseas, 1926–1935,” *Cross-Currents: East Asian History and Culture Review* 6, no. 2 (2017): 558–82, <https://doi.org/10.1353/ach.2017.0019>.

Japanese students was immensely difficult. Along with room and board, students received clothing, shoes, textbooks, and stationery.”<sup>9</sup>

Recounting the level of support and the attitudes that the Chinese students encountered, Zheng Chaolin noted that “Among the Soviet masses, young people accepted the [need for] revolutionary education, and there was no pretentiousness. But ordinary people still had the concept of rejecting foreigners.”<sup>10</sup> Zheng also spoke of having several opportunities for cultural interchange and understanding. The ideological education the Chinese received, according to Zheng, was not in-depth, and there was no serious attempt to train the Chinese to learn the Russian language.

Few of those who arrived before 1924 were targets of the charged political atmosphere that would emerge in Moscow during the late-1920s – a fate that awaited *other* Chinese comrades who arrived after the mid-1920s. As Pantsov and Spichak concluded about these Soviet international educational institutions: “A history of the Chinese international schools in the USSR highlights the significance of the Soviet factor in the ideological and political evolution of the Chinese Communist movement in the 1920s and 1930s. The Soviet Communists tried to create the CCP cadres in their own image and likeness. To pursue this goal they financed, directed, and supervised the Chinese Communists. For this reason, the Stalinization of the Bolshevik Party that took place from 1924 to 1929 had a tremendous impact on the CCP. The Chinese students in Moscow were the first to feel it.”<sup>11</sup>

Most of the Soviet Returned Leaders explored in this study were exemplars of adaptive revolutionary strategies and strongly influenced by Lenin, particularly in the concepts of party vanguard leadership, party discipline, and the need for a world revolution. As one of the most successful early revolutionaries, Zhao Shiyan proclaimed, “Lenin was the author of the ‘October Revolution,’ – the start of the World Revolution, [he is] the leader and guide for the entire world’s proletariat, peasant masses, and all those who are oppressed. *Currently, Leninism is our banner for the oppressed; Leninism is our weapon; the entire world revolution – is our responsibility* [Emphasis Zhao’s].”<sup>12</sup> Returning from the Soviet Union these CCP revolutionaries would continue to work towards a Bolshevik model of revolution in China.

---

9 See Fowler, *Japanese and Chinese Immigrant Activists*, 70.

10 Zheng Chaolin, Interview, October 10, 1985, Beijing.

11 Alexander V. Pantov [sic] and Daria A. Spichak, “New Light from the Russia Archives.”

12 Zhao Shiyan, “Lieningzhuyi zhi lilun yu shiji” 列寧主義之理論與實際 [The Theory and Practice of Leninism], in *Zhao Shiyan wenji* 趙世炎文集 (Chengdu: Sichuan remin chubanshe, 1984), 393–94.

### 3. Composition, Features, Attributes, and Limitations of the SRL Dataset

The SRL is a dataset of 133 individuals and 202 attributes from the Chinese Biographical Database (CBD) that includes 2,109 individuals and 840 attributes. The CBD was created in 1997–1998 and was available online for queries and reports until 2006. Current CBD work revolves around several datasets to analyze various individuals and groups who had common affiliations. The categories of attributes include: 1. Basic Biographical information (full name, Chinese characters, traditional and simplified, birth year, birth date, death year, lifespan, birth city, birth province, latitudes/longitudes for birth cities, gender, web links, and 2,093 comments); 2. Career; 3. Affiliations; 4. Education; 5. Positions; 6. Youth Activities; and 7. Historical Events. The CBD also includes other data tables such as Alternative Names, Family, Locator, and Sources, which have multiple citations for most individuals.

The CBD and SRL use 2-mode data. This is similar to the Southern Women study by Davis et al. published in 1941.<sup>13</sup> Two-mode data are simply a table of individuals in rows and attributes in columns which, if graphed as a network, produce two types of nodes or points, i.e., individuals or attributes, but with connections only to the other type and not to the same type, e.g., there would be no data on person versus person since this was not recorded in the original table. Conversion to a 1-mode table thus results in only individuals versus all other individuals, or attributes versus all other attributes. Since this is a biographical based network study, the simple choice was to use those two- or one-mode measures that yielded the best concordance with history.<sup>14</sup>

The SRL dataset has two primary subgroups: A Euro-Soviet group (N = 64) who had experiences in Europe and the Soviet Union; and a Soviet group (N = 69) who only traveled to the Soviet Union for training. Because of the importance of these revolutionaries to Chinese history, the idea is to understand the network

---

13 Allison Davis et al., *Deep South: A Social Anthropological Study of Caste and Class* (Chicago: University of Chicago Press, 1941). For a deeper insight about how this data has been utilized, there are 21 studies reviewed in Linton Freeman, "Finding Social Groups: A Meta-Analysis of the Southern Women Data," in *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, by Ronald L. Breiger, Kathleen M. Carley, and Philippa Pattison (Washington, D.C.: National Academies Press, 2003), 1–39.

14 Stephen P. Borgatti, "Two-Mode Concepts in Social Network Analysis," *Encyclopedia of Complexity and System Science* 6 (2009): 8279–91. For a more in-depth discussion of the CBD, including an extended discussion of two-mode data, documentation, and composition, see "Biography for Historical Analysis: A Chinese Biographical Database" in this issue of the *Journal of Historical Network Research*. The CBD [Chinese Biographical Database] should not be confused with the CBDB [China Biographical Database], which is developed by Harvard University, Academia Sinica, and Peking University and focuses on pre-modern China.



by analyzing a large number of attributes. Among the general features of the SRL dataset are:

- The largest subgroup within the SRL are the 121 members of the Chinese Communist Party. Over a third of these individuals were Central Committee members and 22 percent joined the Politburo.
- There are 11 females (8 percent) in the SRL.
- While KUTV has 104 of the 133 SRL individuals, there were matriculations at other Soviet institutions such as 10 individuals who attended the International Lenin School.
- There are 31 members of the Chinese Nationalist Party (國民黨 GMD) in the SRL, which includes Jiang Jieshi (蔣介石 1887–1975), and his son Jiang Jingguo (蔣經國 1910–1988). Another GMD member was Chen Chunpu (陳春圃 1900–1966) who followed Wang Jingwei.<sup>15</sup> The majority of the GMD members in this group were Communists who joined because of the United Front (1924–1927), which was a collaboration between the GMD and the CCP. The collaboration's main focus was the Northern Expedition (1926–1927), a military expedition to eradicate the warlords.
- There are nine members of the so-called “Twenty-eight Bolsheviks”: Chen Changhao (陳昌浩 1906–1967), Qin Bangxian, Shen Zemin (沈澤民 1900–1933), Wang Jiaxiang, Wang Ming, Wang Shengdi (汪盛荻 1899–1950), Xia Xi (夏曦 1901–1936), Yang Shangkun, and Zhang Wentian. These leaders had a great impact, but based on the dates and locations of leadership, as demonstrated by Thomas Kampen, it does not appear that they were either as efficacious as a block of leaders or necessarily identified exclusively with their “group”, as is often depicted.<sup>16</sup> The current study supports Kampen's analysis.
- There are seven Trotskyists including: Zheng Chaolin, Yin Kuan (尹寬 1897–1967), Liu Renjing (劉仁靜 1902–1987), Wang Fanxi (王凡西 1907–2002), Peng Shuzhi (彭述之 1895–1983), Shi Yisheng (施益生 1902–1993), and Wang Zekai (汪澤楷 1894–1959). Although limited by disunity and incarcerations, the Trotskyists played important roles in both ideological and labor activities throughout the 1920s and 1930s.<sup>17</sup>

15 Wang Jingwei led the leftist GMD and later headed the collaborationist government during WWII. Chen Chunpu was one of his chief officials.

16 Thomas Kampen, *Mao Zedong, Zhou Enlai and the Evolution of the Chinese Communist Leadership* (Denmark: NIAS Pub., 2000). Kampen argues that the focus of internal CCP dynamics during the Jiangxi period should be more on the role of leaders such as Zhou Enlai or Zhang Wentian, and not a dichotomy between the so-called Twenty-eight Bolsheviks and Mao Zedong. Kampen clearly demonstrates this group had powerful leaders, but they were not necessarily a group that acted as a unified bloc.

17 For overviews of Chinese Trotskyites, and particular attention to activities in the Soviet Union see: Zheng Chaolin, *An Oppositionist for Life*; Gregor Benton, *Prophets Unarmed: Chinese Trotskyists in Revolution, War, Jail, and the Return from Limbo*, Historical Ma-

Every individual in a network potentially has a very large number of attributes that determine their position in an ever-evolving dynamic network. A systematic exploration of all these attribute correlations is desirable but is limited by the human mind in terms of calculation. Limiting these attributes to a few, or even one, necessitates subjective choices and therefore weighting. The opportunity here is that the greater the number of attributes the greater the statistical strength. Furthermore, choosing obvious differentiating attributes as primary or sole characteristics for a network may miss minute or inconspicuous attributes that could be critical to a network structure and its operation. This study uses the complete biographical record with no *a priori* weighting or elimination of rare attributes in order to let the analyses reveal the network *in toto*, so that one can then explore what was revealed in terms of importance.

A current limitation of the SRL at this time is the absence of time series data. The SRL is also not a full population study of all those who studied or attended conferences in the Soviet Union. However, the SRL includes a broad sample of medium and lesser-known individuals from various groups as well as many prominent political leaders (see Appendix 2 for full listing).

## 4. Quantitative and Geospatial Analyses of the SRL

### 4.1 Analysis of Birth Year, Death Year, and Lifespan

A brief statistical analysis of birth year, death year, and lifespan illustrates the temporal cohesiveness of this cohort. Table 1 shows the range of birth year within 1–2 years of each other for both groups, with a later death year for the Soviet group.

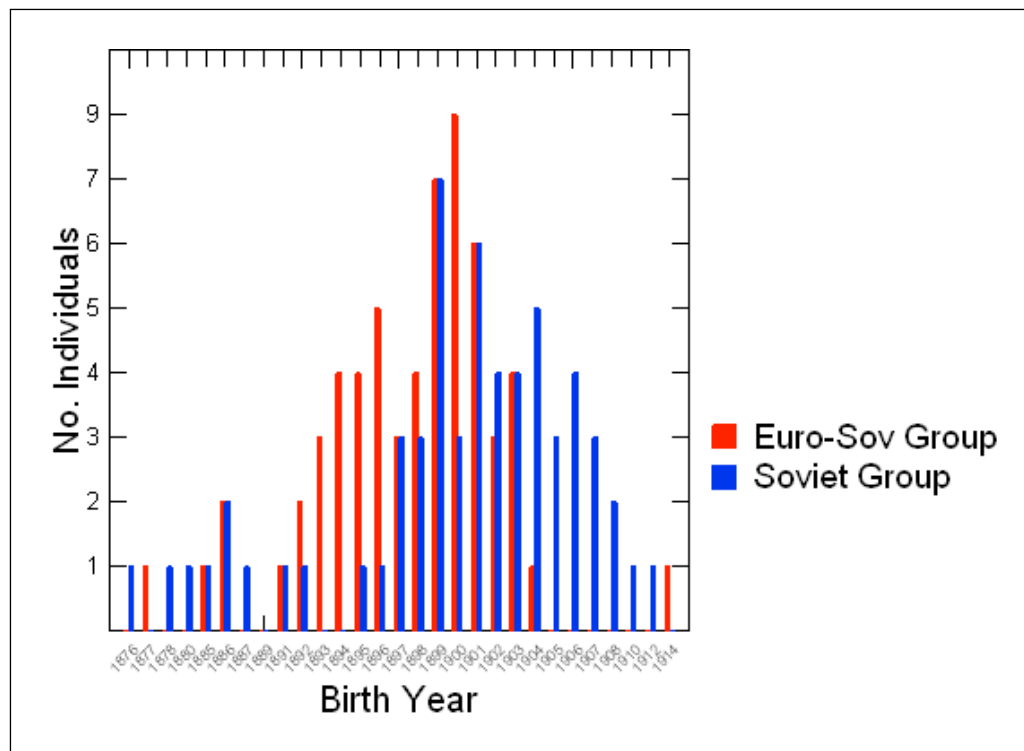
The distribution in Fig. 1 shows that the Soviet group was generally a few years younger than the Euro-Soviet group. For example, after 1904 there are 19 individuals born in the Soviet group and only 1 member of the Euro-Soviet group. Perhaps the most dramatic difference is in Fig. 2 where there is a relatively large segment of the Euro-Soviet group who died between 1924 and 1931, especially during the capture and execution of CCP members in the April 12<sup>th</sup> coup of 1927. Twenty-one individuals – one-third of the Euro-Soviet cohort – perished, often after being tortured. Eleven members of the Soviet group, or 17 percent, died during the same period. In contrast, after 1931 the Soviet group members or 22 percent (N = 17) of their members died in encirclement campaigns, war against the

---

terialism Book Series (Leiden, The Netherlands: Brill, 2015); Gregor Benton, *China's Urban Revolutionaries Explorations in the History of Chinese Trotskyism, 1921–1952* (Atlantic Highlands, N.J.: Humanities Press, 1996).

Group No.	Individuals	Min	Max	Median	Mean	SD
Birth Year-Euro-Soviet	61	1877	1914	1899	1897	5.527
Birth Year-Soviet	69	1876	1912	1901	1899	7.239
Death Year-Euro-Soviet	58	1924	1998	1961	1957	26.29
Death Year-Soviet	69	1927	2008	1967	1963	24.97
Lifespan-Euro-Soviet	57	26	97	68	61	25.33
Lifespan-Soviet	69	26	101	69	63	23.47

**Tab. 1** Comparison of birth year, death year, and lifespan for Euro-Soviet and Soviet groups in the SRL dataset. The Euro-Soviet group had missing data for individuals: birth year = 3, death year = 6, and lifespan = 7.



**Fig. 1** SRL comparison of Euro-Soviet and Soviet groups for birth year.

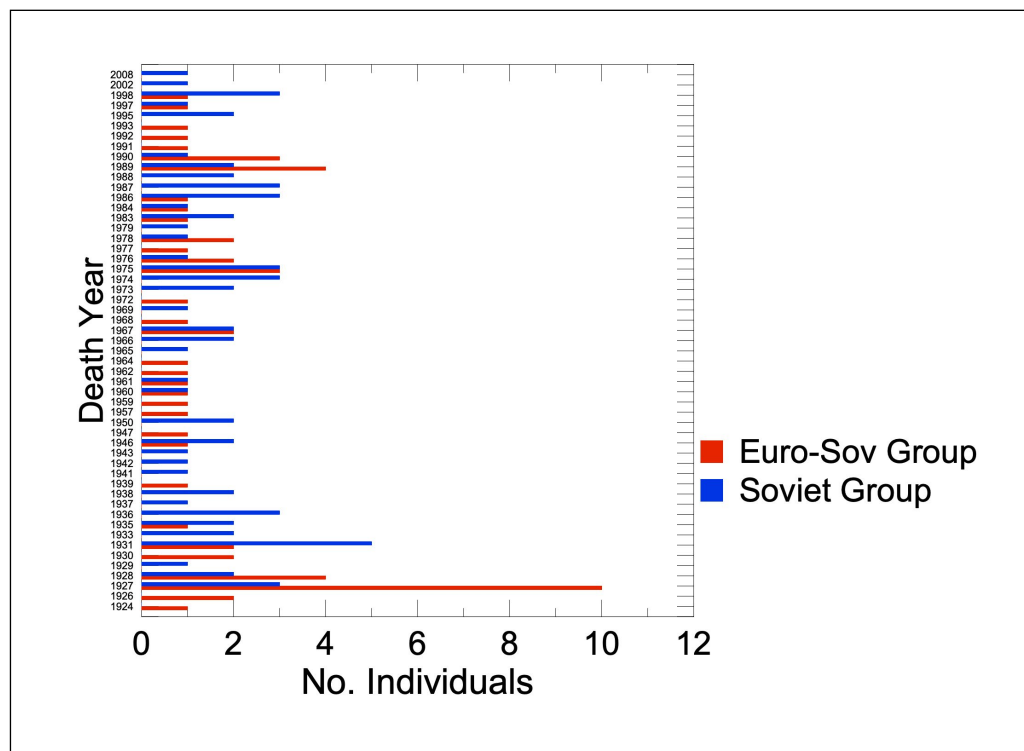
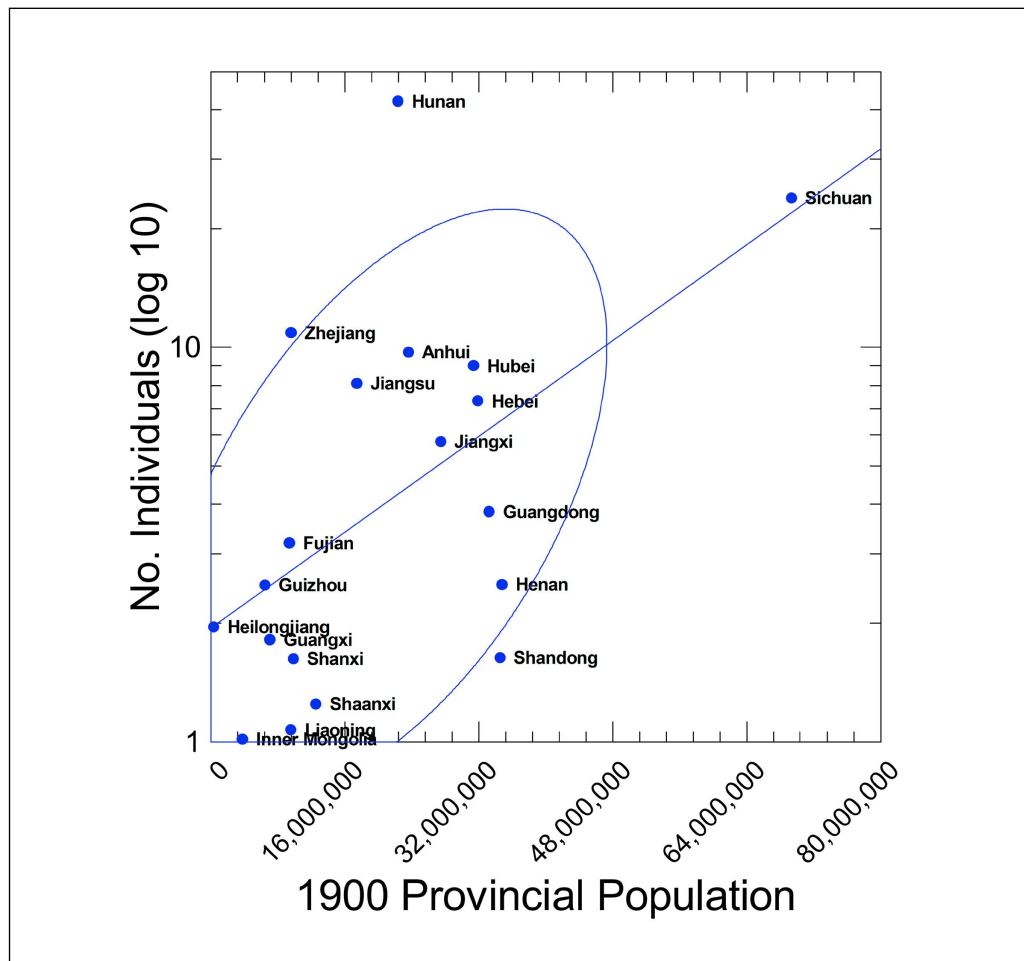


Fig. 2 SRL comparison of Euro-Soviet and Soviet groups for death year.

Japanese, and the Civil War, while the Euro-Soviet group lost only 4 individuals or 6 percent of their members. How were individuals impacted by the loss of their comrades during these periods? They likely had no time for grief or retraining but were concerned with survival and strategic adaptation; however, the sense of loss appears to be profound, as can be seen from commemorations at the time and decades later.

#### 4.2 Geospatial Analysis of Regional Origins and Birth Cities

In examining the Soviet Returned Leaders, in general, provincial birth origin did correlate with the provincial population and showed a fairly consistent contribution from most provinces. A scatterplot of individuals from their respective provinces versus population showed the general relationship of increased participation with increases in provincial population (see Fig. 3). Performing a regression analysis (least squares) between the total number of individuals from their province and the provincial population demonstrates a very significant relationship ( $p$ -value = 0.041). It is notable that Hunan ( $N = 34$ ) is an outlier, as it had 600% higher participation than other provinces. Sichuan ( $N = 25$ ) had significant leverage due to its much larger population.

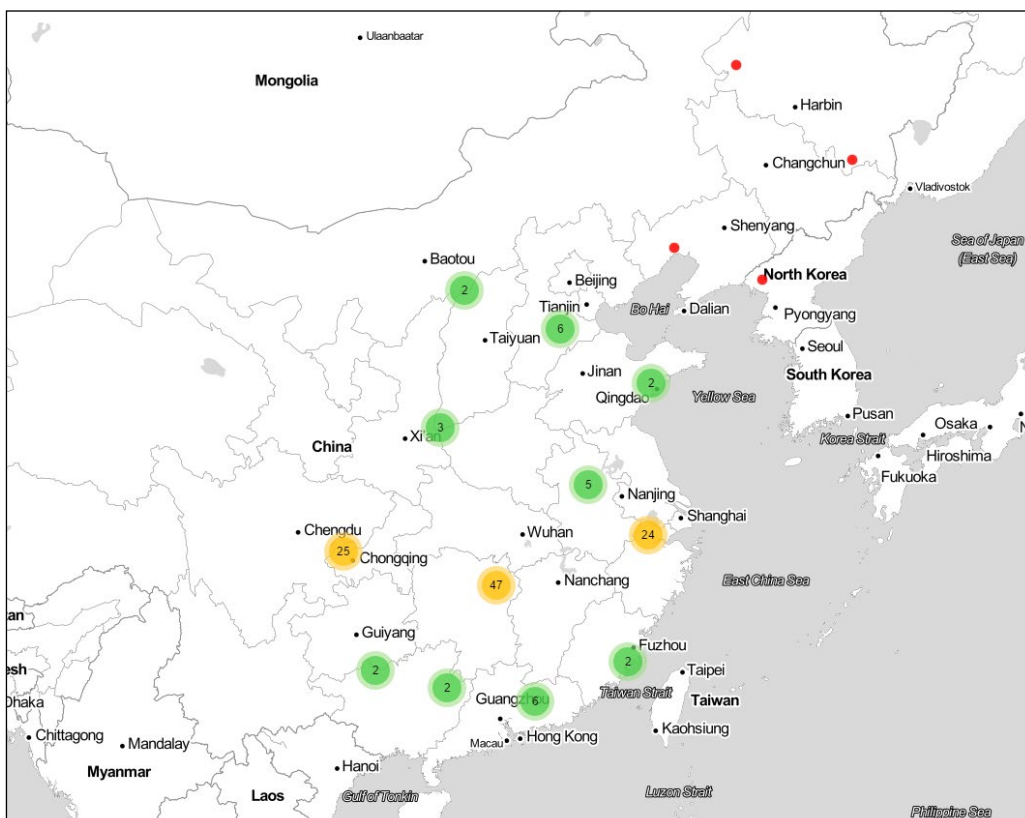


**Fig. 3** Individual participation ( $N = 132$ ) vs. 1900 provincial population. Confidence sample ellipse chosen at  $p = 0.6827$  with linear regression line. Source: Jan Lahmeyer, "Population Statistics: Historical Demography of All Countries, Their Divisions and Towns," 2006, <http://www.populstat.info>, retrieved June 30, 2019.

One possibility for the provincial role of Hunan and Sichuan might have been the clustering of participants at high population centers. Comparing the physical sizes of Hunan and Sichuan (respectively 210,000 and 485,000 square kilometers) with the 1900 population, one obtains a respective density of 106 and 142 individuals per square kilometer. There was an unusual number of Hunanese participants clustered around the broader Changsha region, while the Sichuan participants were more dispersed. In fact, Changsha and closely surrounding regions have the highest rate of the Soviet Returned Leaders in this cohort ( $N = 47$ ) as can be observed in the cluster map in Fig. 4.

In examining the leadership from Hunan and Sichuan in historical terms of the Soviet Returned Leaders, it is clear that Hunan is an outlier in terms of the numbers of people involved who were sustained in their leadership roles throughout the twentieth century, particularly in the 1920s through the mid-1930s, and for some lasting beyond 1949 (e.g., Liu Shaoqi, Li Fuchun, and Cai Chang). The pattern of leadership roles for the Sichuan contingent of the Soviet Returned Leaders also reflects the high leverage of Sichuan (e.g., Deng Xiaoping, Nie Rongzhen, and Zhu De).

Figure 4 displays the distribution of birth city clusters of individuals. There are three centers where birth cities cluster together. From east to west in the middle of China there are 24 individuals from Jiangsu/Zhejiang, 47 individuals from Hunan and surrounding regions, and 25 Sichuanese from Chengdu/Chongqing. Provinces are large areas, and individual birth cities may be proximate to a border or be composed of cities clustered around a river, as is the case with birth cities along the Xiang River near Changsha. Birth cities can play an important role for individuals in terms of proximity to education and youth group activities in historical networks.



**Fig. 4** Birth Cities clustered distribution of individuals for SRL (N = 130). Each red point represents a single individual. Three individuals missing.

## 5. Network Analyses of the SRL

### 5.1 Centralities

The centrality analyses shown in Table 2 are from 133 individuals by 202 attribute data (2-mode) and display the top 20 individuals sorted in descending order for the four measures of degree, Eigenvector, closeness, and betweenness centrality.<sup>18</sup>

There are some suggestive patterns that emerge in this analysis. First, Zhou Enlai is the highest ranked leader for all four measures and is followed in the top five ranks by Liu Shaoqi, Deng Xiaoping, and Zhu De, with Nie Rongzhen placed second in the first three metrics. The Euro-Soviet members consistently ranked higher in the top ten of the first three measures. Within the top ten rankings of the four centralities the Soviet group, Liu Shaoqi is joined by Zhang Wentian and Qu Qiubai. The betweenness measure, which measures how frequently an individual is placed in between other nodes has different rankings in the fourth column of Table 2. Particularly intriguing is the inclusion of GMD leaders Jiang Jingguo, Chen Chunpu, and Cao Chengde (曹承德 1903–1978), a female activist who later went to Taiwan.

As mentioned, Zhou Enlai is the most central individual in terms of all four centralities. He appears to function (as will be shown in the graphs below) in the network as a “boundary spanner,”<sup>19</sup> someone who was a key resource for communicating with other groups. In his research on organizational leadership teams, Michael Tushman explained, “Individuals filling these roles [as boundary spanners] are capable of translating contrasting coding schemes and therefore of acting as boundary spanners between the work unit and external information areas. Thus, information may flow...in a two-step fashion through a set of key persons who channel this information to their colleagues...These opinion leaders were sought after for advice and, in turn, influenced the decisions of the less active members of the social system.”<sup>20</sup> Certainly, Zhou Enlai personified this role of spanning communication and influencing others throughout his career as a journalist in the early 1920s traveling throughout Europe, as a conduit to the Com-

---

18 Stephen P. Borgatti and Martin G. Everett, “Network Analysis of 2-Mode Data,” *Social Networks* 19, no. 3 (August 1, 1997): 243–69, [https://doi.org/10.1016/S0378-8733\(96\)00301-2](https://doi.org/10.1016/S0378-8733(96)00301-2).

19 Discussions of Zhou Enlai and the importance of Cai Chang and Li Fuchun were presented at a workshop, Marilyn Levine, “Revolutionary Roads: An Integrative Analysis Utilizing a Chinese Biographical Database” (Workshop, ERC ENP-China Project “Elites, Knowledge, and Power in Modern China,” Aix-en-Provence, France, October 7, 2019), <https://enepchina.hypotheses.org/>.

20 Michael L. Tushman, and Thomas J. Scanlan “Special Boundary Roles in the Innovation Process,” *Administrative Science Quarterly* 22, no. 4 (December 1977): 590–91.

Individual	Degree	Individual	Eigenvector	Individual	Closeness	Individual	Betweenness
<i>Zhou Enlai</i>	0.30542	<i>Zhou Enlai</i>	0.13697	<i>Zhou Enlai</i>	0.61873	<i>Zhou Enlai</i>	0.06836
<i>Nie Rongzhen</i>	0.21182	<i>Nie Rongzhen</i>	0.13329	<i>Nie Rongzhen</i>	0.58920	<i>Zhu De</i>	0.03495
<b>Liu Shaoqi</b>	0.20690	<i>Deng Xiaoping</i>	0.12623	<b>Liu Shaoqi</b>	0.58772	<i>Deng Xiaoping</i>	0.03247
<i>Deng Xiaoping</i>	0.20197	<b>Liu Shaoqi</b>	0.12377	<i>Deng Xiaoping</i>	0.58625	<b>Liu Shaoqi</b>	0.03026
<i>Zhu De</i>	0.20197	<i>Cai Chang</i>	0.12370	<i>Zhu De</i>	0.58625	<b>Jiang Jingguo</b>	0.02971
<i>Cai Chang</i>	0.18227	<i>Li Fuchun</i>	0.12359	<i>Cai Chang</i>	0.58045	<b>Wu Liangping</b>	0.02802
<b>Zhang Wentian</b>	0.16749	<i>Zhu De</i>	0.12077	<b>Zhang Wentian</b>	0.57617	<b>Qu Qiubai</b>	0.02568
<i>Li Fuchun</i>	0.16256	<b>Zhang Wentian</b>	0.11227	<i>Li Fuchun</i>	0.57475	<b>Zhang Wentian</b>	0.02427
<b>Dong Biwu</b>	0.15764	<b>Dong Biwu</b>	0.10954	<b>Dong Biwu</b>	0.57335	<b>Liu Yaxiong</b>	0.02294
<b>Qu Qiubai</b>	0.15764	<b>Zhang Guotao</b>	0.10747	<b>Qu Qiubai</b>	0.57335	<b>Cao Chengde</b>	0.01902
<b>Ye Jianying</b>	0.14778	<i>Li Weiha</i> n	0.10746	<b>Ye Jianying</b>	0.57056	<i>Nie Rongzhen</i>	0.01883
<i>Li Weiha</i> n	0.14286	<i>Wang Ruofei</i>	0.10741	<i>Li Weiha</i> n	0.56917	<b>Kang Sheng</b>	0.01801
<b>Zhang Guotao</b>	0.14286	<b>Qu Qiubai</b>	0.10655	<b>Zhang Guotao</b>	0.56917	<b>Guo Huaruo</b>	0.01697
<i>Zhao Shiyan</i>	0.13793	<i>Li Lisan</i>	0.10470	<i>Zhao Shiyan</i>	0.56780	<b>Han Guang</b>	0.01618
<b>Ulanfu</b>	0.13300	Guo Longzhen	0.10353	<b>Ulanfu</b>	0.56643	<b>Ulanfu</b>	0.01588
Guo Longzhen	0.12808	<b>Ren Bishi</b>	0.10351	Guo Longzhen	0.56506	<i>Chen Qixiu</i>	0.01587
<b>Jiang Jingguo</b>	0.12808	<b>Ulanfu</b>	0.10302	<b>Jiang Jingguo</b>	0.56506	<b>Chen Boda</b>	0.01554
<i>Li Lisan</i>	0.12808	<b>Zhang Tailei</b>	0.10279	<i>Li Lisan</i>	0.56506	<i>Zhao Shiyan</i>	0.01439
<i>Wang Ruofei</i>	0.12808	<b>Yang Shangkun</b>	0.10162	<i>Wang Ruofei</i>	0.56506	<i>Cai Chang</i>	0.01418
<b>Han Guang</b>	0.12315	<i>Lin Wei</i>	0.10048	<b>Han Guang</b>	0.56370	<b>Chen Chunpu</b>	0.01359

**Tab. 2** Four normalized centralities for Soviet Returned Leaders group obtained from 2-mode analyses showing highest 20 individuals. Centralities and names are sorted for highest scores. Italicized names are Euro-Soviet individuals and bolded names are Soviet individuals.



munist International in the 1930s, exhibiting political and military leadership from the mid-1920s through the 1940s, and in his role as the Foreign Minister expanding the PRC's legitimacy on the global stage. These positions attest to his capabilities to go beyond normal boundaries and to communicate within complex situations.<sup>21</sup>

There are also some individuals with high centralities that should be explored further, such as Nie Rongzhen,<sup>22</sup> who may have played a more significant role in CCP events than is commonly recognized, and Liu Shaoqi,<sup>23</sup> who is acknowledged as a key leader, but understudied. Particularly notable is the unassuming, yet nonetheless important wife-husband team, Cai Chang and Li Fuchun. Both Cai Chang and Li Fuchun have a high degree centrality and a high Eigenvector centrality, which indicates that a node is connected to individuals who are themselves highly connected. They also have very high closeness measures which emphasize the distance of an actor to others in the network by focusing on the distance from each actor to all others. However, in examining betweenness there is a lower ranking for Cai and Li. Betweenness is often tied to the idea of being a power broker. Therefore, as Cai Chang is ranked 19 ( $N = 0.1418$ ) and Li Fuchun is ranked 37 ( $N = 0.0095$ ) this might explain why this couple have not been extensively studied or remarked upon as a "power couple." Research might be conducted on how Cai and Li actually exercised their power and advantages within their networks.<sup>24</sup> Subsequent measures below will also demonstrate the impor-

- 
- 21 Zhou Enlai's long career and importance to the CCP has generated many books and articles, and he is cited in thousands of articles on pre- and post-1949 Chinese history. The general consensus is that Zhou was a key pragmatist, during a turbulent regime, who provided an anchor of rationality, but managed to maintain his position, by not attempting to take supreme power. This general idea is supported by this study, which attempts to deepen the understanding of Zhou Enlai's relationships with other leaders in the important Soviet returned leaders' network. For some sample biographies see: Kai-yu Hsu, *Chou En-Lai: China's Gray Eminence*. (Garden City, N.Y.: Doubleday, 1968); Chae-Jin Lee, *Zhou Enlai: The Early Years* (Stanford, Calif: Stanford University Press, 1994); Dick Wilson, *Zhou Enlai: A Biography* (New York, N.Y.: Viking, 1984); Gao Wenqian, Peter Rand, and Lawrence R. Sullivan, *Zhou Enlai: The Last Perfect Revolutionary* (New York: Public Affairs, 2008).
  - 22 Nie Rongzhen, *Nie Rongzhen Huiyilu (Memoirs of Nie Rongzhen)*, 3 vols. (Beijing: People's Liberation Press, 1983); Zemin Jiang, Interview, Beijing, interview by Marilyn Levine and Liu Guisheng, October 25, 1985.
  - 23 Lowell Dittmer, *Liu Shao-Ch'i and the Chinese Cultural Revolution: The Politics of Mass Criticism* (Berkeley: University of California Press, 1974). Although mentioned, mostly as a key victim of the purge during the Cultural Revolution, Liu Shaoqi has not been extensively studied. His role in developing political training and policy decisions, and as can be seen in this study, his high centrality scores show he was a very important CCP leader.
  - 24 In the multivariate analysis dendrogram (data not shown), Cai Chang and Li Fuchun form their own dyad when measuring distance via similarities of attributes. What is notable is that spousal information was not used as an attribute in these analyses. Ref-

tance of Cai and Li, who were present in all phases of the CCP revolution and post-1949 regime, as well as holding high-level party posts.

## 5.2 Network Cohesion, Scale, and Core-Periphery Analyses

The remainder of SRL analyses will be conducted by converting 2-mode to 1-mode, a conversion that results in only individual-individual results. All 1-mode matrices in this research were created via sums of cross-products. The sums of cross products yielded the best concordance with the historical record compared to ten other possible measures (Bonacich '72 was second best).<sup>25</sup>

Network cohesion values for the SRL dataset were computed as shown in Table 3, where the results include a short average path length, higher than typical clustering coefficient (due to an above average number of clusters or hubs), and a characteristically small world index.<sup>26</sup>

The small world value of 1.38 suggests a scale-free network. In order to test this, plots were made of the relative frequency (proportion) of the degrees for the individuals, which can be seen in Fig. 5 (panels A, B), and they display the characteristic right-hand tail with the highest degree individuals. The  $\alpha$  exponent was 1.6 determined by best fit, which was in good agreement with the observed frequency with least-squares linear regression ( $p = 0.000$ ). The linearity started breaking down at degree 23–24. These panels demonstrate the power law of a scale-free network. A scale-free network should also show linearity in a log-log plot of high degree individuals in the right-hand tail (Fig. 5C).<sup>27</sup> The large right-hand tail contained eight high centrality individuals comprising Zhou Enlai, Liu Shaoqi, Nie Rongzhen, Deng Xiaoping, Zhu De, Cai Chang, Li Fuchun, and Zhang Wentian. Figure 5C also displays a secondary set of moderately unique high degree individuals, who are important hubs: Dong Biwu (董必武 1886–1975), Qu Qiubai, Ye Jianying, Zhang Guotao (張國燾 1897–1979), Li Weiha (李維漢 1896–1984), Zhao Shiyan, Li Lisan, Ulanfu (烏蘭夫 1906–1988), Jiang Jinguo, and Ren Bishi (任弼時 1904–1950).

---

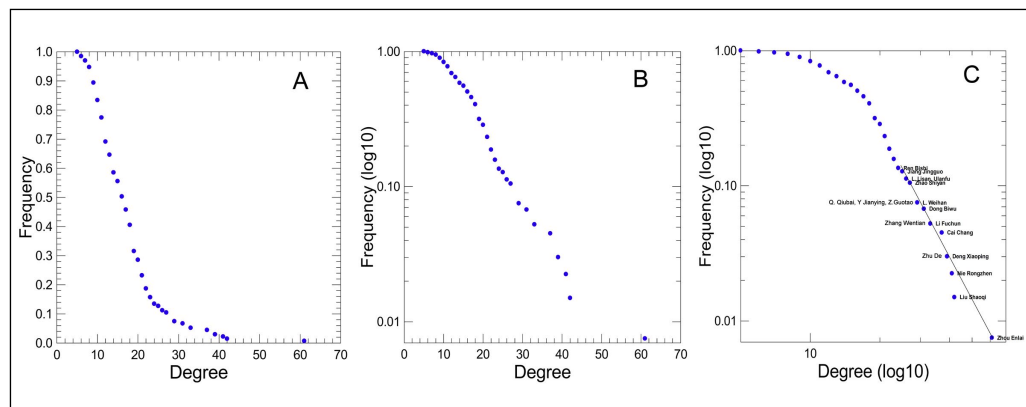
erences to Cai Chang and Li Fuchun are in the context of CCP historical events, or encyclopedia entries, and Chinese historical dictionaries, but not focused biographies. The evidence here indicates that they are important players in this SRL dataset.

25 P. Bonacich, "Factoring and Weighting Approaches to Status Scores and Clique Identification," *Journal of Mathematical Sociology* 92 (1972): 1170–82.

26 Duncan Watts and Steven Strogatz, "Collective Dynamics of Small-World Networks," *Nature* 393 (1998): 440–42.

27 A scale-free network should follow the general equation of  $p_k = Ck^{-\alpha}$  where  $p_k$  is the relative proportion of degree  $k$ ,  $C$  a normalizing constant, and  $\alpha$  is the networks characteristic exponent.

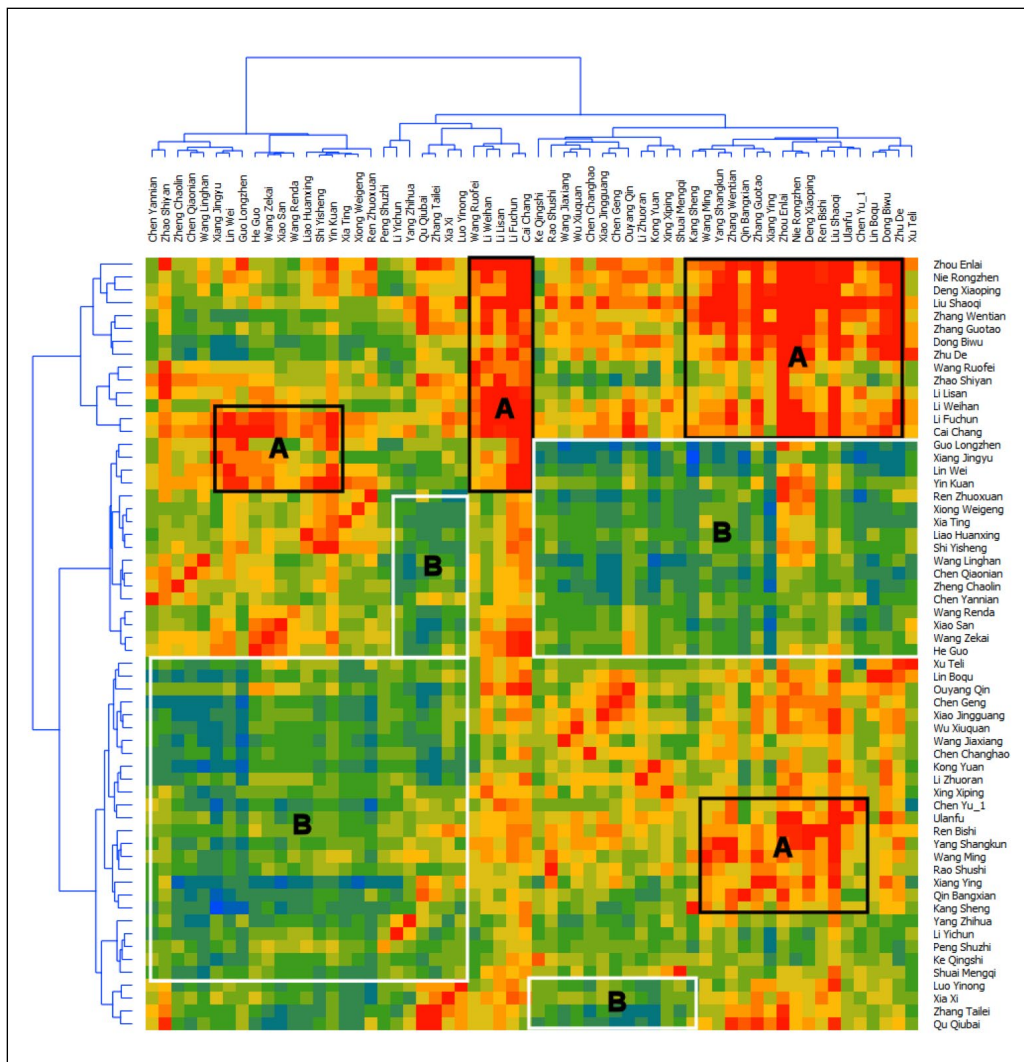
Average Path Length	1.003
Clustering Coefficient	0.9969
Small World Index	1.38
Density	0.9967

**Tab. 3** SRL network values.**Fig. 5** Degree and frequency proportionality. (A) Degree vs. frequency curve. (B) Degree vs. log of frequency. (C) Degree vs. frequency (log-log).

These high degree individuals provide network shortcuts that not only increase the total communication and transfer speed but also link strangers together by shorter links which is important for participants facing difficult situations in their revolutionary activities. The presence of high degree individuals is created by two major dynamic processes, namely preferential attachment and fitness. A preferential attachment process is based on nodes preferring to link with a node that has been shown to attract other nodes,<sup>28</sup> while fitness can be defined as a node possessing traits that attract more nodes at the expense of their being less fit.<sup>29</sup> Both preferential attachment and fitness can change, which may be seen in growing as well as static networks. Not all individuals are equally successful in acquiring links. In this article, these would be the high degree centrality individuals like Zhou Enlai, Nie Rongzhen, Liu Shaoqi, Deng Xiaoping, and Zhe Du (see Table 2), who have far more links than other individuals.

28 Albert-Laszlo Barabasi and Reka Albert, "Emergence of Scaling in Random Networks," *Science* 286 (1999): 509–12.

29 G. Caldarelli et al., "Scale-Free Networks from Varying Vertex Intrinsic Fitness," *Physical Review Letters* 89 (December 1, 2002): <https://doi.org/10.1103/PhysRevLett.89.258702>.



**Fig. 6** SRL core degree centrality linkage levels for those individuals. Dark red, orange and yellow colors equal the highest density of connections, while green and blue are at the lower end of density levels. Boxes highlighted by black and marked A are the high intensity hotspots and the boxes highlighted by white and marked B are medium intensity hotspots.

Scale-free networks are more resilient to failure – meaning that the network is more likely than a random network to stay connected after the removal of randomly chosen nodes. This means that the network quickly disintegrates when nodes are removed according to their degree. Albert et al. claim that the Internet and WWW are highly resilient to random removal of nodes, but are highly vulnerable to deliberate attack on the nodes of highest degree.<sup>30</sup> In addition, with a scale free network one finds greater perturbation resistance, and the Soviet returned leaders network certainly survived many significant perturbations.

Key player analysis<sup>31</sup> of the SRL data found that network disruption could be achieved by either removing the optimal communication diffusers Cai Chang, Cai Hesun, and Qin Bangxian which created maximal fragmentation (0.045) and heterogeneity (0.045) or by removing Cai Chang, Cao Chengde, Deng Xiaoping, Li Fuchun, Nie Rongzhen, Shi Yisheng, Su Zhaozheng (蘇兆征 1885–1929), Xie Zeyuan (謝澤沅 n.d.), Yang Lin (楊林 1898–1936), Yang Shangkun, Yin Kuan, and Zhou Enlai who were key in terms of their linkages to as many others as possible. Removing key nodes such as Cai Hesun was quite disruptive along with the numerous deaths in 1927–1931, which caused the SRL network to react by going underground and into the countryside to protect itself. In terms of historical change in China during the twentieth century, this scale-free network evolved and survived as nodes were killed or assassinated.

Core-periphery analysis examines the individuals who are highly connected as well as those who are less connected and peripheral. Core individuals are well connected to other core individuals and to peripheral individuals, while peripheral individuals are not strongly connected to each other.<sup>32</sup> The intensity of relationships between the 60 core individuals is displayed in Fig. 6, which shows several high and medium hot spots (A, B) as subgroups within the core that are highly connected. The high (A) subgroups are marked by black boxes and indicate the highest linkages, while the medium high (B) subgroups are marked by white boxes. These significant subgroups are identified by individual names and allow one to see which clusters interact more with others.

The centrality, scale, and core-periphery analyses suggest distinct subgroup linkages that contain leadership hubs. A look at Table 2, and Figs. 5 and 6 reveals several individuals known for political and/or military leadership. In the history

---

30 Reka Albert, Hawoong Jeong, and Albert-Laszlo Barabasi, "Error and Attack Tolerance of Complex Networks," *Nature* 406 (August 3, 2000): 378–82.

31 Stephen P. Borgatti, "Identifying Sets of Key Players in a Social Network," *Computational and Mathematical Organization Theory* 12, no. 1 (April 2006): 21–34, <https://doi.org/10.1007/s10588-006-7084-x>.

32 Stephen P. Borgatti and Martin G. Everett, "Models of Core/Periphery Structures," 21, no. 4 (2000): 375–95; Stephen P. Borgatti, Martin G. Everett, and Jeffrey C. Johnson, *Analyzing Social Networks*, 2nd ed. (Los Angeles: Sage Publications, 2018).

of the Chinese revolution the changes in the CCP power configurations were dramatic from the mid-1920s onwards and it is important to note that many of the SRL individuals identified in the above analyses survived as powerful leaders.<sup>33</sup>

### 5.3 Network Internal Structure: Triad Census Analysis, Subgroups, Ego Networks, and Pruning Tables

Triad analysis provides one method for studying the internal structure of a network. Triads of three individuals are the smallest possible network that contains more than two individuals. Triads are also where one sees the simplest occurrence of hierarchy. In 1908, Georg Simmel proposed that a triad reduces the individuality of its members to fully connected triad norms, as information is shared which creates trust with feedback, and conflicts between individuals can be mediated by the third individual, so that the triad behaves as a group rather than an individual. This also means that less connected triads may be less bound by group norms.<sup>34</sup>

A triad census enumerates all of the patterns of the actual links at the three-individual level. In the SRL group the network ties are undirected, which means there only are four possible triad types out of the 16 possible. The articulation of triadic analysis structure was developed by Holland and Leinhardt.<sup>35</sup> The triad types for undirected ties are: 003, which contains three isolated ties; 102, where two ties are not connected to the third; 201, where the third tie does not occur, creating a hole in the network that may or may not change over time; and 300, where three ties are fully connected creating a closed structure. The four types of undirected triadic relations are illustrated in Fig. 7, with the results presented in Table 4.<sup>36</sup>

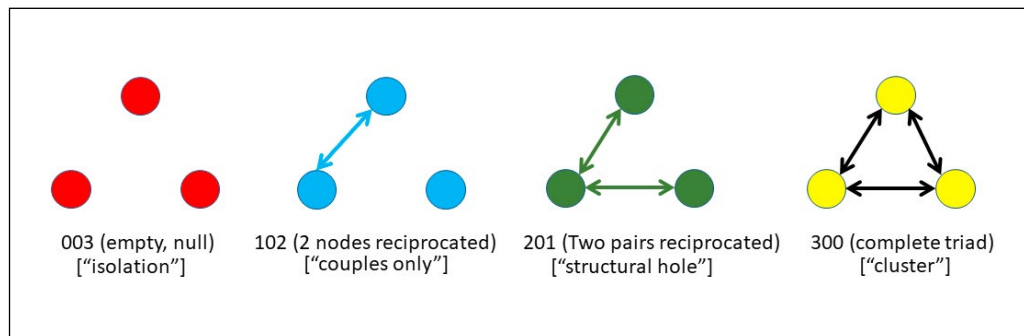
In Table 4, it can be observed that although the ratios of observed versus expected for types 102 and 003 are large, this was due to very few triads and low expected

33 Victor Shih, Shan Wei, and Mingxing Liu, "Gauging the Elite Political Equilibrium in the CCP: A Quantitative Approach Using Biographical Data," *The China Quarterly*, no. 201 (2010): 79–103, <http://www.jstor.org/stable/20749349>.

34 Kurt H. Wolff, *The Sociology of Georg Simmel* (The Free Press., 1950), <http://archive.org/details/sociologyofgeorg030082mbp>.

35 Paul W. Holland and Samuel Leinhardt, "A Method for Detecting Structure in Sociometric Data," *American Journal of Sociology* 76, no. 3 (1970): 492–513. Also see Tore Opsahl, "Triadic Closure in Two-Mode Networks: Redefining the Global and Local Clustering Coefficients," *Social Networks*, no. 35 (2013): 159–67.

36 Triads are tabulated by computer for all possible triplets in a network (383,176 for these data). The triads code types, as articulated by Holland and Leinhardt, consist of three parts: the number of mutual, asymmetric, and null pairs linkages that they contain. Thus, a 201 triad contains two mutual, no asymmetric, and one null pair.



**Fig. 7** The four possible triadic analysis structures for undirected ties.

Triad Type	Observed(ni)	Expected (ei)	Ratio (ni-ei)/ei
102	128	0.00	940.6 <sup>a</sup>
003	2	0.00	381.8 <sup>a</sup>
201	3,537	12.39	284.6
300	379.6 <sup>a</sup>	375.8 <sup>a</sup>	0.01

**Tab. 4** Triad analysis of SRL via Pajek. <sup>a</sup> = in thousands.

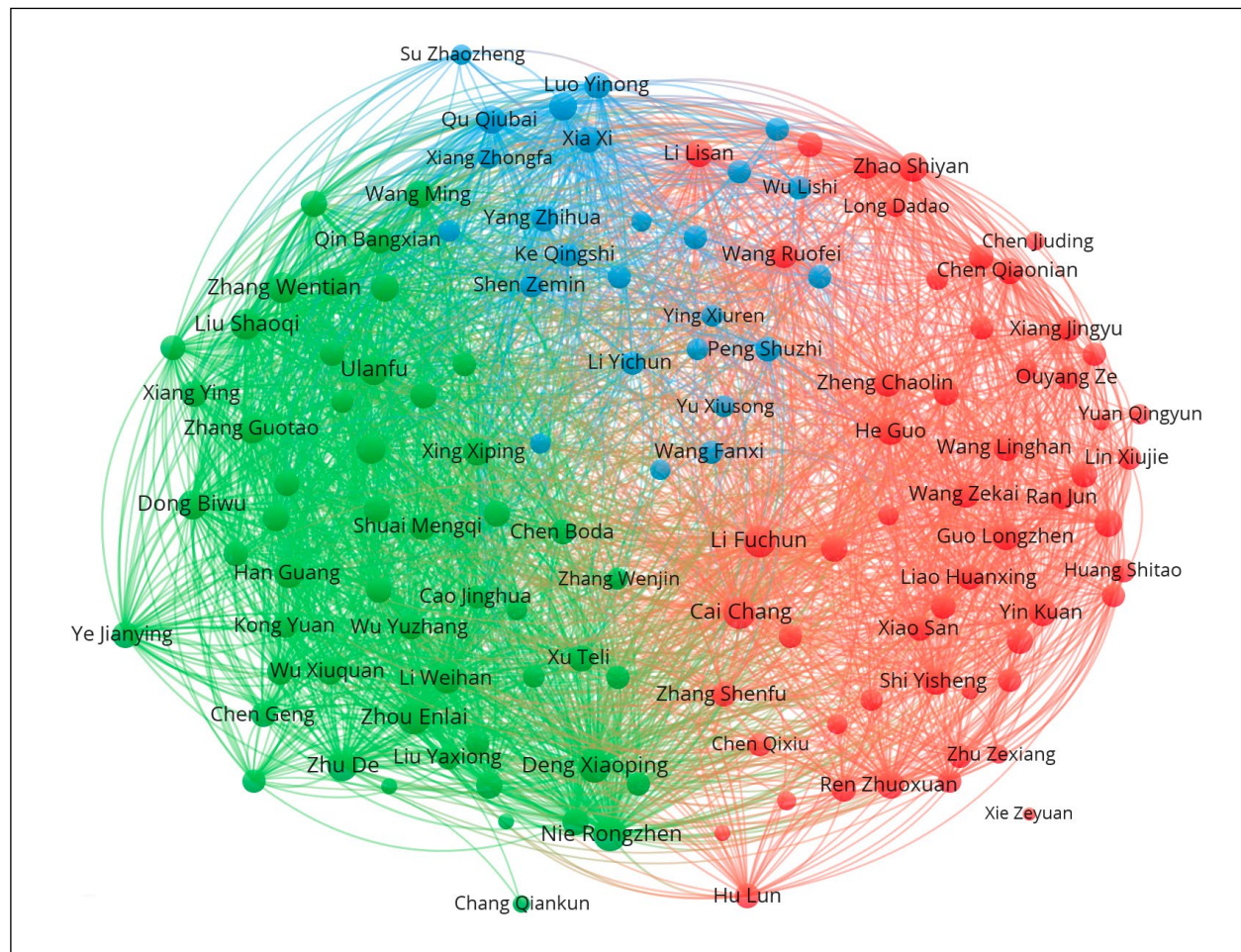
values.<sup>37</sup> The triad type 300, a fully closed cluster represents over 90 percent of the triads in this analysis. Type 300 results are not unexpected due to the strong connections between many SRL individuals. Type 201 (structural hole) is interesting, as it provides for a gatekeeper or broker (Fig. 7) who regulates information to others, or possibly the remainder of the network.<sup>38</sup> It is not clear what the unexpected occurrence of type 102 (couples only) means although the ni/ei ratios with zero expected values make interpretation difficult. The reported chi-square was highly significant ( $p = 0.001$ ) although the presence of zero expected values urges caution.

Another method for detecting internal structure is through visualization of similarities (VOS) cluster detection. This can be seen in Fig. 8, which displays a configuration of the SRL network with colored clusters. The aim of VOS is to pro-

<sup>37</sup> A triad census of a network simply reports the triad type code, the observed number of that particular triad code (ni), the expected number (ei) for that triad code, the ratio of observed to expected number, the model(s) for the network, and a chi-square. The expected number is calculated based on network size and number of linkages.

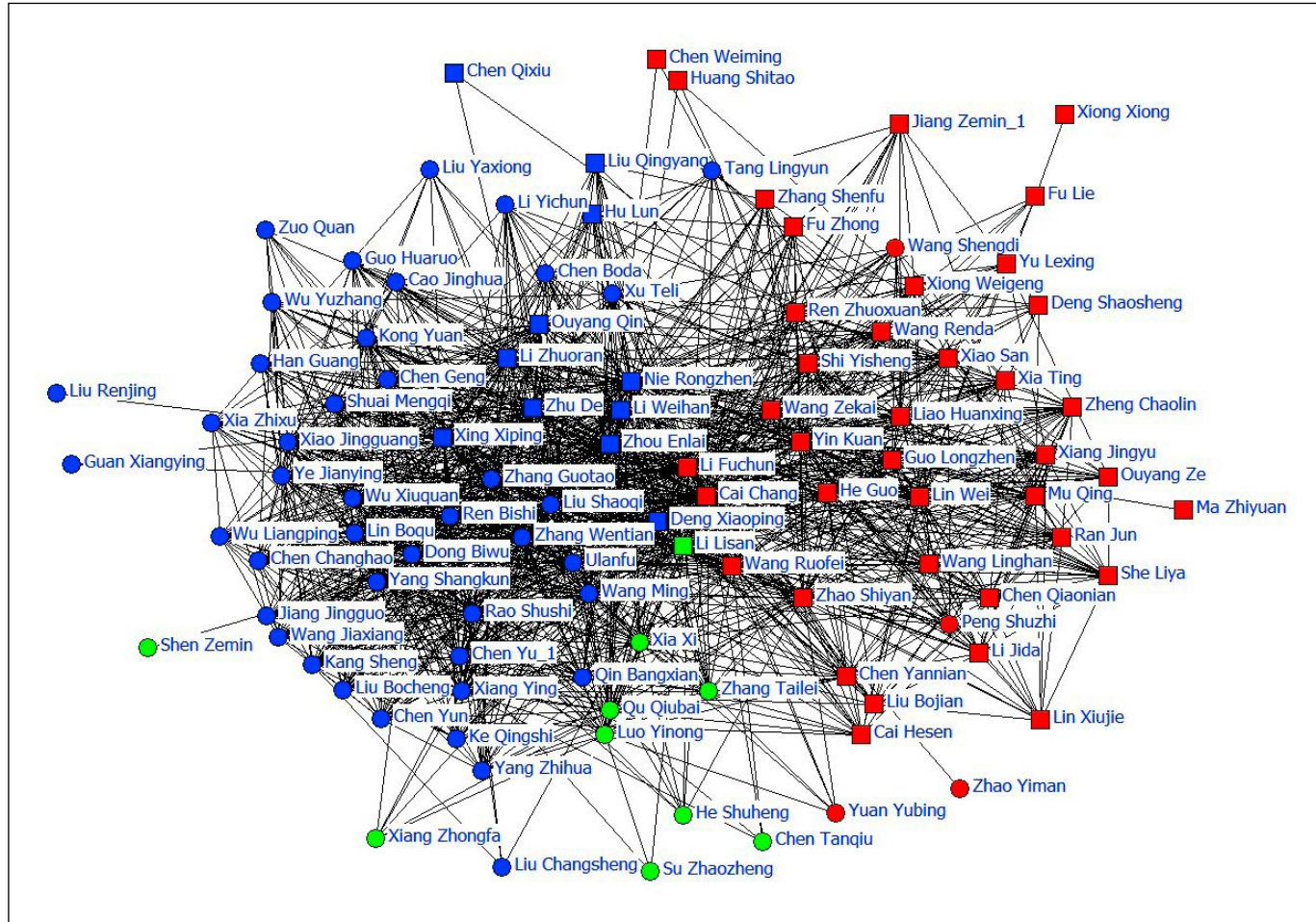
<sup>38</sup> Ronald S. Burt, *Structural Holes: The Social Structure of Competition* (Cambridge, Mass.: Harvard University Press, 1995).





**Fig. 8** SRL network graph with VOS internal clusters. Pruned to 2,500 ties for visibility, cluster resolution = 1.00, normalization via association strength.





eISSN: 2535-8863  
DOI: 10.25517/jhnr.v5i1.125

vide a low-dimensional visualization in which objects are located in such a way that the distance between any pair of objects reflects their similarity as accurately as possible. It is similar to multi-dimensional scaling but with less distortion due to indirect similarities of third objects and a tendency to locate objects close to their ideal coordinates.<sup>39</sup> The VOS graph in Fig. 8 displays three primary groups: a Soviet subgroup (green), a Euro-Soviet subgroup (red), and a smaller, intermediate group that with one exception is a Soviet subgroup (blue).

A second view of the SRL subgroups is presented in Fig. 9, which is obtained by a different analytical technique of Louvain clustering. One can observe results similar to Fig. 8 in the separation of the Louvain subgroups, i.e., that they are dominated by the Soviet (blue) and Euro-Soviet (red) individuals, with a small, intermediate group placed in between the two that is entirely composed of Soviet individuals (green). Zhou Enlai, Deng Xiaoping, Nie Rongzhen, and Zhu De are in the central cloud of the Soviet subgroup, while Cai Chang and Li Fuchun are in the central cloud of the Euro-Soviet subgroup. The square shapes represent affiliation with the ECCO and almost all ECCO members are in the Euro-Soviet red Louvain subgroup.

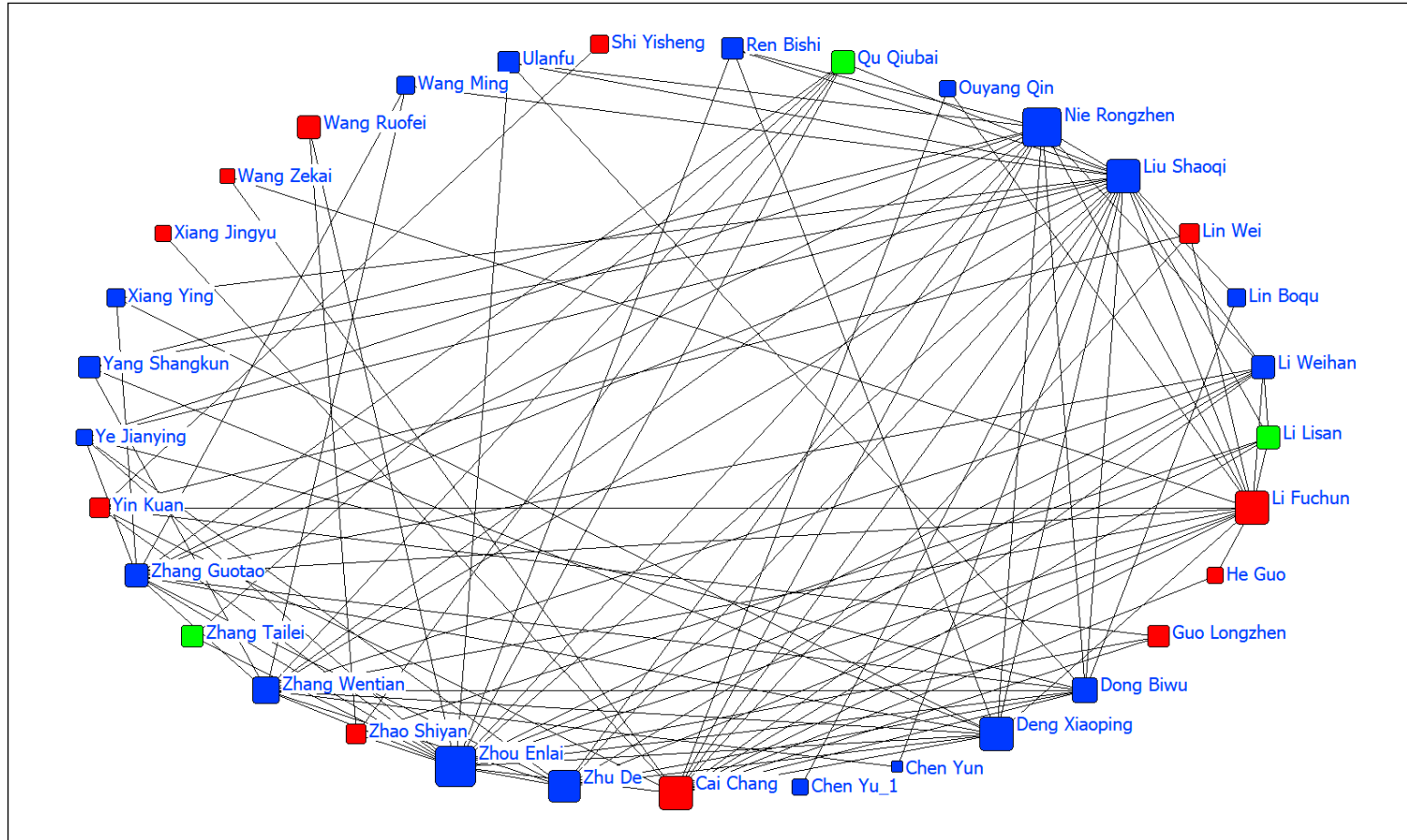
Figure 10 allows a further examination of the internal structure of the SRL network. It displays the 33 highest degree individuals from Table 2 in a circular layout. The colors again, represent the three Louvain subgroups, and the node sizes represent Eigenvector centralities, where one can observe more ties from those with higher Eigenvector values. In addition to well-known leaders, these 33 individuals include important early CCP figures who died in the twenties, such as Xiang Jingyu, Qu Qibuai, and Zhang Tailei, or leaders who were not widely known, such as the Trotskyites Wang Zekai, Yin Kuan, and Shi Yisheng. Yet their positions in the network demonstrate their ties to key leaders.

These subgroup graphs allow us to reexamine leadership perceptions. For example, Wang Ming is historically portrayed as the key leader of the so-called Twenty-eight Bolsheviks. Yet these measures demonstrate that Zhang Wentian has more linkages, which is made clearer when one views the ties in Zhang Wentian's and Wang Ming's ego networks displayed in Figs. 11 and 12. Zhang Wentian displays more linkages and a denser network than Wang Ming. Zhang is also much higher ranked in the centrality measures above, as well as the pruning table (see Table 5). This accords with the revised theories of the returned Russian leadership situation as presented by Thomas Kampen.<sup>40</sup>

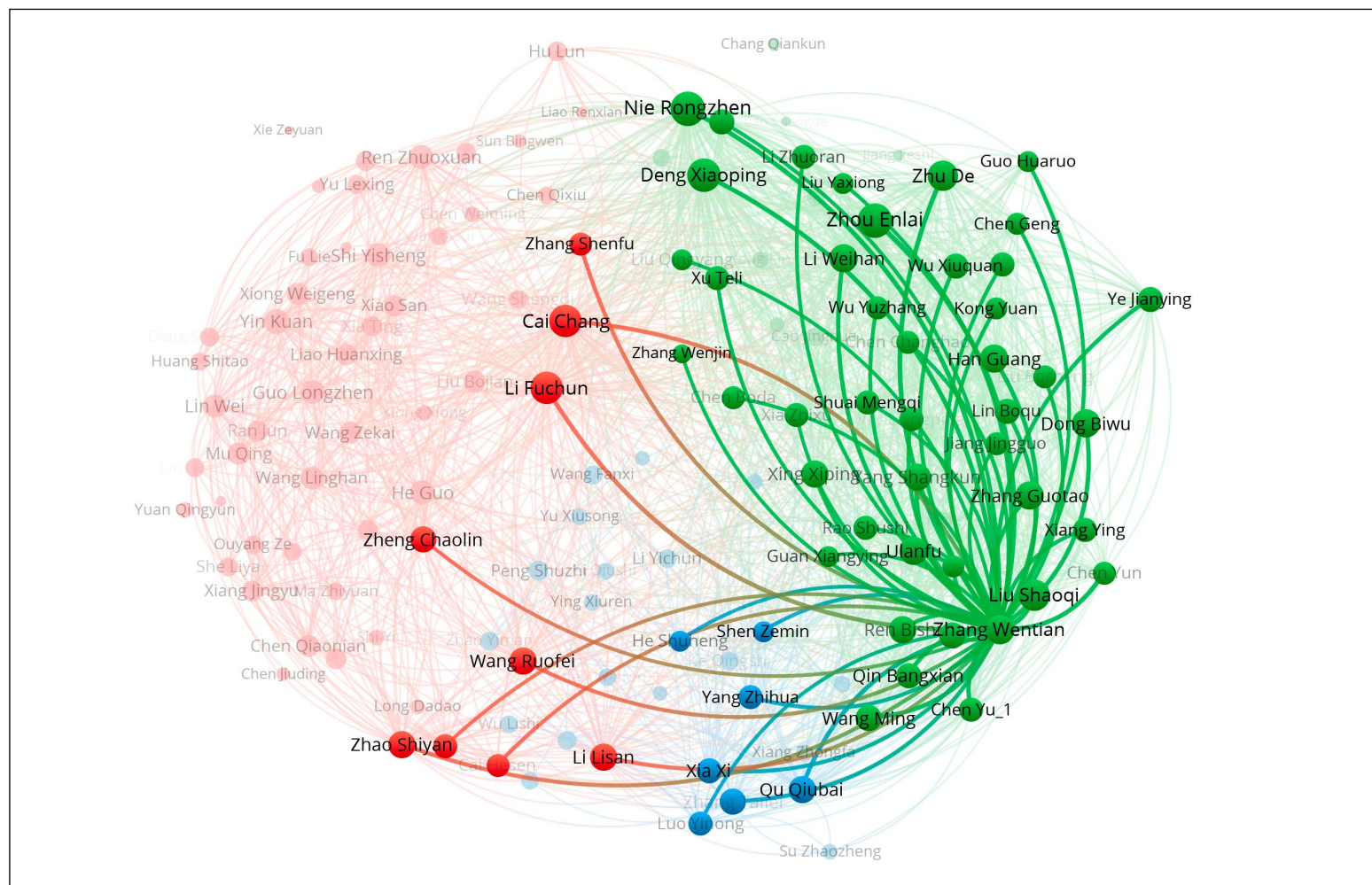
---

39 N. J. Van Eck and L. Waltman, "VOS: A New Method for Visualization Similarities between Objects," in *Advances in Data Analysis: Proceedings of the 30th Annual Conference of the German Classification Society*, ed. H. J. Lenz and R. Decler (Heidelberg: Springer, 2007), 299–306.

40 Kampen, *Mao Zedong, Zhou Enlai and the Evolution of the Chinese Communist Leadership*.

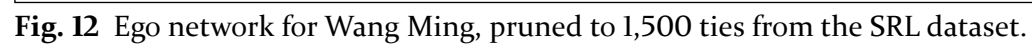


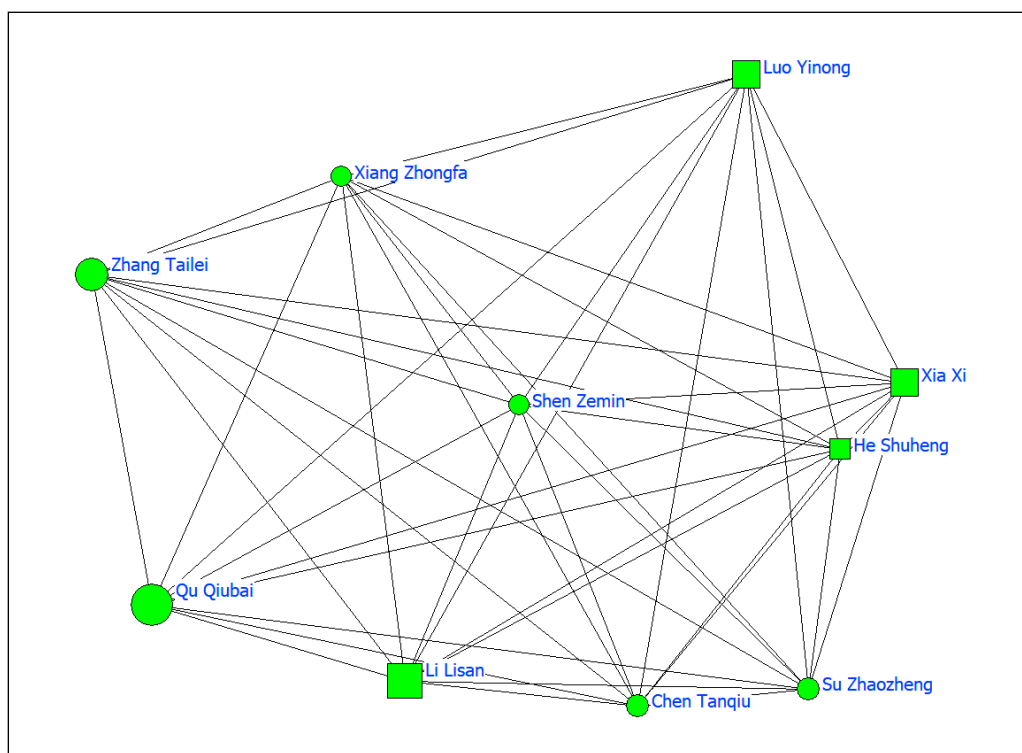
**Fig. 10** Network graph of SRL circular layout, individuals who have greater than 16 ties, equaling 33 individuals, 166 ties, Louvain subgroups by colors (green = 1, blue = 2, red = 3), Eigenvector centrality = symbol size. Note: Chen Yu\_1 is numbered as a unique biographical identifier for multiple same names.



**Fig. II** Ego network for Zhang Wentian, pruned to 1,500 ties from the SRL dataset.





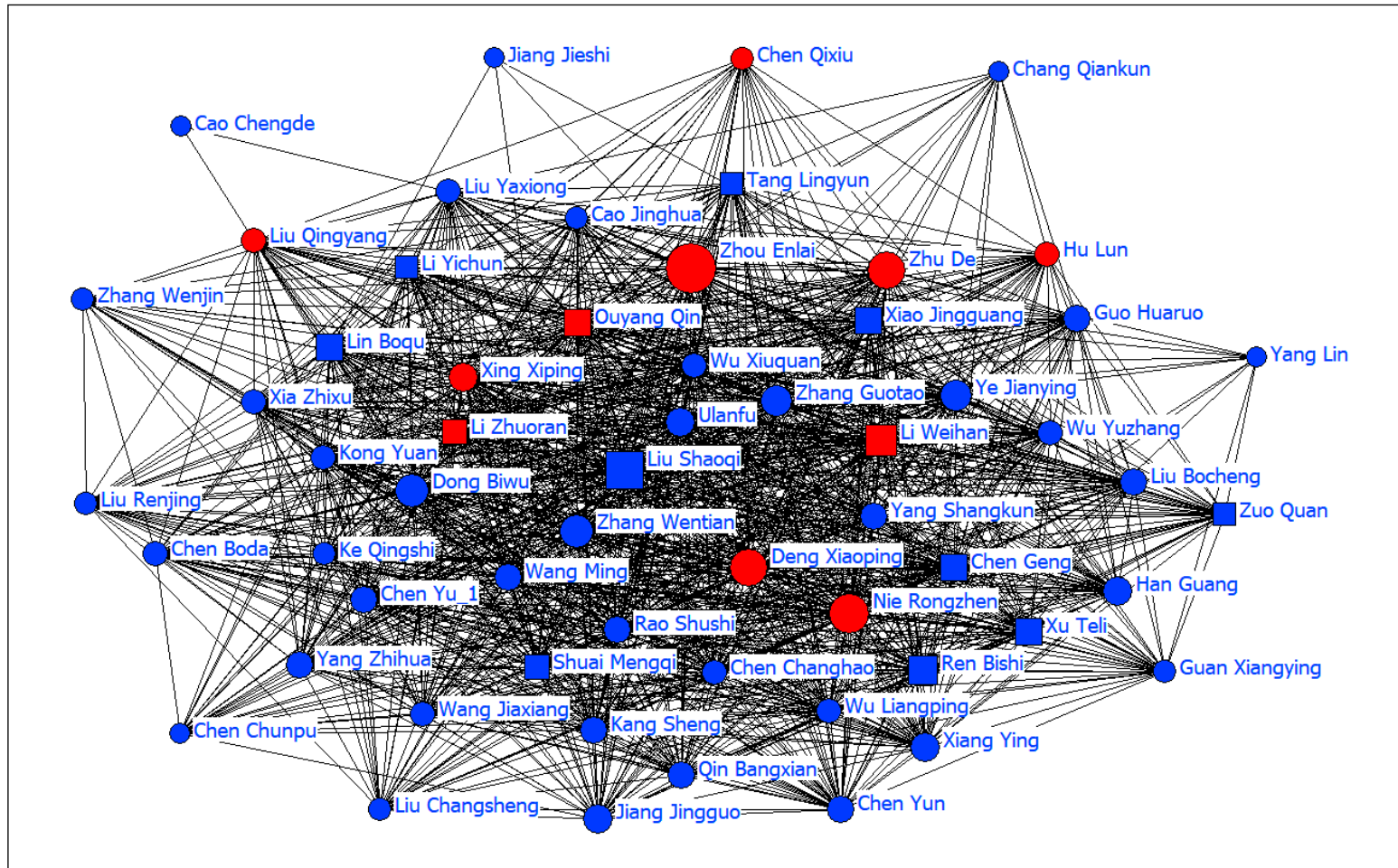


**Fig. 13** Network graph of Louvain subgroup 1, containing 10 individuals, 90 ties, degree centrality = symbol size, and Hunan provincial origin = square.

Figures 13–15 display each of the three SRL Louvain subgroups. Each graph displays attribute data that deepens our understanding of the individuals who are profiled by the graph. Besides displaying network graphs (total or subgroups) it is illuminating for historians to be able to add historically relevant information to a network graph. This approach expands the visualizations by understanding subgroup affiliations based on attributes. Currently this is done by node attributes (size, shape, color, rim), and similarly with linkages. Results with the SRL data show that by choosing attributes that are fairly common (political affiliations, provincial origin, common education institutions), the results yield strong correlations with each respective subgroup.

Fig. 13 displays the Louvain subgroup 1, containing 10 individuals who were primarily central leaders in the Chinese Communist Party and the Youth League, who either died by the early 1930s and 1940s or were out of power by that time.

Figures 14 and 15 with Louvain subgroups 2 and 3 contain the Soviet and Euro-Soviet subgroups respectively. Although there is some overlap, this segmentation between these two groups is a significant finding of this study. It is intriguing that four of the highest-ranking Euro-Soviet individuals are located in the Soviet



**Fig. 14** Network graph of Louvain subgroup 2, containing 59 individuals, 11,740 ties, degree centrality = symbol size, provincial origins = shape (circle = Sichuan, square = Hunan), and ECCO membership = red.

subgroup (Fig. 14), namely Zhou Enlai, Deng Xiaoping, Nie Rongzhen, and Zhu De. The ECCO affiliation is represented by the color red, and all four of the Euro-Soviet leaders have high degree centralities.

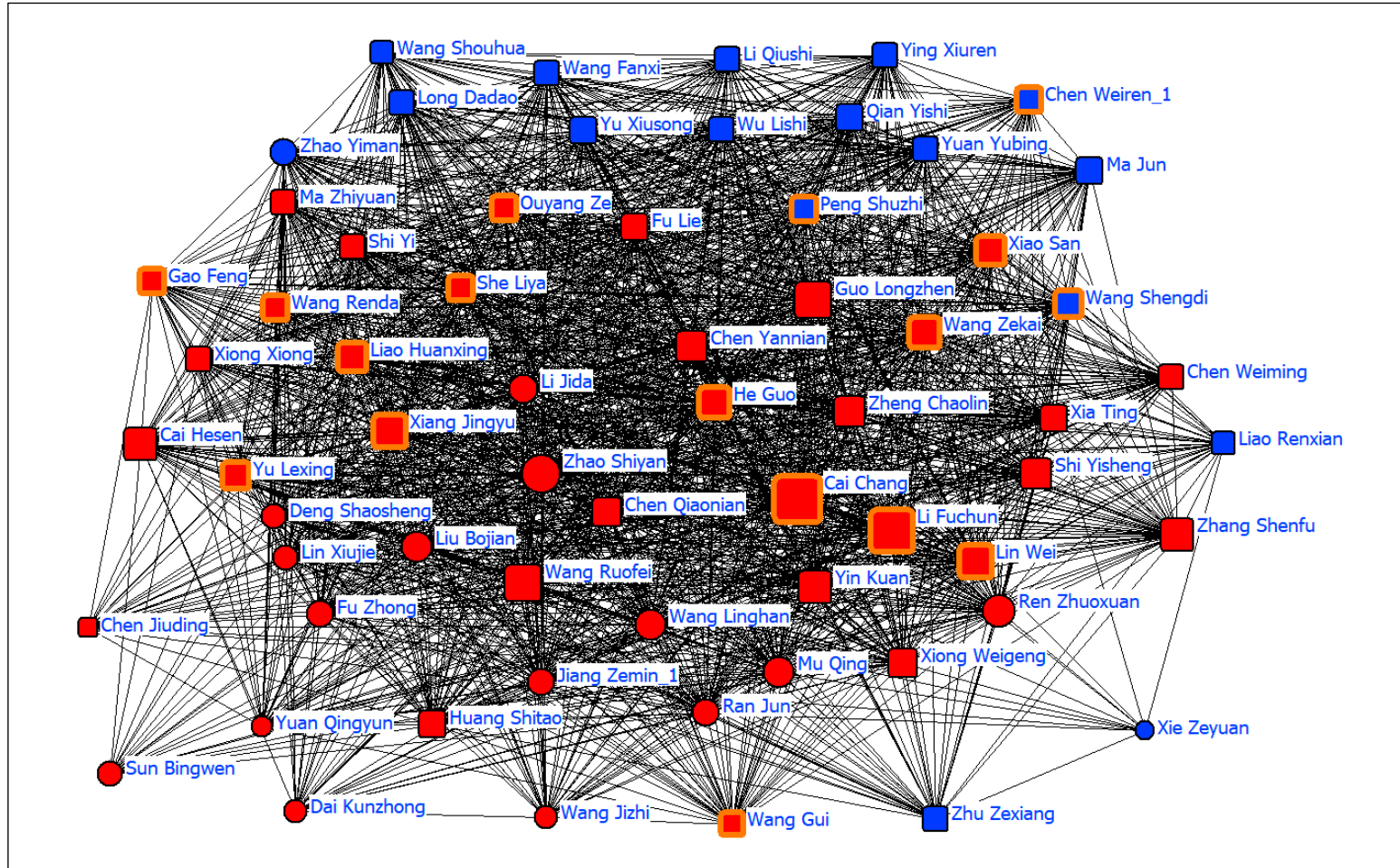
In Fig. 15, Li Fuchun and Cai Chang have the highest degree centrality in this largely Euro-Soviet subgroup. There is also a large contingent of Hunan individuals in Fig. 15, many of whom were radicalized together and may have adhered to lifelong bonds. It also is notable that three of the four higher-ranked ECCO members in Fig. 14 are from Sichuan. What is the role of provincial origin and political affiliation in keeping ties strong within these Louvain subgroups? Although the entire network has close linkages, these subgroup clusters suggest the importance of commonality of experiences and regional background. Finally, in Fig. 15 there are 6 of the 7 Trotskyites, even though several of them were not in the Euro-Soviet group. The Twenty-eight Bolsheviks are not so linked in the Louvain analysis, with two members in Subgroup 1, six in subgroup 2 (the Soviet group) and one in subgroup 3 (data not shown).

In addition to adding attributes to the visualization of subgroups, a second proposed method to better understand a network is the utilization of *pruning tables*, which slowly prune away the number of ties needed to remain in the network, one link at a time. Often displaying entire networks, unless quite small, results in difficult to interpret “hairballs.” It has been suggested that the slow dissection of a network by removing one required linkage at a time can be used to gain a better understanding of the structural relationships. These tables easily reveal at what stage individuals or attributes leave or remain in the network.

A pruning table approach can be used for either individuals or the original attributes. After converting the attribute data to 1-mode (attribute vs attribute), it can then be pruned stepwise by deleting the lower degree (linkage) attributes to reveal a graph of the highest degree attributes. In the SRL 1-mode network of individuals (Table 5) we see that the table begins with 17,498 ties in the total network and that no change occurs until those with 5 or fewer ties are removed. That person is Chen Jiuding (陳九鼎, n.d.) who has the least number of known attributes in the SRL dataset. Chen is one of twelve ECCO members who first went to Moscow and studied at KUTV. The table then skips to another level, which requires more than 7 ties before more individuals are eliminated from the network. It is intriguing to see how many people are not eliminated through later tie removal, in particular levels 18 to 30. At the 18-tie pruned level there are only 90 ties, and 17 remaining individuals, who take until level 30 to eliminate the final two individuals, namely Nie Rongzhen and Zhou Enlai. The pruning tables provide a valuable framework for easily clarifying the entire network structure.

As demonstrated in Fig. 16, one can see a progression of six visualizations at tie levels from 5 to 21. Panel A shows a dense center of nodes along with some peripheral individuals, where the lower degree individuals are gradually pruned



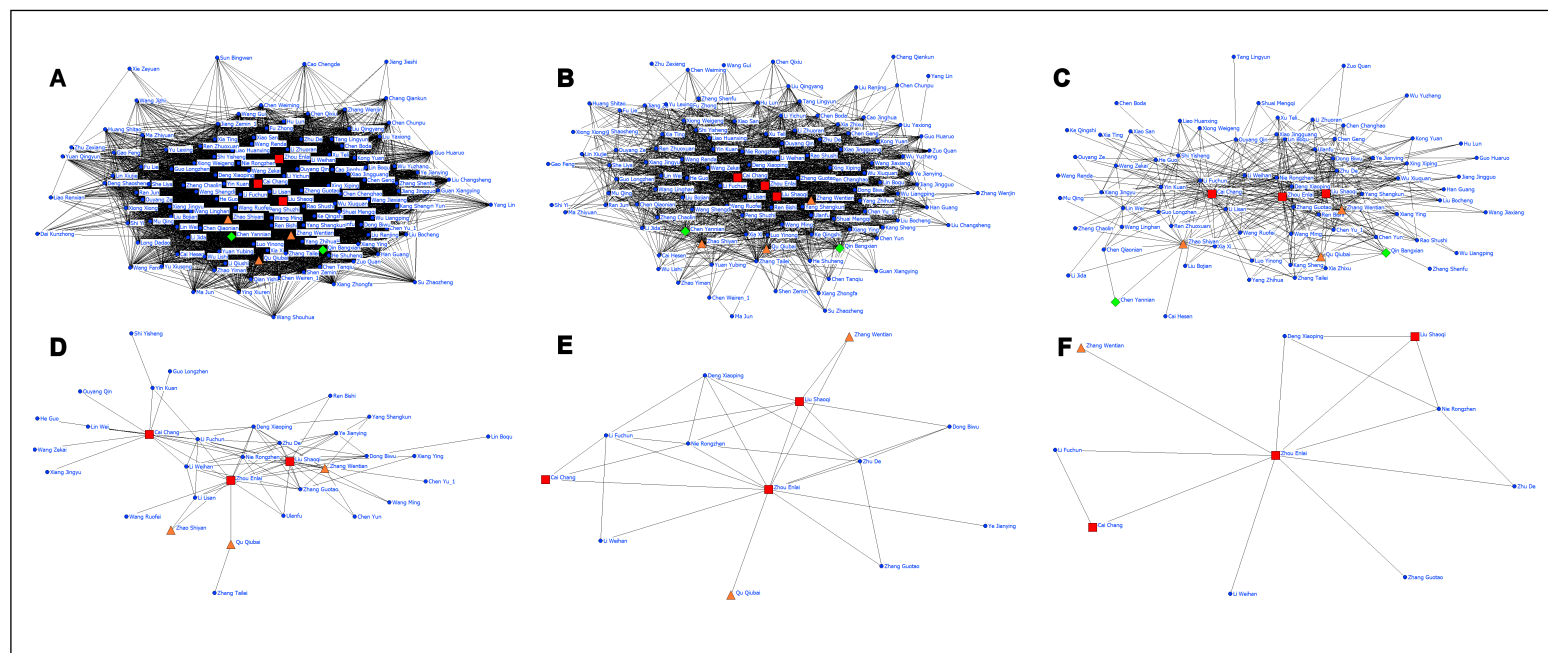


**Fig. 15** Network graph of Louvain subgroup 3 containing 64 individuals, 14,308 ties, degree centrality = symbol size, provincial origins = shape (circle = Sichuan, square with orange rim = Hunan), and ECCO membership = red.

Ties Pruned	Network Ties	Pruned Individuals	Remaining Individuals	Individuals Removed at Each Level
1	17,498	0	133	
2	17,134	0	133	
3	16,132	0	133	
4	14,308	0	133	
5	11,740	1	132	Chen Jiuding
6	8,916	0	132	
7	6,392	5	127	Jiang Jieshi, Liao Renxian, Wang Jizhi, Xie Zeyuan, Yuan Qingyun,
8	4,478	10	117	Cao Chengde, Dai Kunzhong, Li Qius-hi, Long Dadao, Qian Yishi, Sun Bingwen, Wang Fanxi, Wang Shouhua, Ying Xiuren, Yu Xiusong
9	3,032	11	106	Chang Qiankun, Chen Chunpu, Chen Weiren_1, Gao Feng, Ma Jun, Shi Yi, Wang Gui, Wu Lishi, Yang Lin, Zhang Wenjin, Zhu Zexiang
10	2,072	6	100	Chen Weiming, Guan Xiangying, Huang Shitao, Liu Renjing, Ma Zhiyuan, Zhao Yiman
11	1,400	12	88	Chen Qixiu, Chen Tanqiu, Deng Shaosheng, Fu Zhong, Lin Xiujie, Liu Changsheng, Peng Shuzhi, Shen Zemin, Su Zhaozheng, Xiang Zhongfa, Xiong Xiong, Yuan Yubing
12	962	11	77	Cao Jinghua, Fu Lie, He Shuheng, Jiang Zemin_1, Li Yichun, Liu Qingyang, Liu Yaxiong, Ran Jun, She Liya, Wang Shengdi, Yu Lexing
13	586	9	68	Cai Hesun, Chen Boda, Ke Qingshi, Mu Qing, Ouyang Ze, Tang Lingyun, Wang Jiaxiang, Wang Renda, Zuo Quan
14	358	15	53	Chen Changhao, Chen Qiaonian, Chen Yannian, Guo Huaruo, Jiang Jingguo, Li Jida, Liu Bojian, Wang Linghan, Wu Liangping, Wu Yuzhang, Xia Zhixu, Xiong Weigeng, Yang Zhihua, Zhang Shenfu, Zheng Chaolin
15	240	11	42	Chen Geng, Han Guang, Hu Lun, Kang Sheng, Li Zhuoran, Liao Huanxing, Shuai Mengqi, Xia Ting, Xia Xi, Xiao San, Xing Xiping

Ties Pruned	Network Ties	Pruned Individuals	Remaining Individuals	Individuals Removed at Each Level
16	166	9	33	Kong Yuan, Liu Bocheng, Luo Yinong, Qin Bangxian, Rao Shushi, Ren Zhuoxuan, Wu Xiuquan, Xiao Jingguang, Xu Teli
17	120	9	24	Chen Yun, Guo Longzhen, He Guo, Lin Boqu, Lin Wei, Ouyang Qin, Shi Yisheng, Wang Ruofei, Wang Zekai
18	90	7	17	Chen Yu_1, Wang Ming, Xiang Ying, Yang Shangkun, Yin Kuan, Zhang Tailei, Zhao Shiyan
19	56	4	13	Li Lisan, Ren Bishi, Ulanfu, Xiang Jingyu
20	38	1	12	Ye Jianying
21	28	2	10	Dong Biwu, Qu Qiubai
22	24	1	9	Zhang Guotao
23	16	2	7	Li Weiham, Zhang Wentian
24	16	0	7	
25	8	0	7	
26	8	1	6	Zhu De
27	4	3	3	Cai Chang, Deng Xiaoping, Li Fuchun
28	2	1	2	Liu Shaoqi
29	2	0	2	
30	0	2	0	Nie Rongzhen, Zhou Enlai

**Tab. 5** SRL Network pruning table with number of diminishing network ties at each level and the remaining individuals.



**Fig. 16** Prunings of network graph. A. 132 individuals with  $> 5$  ties. B. 117 individuals with  $> 8$  ties. C. 77 individuals with  $> 12$  ties. D. 33 individuals with  $> 16$  ties. E. 13 individuals with  $> 19$  ties. F. 10 individuals with  $> 21$  ties. Representative individuals: Chen Yannian and Qin Bangxian, green circles; Zhao Shiyan, Qu Qiubai and Zhang Wentian, orange triangles; Cai Chang, Liu Shaoqi and Zhou Enlai, red squares.

away. For example, Jiang Jieshi is on the periphery and is eliminated from the network at the next pruning level. Jiang in fact had a narrower experience in the Soviet Union and did not, at this stage, interact with most of the SRL. In panel B, one can trace highlighted leaders like Chen Yannian (陳延年 1898–1927) and Qu Qiubai. Chen is clearly connecting to ECCO cohort members, while Qu is in the middle of Soviet individuals and the so-called Twenty-eight Bolsheviks. In panel C, there still are strong linkages for Chen and Qu. Chen is linked to two individuals: Zhao Shiyan (who recruited Chen directly to the ECCO), and Chen's brother Qiaonian (陳喬年 1902–1928), while Qu has thinning but important linkages to leaders such as Zhang Tailei, Chen Yun (陳雲 1905–1995), Zhang Wentian, and Yang Shangkun. Panel D shows that Qu Qiubai and Zhao Shiyan still maintain a place in this network. In addition, there is an interesting placement of Cai Chang, who is centrally linked to a Hunan cluster of five individuals on the left; she also links to other network segments such as Trotskyites, women leaders, and other key leaders. Panel E shows the centrality of Zhou Enlai and Liu Shaoqi. Cai Chang and Qu Qiubai are on the peripheries. Lastly, panel F places Cai in a triangle with her husband, Li Fuchun and Zhou Enlai. Zhang Wentian is on the periphery, while Nie Rongzhen (who is eliminated alongside Zhou Enlai at > 30 ties) is actually more connected than Liu Shaoqi.

The pruning table concept can also be applied to the attributes of a 1-mode network constructed from attribute vs attribute which are the second type of nodes in the original 2-mode data. The results in Table 6 are an excerpt of the 17 highest linked individuals, excerpted from Table 5, along with their 127 attributes, with the rarer attributes not shown. As one might expect, party positions, military positions, and activities are highly ranked attributes for this group of top leaders. Yet some intriguing attributes include the May 4<sup>th</sup> Movement, and the birth city latitude, which when analyzed shows that 70 percent of these individuals lived on or nearby major waterways like the Yangtze or Xiang rivers. The linkage of the attributes is also interesting. In a comparison with a similar size group (N = 15) from the middle of the network (level 13) there was only a 29 percent elimination of attributes at level one, while 45 percent of the attributes were eliminated in the first level for the mid-level group (data not shown).

Isolate Level	No. Network Ties	No. Pruned Attributes	No. Remaining Attributes	Attributes Pruned at Each Level
0	8800	0	127	127 attributes
1–8	3968-176	104	23	Attributes not listed here
9	114	9	14	Civil War_1946_49, CPPCC, Cultural Revolution, Cultural Revolution Purged, ECCO, May 4th Movement, Military Official, NPC, Sino Soviet Relations
10	80	1	13	Yan'an Base Activities
11	46	1	12	Birth City Longitude
12	20	3	9	Birth Year, Death Year, Revolutionary
13	10	3	6	KUTV, Long March, Party Position
14	0	6	0	Central Committee, CCP, Birth City Latitude, Lifespan, Politburo, Red Base Activities

**Tab. 6** Excerpt from SRL Network Pruning Table for the top 17 individuals (127 attributes) with 1-mode network analysis. Results are displayed for number of ties and number of network ties, number of pruned and remaining attributes and names of pruned attributes at each level.

## 6. Conclusions

### 6.1 New Perspectives on CCP Leadership

This study is the first examination of early CCP structure using the SRL dataset from the larger CBD database. These preliminary results indicate potentially new perspectives for CCP leadership at the subgroup, individual, and attribute levels. The data presented demonstrate the distinct profile of birth year, death year, lifespan, and provincial and birth city origin that typified both the Euro-Soviet and the Soviet groups. The comparisons revealed an earlier death and shorter lifespan for the Euro-Soviet group than the Soviet group (see Table 1 and Figs. 1 and 2). The provincial and birth city origin show significant clustering in terms of close propinquity (see Figs. 3 and 4). Subgroup trends between the Euro-Soviet and Soviet groups were also displayed in the network profiles that formed distinctive patterns. It was also shown that this was a scale-free network with high cohesion and clustering, indicating an effective and highly connected network of individuals.

Network analyses generate new ideas for exploring past CCP leadership, as well as its evolution. For example, in the centrality measures, one can observe the

important individuals of influence (degree centrality), leaders who were close to others and able to obtain and convey information (closeness centrality), those who were close to centers of influence (Eigenvector centrality), or those who became power brokers (betweenness centrality). Because of the uncertainty of life and death during a revolution, the speedy flow of communication is critical for dynamic, dangerous situations. In terms of cohesion, the centrality measures, core-periphery, scale-free measures, and triadic census reveal a fast, efficient, and strong network that should be considered a major factor that ensured party survival. How did the close-knit nature of this network allow for adaptability and survival in the future challenges beyond the 1920s? How did the survivors learn to accept new responsibilities or switch roles? How inventive were they in difficult situations, and how were those responses tied to their adaptive experiences while overseas, encouraged by this densely cohesive network? As suggested by these analyses, these are possible research areas for the future.

The extraordinary role of Zhou Enlai demonstrated in numerous measures becomes more essential to understand. Zhou's role was displayed throughout the paper, particularly in the centralities where he was the highest degree, Eigenvector, closeness and betweenness. Likewise, the core individuals heatmap and visualizations demonstrated the strength of Zhou Enlai's position as a key leader. It has already been mentioned that he is clearly a boundary spanner between communities, occupying the highest degree position, his exceptional strength in the centrality measures, and the visualization of his network strengths in the subgroup graphs. But beyond being a conduit for communication, a person of influence and prestige, and a power broker, one must recognize that Zhou worked to maintain his influence but was comfortable with his role. This brings forth the question of leadership and ambition. Others in the network, such as Zhang Guotao, eventually gambled but lost in their attempt to gain more power in the 1930s, but they still played key roles and were important leadership hubs. The power struggles in the late 1920s and 1930s might have demonstrated to Zhou Enlai that he did not want to take the risk of being at the center with a higher position, since he could accomplish his objectives with the power he already wielded. Perhaps later CCP leaders, like Zhao Ziyang (趙紫陽 1919–2005) or Bo Xilai (薄熙來 1949–), may have been more vulnerable as they sought the higher positions and did not have the requisite breadth of network ties to see them through the political fluctuations. In addition, one could suggest that Zhou's experiences and openness to new knowledge, cultures, and ideas while in Western Europe and the Soviet Union definitely had an impact on his ability to adapt to the rapidly changing, complex set of circumstances that characterized the CCP revolution.

Secondly, Liu Shaoqi also deserves much greater study for his stature in the CCP network. In this study he emerged as high in the centrality measures and in the pruning table he was one of the critical final three members remaining. Although he does not have the boundary spanner role of Zhou Enlai, Liu Shaoqi is near the top in all four SRL centrality measures. Liu is important not just for

his positions, but also to study the role of Hunan. Liu Shaoqi later became the first successor planned for Mao Zedong, who also was from Hunan. What role was played by Liu's strengths in general and among his fellow Hunanese? A generally unassuming person, Liu was task oriented and gifted in his organizational skills. Liu Shaoqi provided much needed talent and strength during the 1930s and 1940s in the military leadership and ideological developments of the CCP.

Thirdly, the duo of Cai Chang and Li Fuchun are worth further research as a revolutionary couple. Their high placement in three of the four centrality measures, as well as their role in the Euro-Soviet subgroup demonstrates influence. But we can observe the nuances of their roles by viewing the centrality measures, where they are influential leaders with prestige, high proximity to powerful communities, and closeness to the other individuals in this network, although they are not ranked high in betweenness centrality scores that help to define those who are power brokers. How did this play out in the positions that were held by this couple in their revolutionary careers and post-1949? Did Cai and Li play a role as mediators and important individuals to consult on trouble shooting for their network colleagues? As Hunanese, how did they relate to and rely on Sichuanese long-term comrades such as Deng Xiaoping and Nie Rongzhen, who measured high on all four centralities?

Network analysis allows one to raise questions about leadership within these subgroups. As discussed above, it was demonstrated that Thomas Kampen's assertion was correct in saying that the 28 Bolsheviks were not a discrete group upon their return to China, because they did not in historical fact temporally appear or behave as a block. It was Zhou Enlai who was at the center of political struggles in the late 1920s and early 1930s and he acted in concert with members of the so-called 28 Bolsheviks. Both of these ideas have been supported here by both the quantitative and network analysis of the Soviet Returned Leaders. However, this does not mean that the so-called Twenty-eight Bolshevik leaders like Wang Ming, Zhang Wentian, Wang Jiaxiang, or Qin Bangxian did not have influence or were not core members of the Soviet Returned Leadership. Clearly, they had affinities of ideology and at times ambition, just as other members did in the Soviet Returned Leaders group. It is suggested that a fruitful network analysis should be conducted for the Anhui, Zhejiang, and Jiangsu members of the 28 Bolsheviks. Likewise, as mentioned above, the disproportionate number of executions between 1927 and 1930 from the Euro-Soviet group should be studied in more depth, in particular with a time-series network analysis.

## 6.2 New Perspectives on Historical Network Analytical Tools

The article has suggested that historians might consider additional ways to analyze as well as represent their results to allow for better network clarity and the utilization of prosopographical data. First, it is suggested that by analyzing subgroups and their attribute data one might gain unexpected insights to compare



and contrast important historical patterns. Another technique presented was to develop pruning tables that identify relationships at each tie level. This method was proposed for understanding both networks of individuals and networks of their attributes. Knowing at what level individuals or attributes are eliminated allows the reader to have a more granular knowledge of complex network structures. It was suggested that one might accompany the pruning table with visualizations (as displayed in Fig. 16) that show the remaining individuals and their ties, and how these shifts worked at different levels.

Post-WWI Chinese revolutionaries who adopted their communist ideology in Europe, whether in Western Europe or the Soviet Union, played an important part in the Chinese revolutionary process. Historical network analysis and quantitative history allows us to better define who they were as individuals and as groups. The SRL clearly had a cohesive and effective network of revolutionary leaders who provided dynamic leadership during and after the Chinese revolution of 1949. This study has demonstrated a deeper understanding of these leadership linkages and hopefully raised intriguing questions for future research.

## References

- Albert, Reka, Hawoong Jeong, and Albert-Laszlo Barabasi. "Error and Attack Tolerance of Complex Networks." *Nature* 406 (August 3, 2000): 378–82.
- Bailey, Paul. "The Chinese Work-Study Movement in France." *The China Quarterly*, no. 115 (1988): 441–61. <http://www.jstor.org/stable/654865>.
- Barabasi, Albert-Laszlo, and Reka Albert. "Emergence of Scaling in Random Networks." *Science* 286 (1999): 509–12.
- Belogurova, Anna. "Networks, Parties, and the 'Oppressed Nations': The Comintern and Chinese Communists Overseas, 1926–1935." *Cross-Currents: East Asian History and Culture Review* 6, no. 2 (2017): 558–82. <https://doi.org/10.1353/ach.2017.0019>.
- Benton, Gregor. *China's Urban Revolutionaries Explorations in the History of Chinese Trotskyism, 1921–1952*. Atlantic Highlands, N.J.: Humanities Press, 1996.
- *Prophets Unarmed: Chinese Trotskyists in Revolution, War, Jail, and the Return from Limbo*. Historical Materialism Book Series. Leiden, The Netherlands: Brill, 2015.
- Bonacich, P. "Factoring and Weighting Approaches to Status Scores and Clique Identification." *Journal of Mathematical Sociology* 92 (1972): 1170–82.
- Borgatti, Stephen P. "Identifying Sets of Key Players in a Social Network." *Computational and Mathematical Organization Theory* 12, no. 1 (April 2006): 21–34. <https://doi.org/10.1007/s10588-006-7084-x>.
- "Two-Mode Concepts in Social Network Analysis." *Encyclopedia of Complexity and System Science* 6 (2009): 8279–91.

- Borgatti, Stephen P, and Martin G Everett. "Models of Core/Periphery Structures." *Social Networks* 21, no. 4 (2000): 375–95.
- Borgatti, Stephen P., and Martin G. Everett. "Network Analysis of 2-Mode Data." *Social Networks* 19, no. 3 (August 1, 1997): 243–69. [https://doi.org/10.1016/S0378-8733\(96\)00301-2](https://doi.org/10.1016/S0378-8733(96)00301-2).
- Borgatti, Stephen P., Martin G. Everett, and Jeffrey C. Johnson. *Analyzing Social Networks*. 2nd ed. Los Angeles: Sage Publications, 2018.
- Burt, Ronald S. *Structural Holes: The Social Structure of Competition*. Cambridge, Mass.: Harvard University Press, 1995.
- Caldarelli, G., A. Capocci, P. de Los Rios, and M. A. Muñoz. "Scale-Free Networks from Varying Vertex Intrinsic Fitness." *Physical Review Letters* 89 (December 1, 2002): 258702. <https://doi.org/10.1103/PhysRevLett.89.258702>.
- Chen, San-ching 陳三井. *Qingong jianxue yundong* 勤工儉學運動 [*The Diligent-Work Frugal-Study Movement*]. Taipei: Zhengzhong shuju, 1981.
- Chow, Tse-tung. *The May 4th Movement: Intellectual Revolution in Modern China*. 5th Printing edition. Cambridge, Mass.: Harvard University Press, 1960.
- Davis, Allison, Burleigh B. Gardner, Mary R. Gardner, and W. Lloyd Warner. *Deep South: A Social Anthropological Study of Caste and Class*. Chicago: University of Chicago Press, 1941.
- Dirlik, Arif. *The Origins of Chinese Communism*. New York: Oxford University Press, 1989.
- Dittmer, Lowell. *Liu Shao-Ch'i and the Chinese Cultural Revolution: The Politics of Mass Criticism*. Berkeley: University of California Press, 1974.
- Fowler, Josephine. *Japanese and Chinese Immigrant Activists: Organizing in American and International Communist Movements, 1919–1933*. Rutgers University Press, 2007. <https://doi.org/10.2307/j.ctt5hj7ms>.
- Freeman, Linton. "Finding Social Groups: A Meta-Analysis of the Southern Women Data." In *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, edited by Ronald L. Breiger, Kathleen M. Carley, and Philippa Pattison, 1–39. Washington, D.C.: National Academies Press, 2003.
- Holland, Paul W., and Samuel Leinhardt. "A Method for Detecting Structure in Sociometric Data." *American Journal of Sociology* 76, no. 3 (1970): 492–513.
- Hsu, Kai-yu. *Chou En-Lai: China's Gray Eminence*. Garden City, N.Y.: Doubleday, 1968.
- Jiang, Zemin. Interview, Beijing. Interview by Marilyn Levine and Liu Guisheng, October 25, 1985.
- Kampen, Thomas. "Chinese Communists in Austria and Germany and Their Later Activities in China." *Asian and African Studies* XI, no. 1–2 (2007): 21–30.
- *Mao Zedong, Zhou Enlai and the Evolution of the Chinese Communist Leadership*. Denmark: NIAS Pub., 2000.

- Kemper, Michael. "Red Orientalism: Mikhail Pavlovich and Marxist Oriental Studies in Early Soviet Russia." *Die Welt Des Islams* 50 (2010): 435–76. <http://www.jstor.org/stable/41105362>.
- Kirasirova, Masha. "The 'East' as a Category of Bolshevik Ideology and Comintern Administration: The Arab Section of the Communist University of the Toilers of the East." *Kritika: Explorations in Russian and Eurasian History* 18, no. 1 (March 11, 2017): 7–34. <https://doi.org/10.1353/kri.2017.0001>.
- Lee, Chae-Jin. *Zhou Enlai: The Early Years*. Stanford, Calif: Stanford University Press, 1994.
- Lenin, Vladimir Il'ich. *The Lenin Anthology*. Translated by Robert C Tucker. New York: WW Norton & Company, 1975.
- Levine, Marilyn. "Conducting Research in the French Archives on Chinese Radicalism." *Republican China* 22, no. 2 (April 1997): 93–102.
- "Revolutionary Roads: An Integrative Analysis Utilizing a Chinese Biographical Database." Workshop presentation for the ERC ENP-China Project "Elites, Knowledge, and Power in Modern China." Aix-en-Provence, France, October 7, 2019. <https://enepchina.hypotheses.org/>.
- Levine, Marilyn A. *The Found Generation: Chinese Communists in Europe during the Twenties*. Seattle: University of Washington Press, 1993.
- Levine, Marilyn A., and San-ching Chen. "Communist-Leftist Control of the European Branch of the Guomindang, 1923–1927." *Modern China* 22, no. 1 (1996): 62–92. <http://www.jstor.org/stable/189290>.
- *The Guomindang in Europe: A Sourcebook of Documents*. Berkeley, Calif.: Institute of East Asian Studies, University of California, Berkeley, Center for Chinese Studies, 2000.
- Lin Yusheng. *The Crisis of Chinese Consciousness: Radical Antitraditionalism in the May Fourth Era*. Ann Arbor, Mich.: University of Michigan Press, 2000.
- Nie, Rongzhen 聶榮臻. *Nie Rongzhen Huiyilu 聶榮臻回憶錄 [Memoirs of Nie Rongzhen]*. 3 vols. Beijing: People's Liberation Press, 1983.
- Pantov, Alexander, and Daria A. Spichak. "New Light from the Russian Archives: Chinese Stalinists and Trotskyists at the International Lenin School in Moscow, 1926–1938." *Twentieth-Century China* 33, no. 2 (2008): 29–50. <https://doi.org/10.1353/tcc.0.0001>.
- Pantsov, Alexander. *Bolsheviks and the Chinese Revolution 1919–1927*. Honolulu: University of Hawai'i Press, 2000.
- Sheng, Yueh. *Sun Yat-Sen University in Moscow and the Chinese Revolution; a Personal Account*. International Studies, East Asian Series Research Publication No. 7. Lawrence, KS: Center for East Asian Studies, University of Kansas, 1971.
- Shih, Victor, Shan Wei, and Mingxing Liu. "Gauging the Elite Political Equilibrium in the CCP: A Quantitative Approach Using Biographical Data." *The China Quarterly*, no. 201 (2010): 79–103. <http://www.jstor.org/stable/20749349>.

- Shiu, Wentang. "Les Organisations Politiques Des Étudiants Chinois En France Dans l'entre-Deux Guerres." Université de Paris VII, 1990.
- Tsinghua University Faculty Research Unit on the History of the Communist Party. *Fu Fa Qingong jianxue yundong shiliao* 赴法勤工儉學運動史料 [Documents on the Travel to France Work-Study Movement]. Beijing: Beijing chubanshe, 1979.
- Tushman, Michael L., and Thomas J. Scanlan. "Boundary Spanning Individuals: Their Role in Information Transfer and Their Antecedents." *The Academy of Management Journal* 24, no. 2 (1981): 289–305. <https://doi.org/10.2307/255842>.
- Van de Ven, Hans J. *From Friend to Comrade: The Founding of the Chinese Communist Party, 1920–1927*. Berkeley Calif.: University of California Press, 1991.
- Van Eck, N. J., and L. Waltman. "VOS: A New Method for Visualization Similarities between Objects." In *Advances in Data Analysis: Proceedings of the 30th Annual Conference of the German Classification Society*, edited by H. J. Lenz and R. Decler, 299–306. Heidelberg: Springer, 2007.
- Wang Y. C. *Chinese Intellectuals and the West, 1872–1949*. The University of North Carolina Press, 1974.
- Watts, Duncan, and Steven Strogatz. "Collective Dynamics of Small-World Networks." *Nature* 393 (1998): 440–42.
- Wenqian, Gao, Peter Rand, and Lawrence R. Sullivan. *Zhou Enlai: The Last Perfect Revolutionary*. New York: Public Affairs, 2008.
- Wilson, Dick. *Zhou Enlai: A Biography*. New York, N.Y.: Viking, 1984.
- Wolff, Kurt H. *The Sociology of Georg Simmel*. The Free Press., 1950. <http://archive.org/details/sociologyofgeorg030082mbp>.
- Zarrow, Peter Gue. *Anarchism and Chinese Political Culture*. New York: Columbia University Press, 1990.
- Zhang, Yunhou 張允侯, Yin Xuyi 殷敘彝, and Li Junchen 李峻晨. *Liu Fa Qingong jianxue yundong* 留法勤工儉學運動 [The Travel to France Work-Study Movement]. 2 vols. Shanghai: Shanghai renmin chubanshe, 1980.
- Zhao, Shiyan 趙世炎. "Lieningzhuyi zhi lilun yu shiji" 列寧主義之理論與實際 [The Theory and Practice of Leninism]. In *Zhao Shiyan wenji* 趙世炎文集 [Zhao Shiyan Collected Works] 393–94. Chengdu: Sichuan renmin chubanshe, 1984.
- Zheng, Chaolin 鄭超麟. *An Oppositionist for Life: Memoirs of the Chinese Revolutionary Zheng Chaolin*. Translated by Gregor Benton. Atlantic Highlands, N.J.: Humanity Books, 1997.
- Interview, Shanghai. Interview by Marilyn Levine and Zhu Yuhe, October 29, 1985.
- Interview, Shanghai. Interview by Marilyn Levine and Zhu Yuhe, June 18, 1990.

## Appendix 1 – Abbreviations

Chinese Biographical Database	CBD
Soviet Returned Leaders Dataset	SRL
Toilers of the East University in Moscow	KUTV
Chinese Communist Party	CCP
Chinese Nationalist Party	GMD
Anarchist Party (Surplus Society)	GYS
European Branches of the Chinese Communist Organizations	ECCO
Chinese Social Democratic Party	SDP
European Branch of the Chinese Nationalist Party	EGMD
Chinese Youth Party	QND

## Appendix 2 – Glossary of SRL Individual Names

Cai Chang	蔡暢	Qian Yishi	錢亦石
Cai Hesun	蔡和森	Qin Bangxian	秦邦憲
Cao Chengde	曹承德	Qu Qiubai	瞿秋白
Cao Jinghua	曹靖華	Ran Jun	冉鈞
Chang Qiankun	常乾坤	Rao Shushi	饒漱石
Chen Boda	陳伯達	Ren Bishi	任弼時
Chen Changhao	陳昌浩	Ren Zhuoxuan	任卓宣
Chen Chunpu	陳春圃	She Liya	余立亞
Chen Geng	陳賡	Shen Zemin	沈澤民
Chen Jiuding	陳九鼎	Shi Yi	史逸
Chen Qiaonian	陳喬年	Shi Yisheng	施益生
Chen Qixiu	陳啟修	Shuai Mengqi	帥孟奇
Chen Tanqiu	陳潭秋	Su Zhaozheng	蘇兆征
Chen Weiming	陳微明	Sun Bingwen	孫炳文
Chen Weiren_1	陳為人	Tang Lingyun	唐靈運
Chen Yannian	陳延年	Ulanfu	烏蘭夫
Chen Yu_1	陳郁	Wang Fanxi	王凡西
Chen Yun	陳雲	Wang Gui	王圭
Dai Kunzhong	戴坤忠	Wang Jiaxiang	王稼祥
Deng Shaosheng	鄧紹聖	Wang Jizhi	王極知
Deng Xiaoping	鄧小平	Wang Linghan	王凌漢
Dong Biwu	董必武	Wang Ming	王明
Fu Lie	傅烈	Wang Renda	王人達
Fu Zhong	傅鍾	Wang Ruofei	王若飛
Gao Feng	高風	Wang Shengdi	汪盛荻
Guan Xiangying	關向應	Wang Shouhua	汪壽華
Guo Huaruo	郭化若	Wang Zekai	汪澤楷
Guo Longzhen	郭隆真	Wu Liangping	吳亮平
Han Guang	韓光	Wu Lishi	吳麗石

He Guo	賀果	Wu Xiuquan	伍修權
He Shuheng	何書衡	Wu Yuzhang	吳玉章
Hu Lun	胡倫	Xia Ting	夏霆
Huang Shitao	黃士韜	Xia Xi	夏曦
Jiang Jieshi	蔣介石	Xia Zhixu	夏之栩
Jiang Jingguo	蔣經國	Xiang Jingyu	向警予
Jiang Zemin_1	江澤民	Xiang Ying	項英
Kang Sheng	康生	Xiang Zhongfa	向忠發
Ke Qingshi	柯慶施	Xiao Jingguang	蕭勁光
Kong Yuan	孔原	Xiao San	蕭三
Li Fuchun	李富春	Xie Zeyuan	謝澤沅
Li Jida	李季達	Xing Xiping	邢西萍
Li Lisan	李立三	Xiong Weigeng	熊渭耕
Li Qiushi	李求實	Xiong Xiong	熊雄
Li Weihan	李維漢	Xu Teli	徐特立
Li Yichun	李一純	Yang Lin	楊林
Li Zhuoran	李卓然	Yang Shangkun	楊尚昆
Liao Huanxing	廖煥星	Yang Zhihua	楊之華
Liao Renxian	廖仁先	Ye Jianying	葉劍英
Lin Boqu	林伯渠	Yin Kuan	尹寬
Lin Wie	林蔚	Ying Xiuren	應修人
Lin Xiujie	林修杰	Yu Lexing	余樂醒
Liu Bocheng	劉伯承	Yu Xiusong	俞秀松
Liu Bojian	劉伯堅	Yuan Qingyun	袁慶雲
Liu Changsheng	劉長勝	Yuan Yubing	袁玉冰
Liu Qingyang	劉清揚	Zhang Guotao	張國燾
Liu Renjing	劉仁靜	Zhang Shenfu	張申府
Liu Shaoqi	劉少奇	Zhang Tailei	張太雷
Liu Yaxiong	劉亞雄	Zhang Wenjin	章文晉
Long Dadao	龍大道	Zhang Wentian	張聞天
Luo Yinong	羅亦農	Zhao Shiyan	趙世炎
Ma Jun	馬駿	Zhao Yiman	趙一曼
Ma Zhiyuan	馬志遠	Zheng Chaolin	鄭超麟
Mu Qing	穆清	Zhou Enlai	周恩來
Nie Rongzhen	聶榮臻	Zhu De	朱德
Ouyang Qin	歐陽欽	Zhu Zexiang	朱澤祥
Ouyang Ze	歐陽澤	Zuo Quan	左權
Peng Shuzhi	彭述之		



MARCUS BINGENHEIMER

# The Historical Social Network of Chinese Buddhism

*Journal of Historical Network Research* 5 (2021) 233–247

**Keywords** Chinese Buddhism, Biographies of Eminent Monks, Dharma Drum Person Authority Database

**Abstract** The Historical Social Network of Chinese Buddhism is a large dataset consisting of c. 17,500 actors and c. 25,000 links. The data was collected between 2007 and 2020 as part of various projects at Dharma Drum, a Buddhist organization in Taiwan. It is based on two main sources: marked-up biographical literature, and the Buddhist Person Authority Database. The main component of the network begins in the late 3rd century and ends with actors in the early 20th.

The network serves as a research tool for various levels of historical reflection. On the micro-level, researchers can explore the ego-networks of persons of interest as they were embedded within the larger Buddhist networks of their time. On the meso-level, researchers can focus on certain periods and discern relevant communities and the communication lines between them. On the macro-level, the network can reveal long-term historical structures and provide quantitative evidence to corroborate or refute previous assumptions about *longue durée* trends within Chinese Buddhism.

This article relates how and why the data was collected, addresses its problems and limitations, as well as its potential for future research. The Historical Network of Chinese Buddhism is extensible, e.g. by including Korean and Japanese actors, or by importing data from the China Biographical Database Project or Wikidata.

The dataset, as well as its sources, is published with documentation on GitHub under a Creative Commons Attribution-Share Alike 4.0 license and is ready for use in open-source social network analysis tools such as Gephi or Cytoscape.

## 1. Introduction\*

In historical research, the word “network” is used both figuratively and in a formal sense.<sup>1</sup> On the one hand, a figurative, informal approach to “networks”, more or less densely theorized, can be useful wherever the focus is on groups, rather than individuals, regions, or periods.<sup>2</sup> It has the advantage that the network metaphor can be easily expanded to include places and things (like in Actor-network Theory (ANT)), and that data does not need to be in a computable format. In fact, computation does not need to enter the research process at all. The formal approach, on the other hand, requires the systematic collection of computable network data, which can be analyzed with metrics derived from graph theory, and visualized with the help of layout algorithms.<sup>3</sup> It produces testable, quantitative evidence to support or refute historical arguments. Arguably, formal network analysis scales better to larger amounts of information over longer time periods and can help to discover patterns that are not revealed by a close reading of sources. As with the other papers in this special issue, the focus here is on formal social network analysis.

---

\* **Acknowledgements:** I am grateful to Laurent Van Cutsem and the editors of this special issue for making valuable suggestions on an earlier draft of this article.

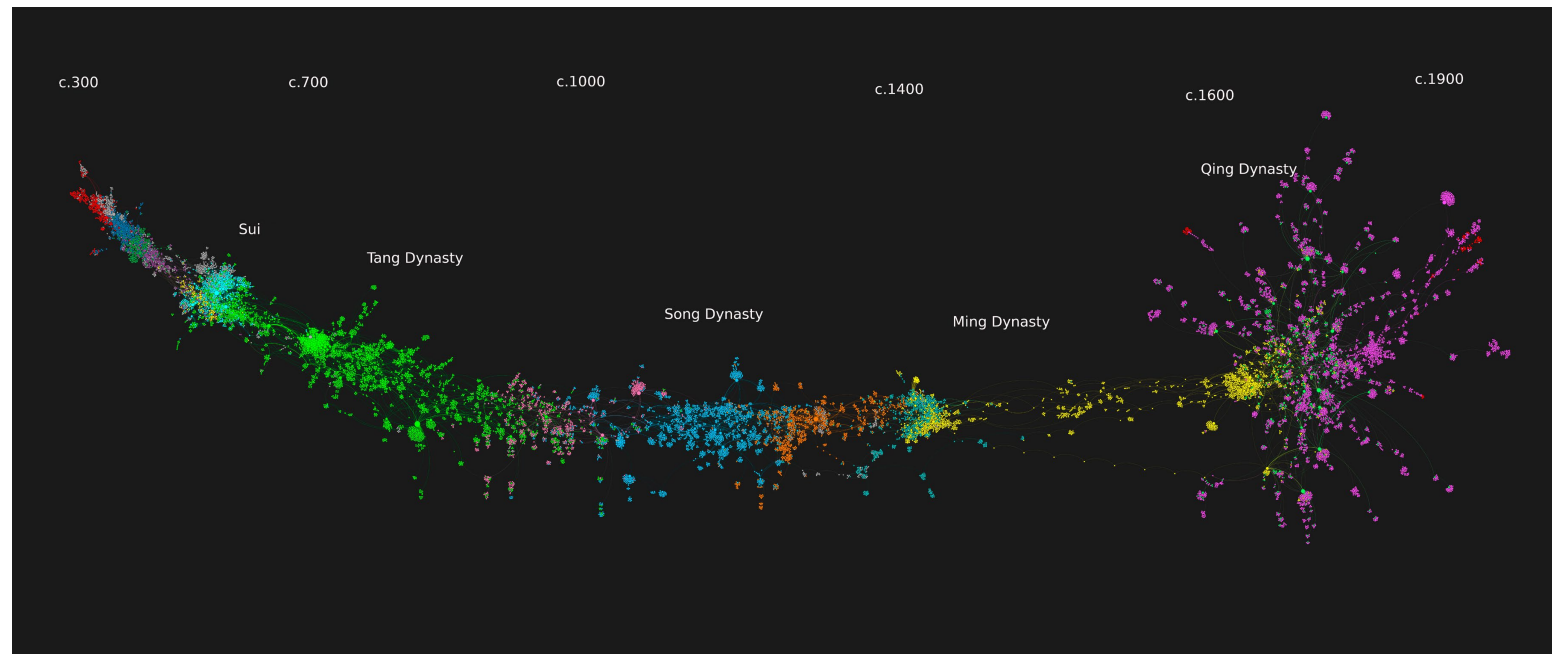
**Corresponding author:** Marcus Bingenheimer (Temple University), m.bingenheimer@gmail.com

1 For archaeology, the distinction I assert here was articulated by Knappett. Carl Knappett ed., *Network Analysis in Archaeology: New Approaches to Regional Interaction* (New York: Oxford University Press, 2013). Marx mentions the same distinction in his overview of historical network studies. Christian Marx, “Forschungsüberblick zur historischen Netzwerkforschung. Zwischen Analysekatgorie und Metapher,” in *Handbuch Historische Netzwerkforschung*, ed. Düring et al. (Münster: Lit Verlag, 2016), 83.

2 This is how the term is used in the history of religions e.g., by Jennifer Eichman for Buddhist networks in the late Ming, or by Anna Collar for cult networks in the Roman Empire. Jennifer Jennifer Eichman, *A Late Sixteenth-Century Chinese Buddhist Fellowship: Spiritual Ambitions, Intellectual Debates, and Epistolary Connection* (Leiden: Brill, 2016); Anna Collar, *Religious Networks in the Roman Empire – The Spread of New Ideas* (Cambridge: Cambridge University Press, 2013).

3 For examples see many of the contributions in Knappett, *Network Analysis*, Tom Brughmans, Anna Collar, and Fiona Coward eds., *The Connected Past – Challenges to Network Studies in Archaeology and History* (New York: Oxford University Press, 2016), and Marten Düring, Ulrich Eumann, Martin Stark, and Linda von Keyserlingk eds., *Handbuch Historische Netzwerkforschung* (Münster: Lit Verlag, 2016).





**Fig. 1** The main component of the network, laid out and rotated into an approximate timeline (ranging from c. 300–1950 CE).

## 2. Dataset – Overview and Metrics

The Historical Social Network of Chinese Buddhism is available under a CC Attribution-Share Alike license at [https://github.com/mbingenheimer/ChineseBuddhism\\_SNA](https://github.com/mbingenheimer/ChineseBuddhism_SNA). The data was collected between 2007 and 2020 as part of various projects at Dharma Drum, a Buddhist organization in Taiwan. It is based on two main sources: marked-up biographies of eminent Buddhists, and the Dharma Drum Buddhist Studies Person Authority Database (see Sec. 3 below). All links are referenced to a *source text* contained in the CBETA corpus.<sup>4</sup>

As of July 2020, the Historical Social Network of Chinese Buddhism consists of c.17,500 *actors* (nodes) and 25,300 *links* (edges). The latter are the (summed) result of 31,900 recorded connections between single actors (see below for a definition of “connection” in this context).

*Edge weights* are the sum of the connections between two nodes. They are of little importance in this dataset. Indeed, one single historical interaction can be recorded in various sources (or different passages in the same text) and therefore result in several connections. In other words, a link weight of 5 does not necessarily mean there are records of five *different* interactions between two actors. It might be based on one single event that was recorded in five different passages. Conversely, it might also be the result of two, three, four, or indeed five different interactions. Thus, edge weights have little analytic value in this dataset, beyond serving as a rough indicator for how prominently the interactions between two actors appear in the sources.

The *network density*, usually the fraction between the actual links and all possible links, is also of little use, even for comparative purposes, in diachronic networks that span several generations. By contrast to synchronous networks, in a diachronic historical social network actors can only form connections with contemporaries, which in our case excludes most other actors. Sub-graphs representing different time periods, however, might be gainfully compared with this metric, with their density indicating the level of activity on record for the period.

The *network diameter* of 55 is highly significant as it correlates exactly with the number of generations in the main component. The network diameter is the longest geodesic (i.e., the longest shortest path) through a network. The main component in the Historical Social Network of Chinese Buddhism begins with actors born around 250 CE and runs to actors born around 1950 CE, a period of 1700 years. The main component contains c. 15,500 of the total c. 17,500 actors (~87%), and c. 24,200 of the total 25,300 links (~96%). Dividing the period

---

4 The CBETA corpus of Buddhist texts is available at <https://github.com/cbeta-git/xml-p5a> (raw data), or at <https://cbetaonline.dila.edu.tw> (user interface).

range of the main component by the network diameter, one arrives at ca. 30 years (1750 years / 55), which is the mean for human biological generations.<sup>5</sup> That the network diameter reflects the number of generations expected for the time span it attempts to model is encouraging. This indicates that the main component contains no grave errors of the type that connects actors who could not have met, because their life spans did not overlap. A mean of 30 years also models teacher-student generations reasonably well, assuming that teachers were on average between 20 and 40 years older than their students (who were able to receive full-ordination at age twenty).

*Diachronic (Historical) spread.* The network models interactions of Chinese (and some Indian and Korean) actors associated with Buddhism from the 2<sup>nd</sup> to the 20<sup>th</sup> century, i.e., from the Han Dynasty to the People's Republic of China. However, neither the earliest nor the most recent centuries are well represented. No attempt has been made to collect data on post-1911 Buddhism, and most of the early Chinese Buddhists of the late Han and three Kingdoms period (c.150–250 CE) are not connected to the main component. Although the data distinguishes between more than 20 different dynastic periods, including transition periods (e.g. Ming-Qing 明清), almost 80% of the actors are assigned to the following eight periods:

Periods	Approximate percentage of actors in the network
Qing 清 (1644–1911)	26%
Tang 唐 (618–907)	20%
Northern Song 北宋 (960–1127)	10%
Ming 明 (1368–1644)	8%
Southern Song 南宋 (1127–1279)	5%
Five Dynasties 五代十國 (907–979)	4%
Yuan 元 (1279–1368)	3%
Sui 隋 (581–618)	3%

\* As defined by the Dharma Drum Person Authority (<http://authority.dila.edu.tw/person/>).

**Tab. 1** Percentage of actors by period.

5 Marc Tremblay and Hélène Vézina, “New Estimates of Intergenerational Time Intervals for the Calculation of Age and Origins of Mutations,” *American Journal Human Genetics* 66 (2000): 651–658.

Network *centralities* help to identify influential actors. Using an earlier, much smaller version of the network based only on biographies, I have reported how *degree centrality* in biographical literature highlights a different set of actors than *betweenness centrality*.<sup>6</sup> While the top twenty actors according to degree centrality consisted prominently of translators and patrons, betweenness centrality neatly foregrounded Chan and Vinaya masters. Today, an expanded version of the network contains much more information, especially for the second millennium, and here in particular for the Qing Dynasty. In the current network, the distinction between betweenness and degree centrality is less clear. Betweenness centrality is now somewhat distorted by the relative dearth of information about the mid-Ming. This has inflated the betweenness centrality of some otherwise relatively unknown figures in the 15<sup>th</sup> and early 16<sup>th</sup> century who now bridge more densely populated regions of the network before and after the mid-Ming.<sup>7</sup>

The *gender imbalance* in the data is conspicuous. Only about 2% (= c. 390) of the nodes represent women. This imbalance is a reminder that the network is based on sources that, at least with regard to gender, are hardly representative of historical reality (or only in as far as they themselves reflect the inequalities of the past, by what they choose to disregard). There is no reason to assume fewer women than men frequented Buddhist temples and supported Buddhist institutions. Indeed, if the present and the recent past is a guide, female lay-Buddhists seem to be doing more of the day-to-day work in Buddhist communities.

### 3. Dataset – Collection Principles and History

The Historical Social Network of Chinese Buddhism is extracted via a pipeline of scripts from two sources: marked-up biographies and an authority database data.

#### 3.1 Biographical data extracted from TEI/XML files

Starting in 2007, a team of encoders started to add TEI/XML markup to the biographical literature of Chinese Buddhism. The objective was to mark all person and place names and disambiguate them by mapping them to two evolving Name Authority Databases for person and place names. Also marked were dates, which

---

6 See Marcus Bingenheimer, “Who was ‘Central’ for Chinese Buddhist History? – A Social Network Approach,” *International Journal of Buddhist Thought and Culture* 28, no. 2 (Dec. 2018): 45–67. The Version 2018-01 contained c. 6,500 actors, while 2020-07 consists of more than 17,500.

7 Research into such neglected figures as Wuji Mingwu 無際明悟 (1381–1446), Bao Yuetan 寶月潭 (d.u.), Zhidu 智度 (1304–1370), and Cuifeng Deshan 翠峯德山 (1468–1549) is much needed.

were standardized via an East Asian Calendar Authority Database.<sup>8</sup> In addition, the encoders recorded *nexus points*, i.e. TEI markup constructs that encode where people were, when, and with whom. The instructions for the encoders were:<sup>9</sup>

Create a single-person nexus point where the text says:

- Something happened to one single person at a certain time
- Something happened to one single person in a certain place
- Something happened to one single person in a certain place at a certain time

Create a multi-person nexus point where the text says:

- Two (or more) figures interacted (e.g. A becomes student of B, A petitions B, A writes a letter to B, A and B are part of the same group that the emperor presents with purple robes)
- Two (or more) figures interacted in a certain place (or are simply mentioned as being in the same place somehow, which in biographical literature implies they were aware of each other)
- Two (or more) figures interacted at a certain time
- Two (or more) figures interacted in a certain place at a certain time

Table 2 shows which texts were marked-up and how many nexus points were created.

For social network analysis, obviously, only multi-person nexus points are of interest. Only contemporary figures can interact in this sense. Persons and places that are merely mentioned, but not encountered, were not encoded as nexus points (e.g. A remembering person B who visited the same place 200 years earlier, A longing for a place B where he had never been). Though not encoded as a nexus point, mentioned names are still encoded as <persName> or <placeName> in the main body of the text, and might be extracted as part of biographical research into a person.

A special case is the occurrence of non-historical figures in the texts. Some nexus points do include non-historical figures, such as the Bodhisattva Guanyin, who are reported as appearing in dreams or manifesting as human beings. En-

---

8 For details on the markup see Marcus Bingenheimer, Jen-Jou Hung, and Simon Wiles, "Social Network Visualization from TEI Data," *Literary and Linguistic Computing* 26, no. 3 (2011): 271–278. For the creation of the most comprehensive open database of East Asian Calendar dates see Marcus Bingenheimer, Jen-Jou Hung, Simon Wiles, and Boyong Zhang, "Modeling East Asian Calendars in an Open Source Authority Database," *International Journal of Humanities and Arts Computing* 10, no. 2 (2016): 127–144.

9 Expressed as a pseudo regular expression: (person,[time|place|time,place])|(person,person+,[time|place|time,place]?)

Text	Total number of nexus points	Multi-person nexus points
Collection of Records from the Tripiṭaka 出三藏記集 (卷13–15*) c. 515 CE	474	222
Excerpts from ‘Biographies of Famous Monks’ 名僧傳抄 (1卷) c. 514 CE	188	65
Biographies of Nuns 比丘尼傳 (4卷) c. 511 CE	376	194
(Liang Dynasty) Collection of Biographies of Eminent Monks (梁)高僧傳 (14卷) c. 530 CE	1793	958
Continued Collection of Biographies of Eminent Monks (from the Tang) (唐)續高僧傳 (30卷) c. 665 CE	5478	2179
Song Dynasty Collection of Biographies of Eminent Monks 宋高僧傳 (30卷) c. 988 CE	5204	2244
Ming Dynasty Collection of Biographies of Eminent Monks 明高僧傳 (8卷) c. 1617 CE	592	267
Supplement to the Continued Biographies of Eminent Monks 補續高僧傳 (26卷) c. 1641 CE	3974	1806
New Continued Biographies of Eminent Monks 新續高僧傳 (明清 biographies from 65 卷**) c. 1923 CE	2057	1985
Total	20136	9920***

\* Only the chapters containing biographies were marked-up.

\*\* Encoding of parts of this collection is still ongoing. The final tally of nexus points will be slightly higher.

\*\*\* The total does not translate directly into network links as nexus points are in effect hyperlinks, which can connect more than two nodes. Due to the limitations of the .gexf file format, these hyperlinks must be represented as individual links. i.e. a single nexus point involving four people must be represented as six edges. (This according to the formula for the number of edges on complete graphs:  $n(n-1)/2$ , with  $n$  being the number of nodes.)

**Tab. 2** Nexus points in biographical collections.

counters with non-historical beings, or otherwise legendary connections can distort the network considerably. Such events have been flagged in the XML source and are filtered out during the construction of the dataset.

No attempt is made to provide an ontology of events or interactions, i.e. the nexus points do not record what happened (much less why), merely that something took place (at a certain place, at a certain time). This leaves users at times dissatisfied. In defense, I can only say that creating an ontology of events and implementing it consistently was not a realistic goal during the creation of the dataset. Questions of bandwidth, time, funding, and personal tastes aside, there are basic limitations to modeling historical reality via ontologies. Since the data is openly available, interested researchers are welcome to categorize the c. 20,000 nexus points according to their needs.

### 3.2 The Dharma Drum Buddhist Studies Person Authority Database

In 2007, the Library and Information Center began to work on creating authority databases to disambiguate dates, person and place names that appeared in various digitization projects at the Dharma Drum library.<sup>10</sup> The Buddhist Studies Person Authority has grown to comprise information on c. 43,000 persons, mainly from China, but also including a sizable number of Japanese, Korean, and Indian figures.<sup>11</sup> It was created in tandem with markup projects that needed unique identifiers for people and places, but there were also discrete projects to input reference works such as Yuan Chen 陳垣 (1964), which resolves contradicting birth and death dates in the sources. The Person Authority is by far the largest such database in the field of Buddhist Studies, providing unique IDs, detailed, referenced information about dates, attributed works, and a precis with general information about the person. Especially important for SNA is the recording of lineage. Wherever possible, teachers and students of the person in question are listed. The main source for this information regarding the Ming and Qing is Yūkei Hasebe 長谷部幽蹊 (2008), the best researched, most comprehensive listing of lineage affiliation.<sup>12</sup>

Crucial for assembling the Historical Network of Chinese Buddhism, the data of the Person Authority is available for download under a CC license. This allows researchers to combine the lineage information and other data (e.g. life dates, gender, dynasty) from the Person Authority with the c. 10,000 multi-person nexus points from the c. 3000 TEI/XML files of biographical literature. The networks can be easily merged because they use the same person IDs.

- 
- 10 Online at <http://authority.dila.edu.tw/>. Of the three early databases, the East Asian Calendar Database is documented in Marcus Bingenheimer, Jen-Jou Hung, Simon Wiles, and Boyong Zhang, *Modeling*. A fourth, more recent, database allows convenient lookup of catalogs of different canonical editions of the Chinese Buddhist canon. It is a revised version of the Jinglu project (<http://jinglu.beta.org/>) organized by Aming Tu (1953–2016).
  - 11 For the current version of the Historical Network discussed here, Japanese and second millennium Korean figures were filtered out. The Person Authority in principle contains a large number of Japanese monks (10,610), connected by lineage data, as well as Joseon Dynasty Korean Buddhists (920), which could be included. The quality of information on non-Chinese actors is, however, not quite as reliable. There are numerous problems with transcription, romanization, use of variant characters in names, etc.
  - 12 The tables in this edition should be considered the “final version” of Hasebe’s work, who over the course of 40 years has aggregated information on Ming-Qing Buddhism from a vast range of literature from different genres (“Records of Collected Sayings” (*yulu* 語錄), “Records of Lamp Transmission” (*chuandenglu* 傳燈錄), local gazetteers (*difangzhi* 地方志) etc.).

## 4. Dataset – Problems and Limitations

Below I will try to assess the limitations of the *dataset*. This is not to be confused with the overall limitations of historical social network analysis as a *method*. Readers of this journal will have their own, informed opinion on the strengths and weaknesses of historical social network analysis.

### 4.1 Balance

The lineage data from the Person Authority and the biographical information from the markup complement each other with regard to their respective time periods. The five main biographical collections provide relatively dense information until the 10<sup>th</sup> century, while the lineage discourse provides data for the second millennium.

With the varying organization of the data, however, one would expect to already see a different network structure after the Northern Song (12<sup>th</sup> century). In fact, the relatively compact band of the diachronic network “explodes” into lineage strands only with the Ming-Qing transition in the 17<sup>th</sup> century. It remains to be seen exactly why this is the case. In any case, one should remember that information for the two millennia comes from quite different sources, and is structured differently. We have tried to mitigate this to a degree by encoding multi-person nexus points in the biographies of Ming and Qing monks contained in the *Xinxu gaoseng zhuan* 新續高僧傳 (CBETA/B. 151, completed 1923), thus adding some density to the simple tree structure of the lineage data. The addition of the *Xinxu gaoseng zhuan* data was partially successful in that it created greater cohesion for the Qing part of the network. However, even the humongous *Xinxu gaoseng zhuan* could not flesh out the thin network of the mid-Ming (c.1460–1560), the dark age of Chinese Buddhism, about which very little is known. Consequently, our network for this period remains meager.<sup>13</sup> With the exception of the mid-Ming, the main component of the network is now relatively well balanced between the two millennia.

### 4.2 Coverage and Completeness (Recall)

Related to *recall* in data mining terms, coverage is the relationship between the amount of SNA relevant information which has been included in the dataset, and the information in principle available to researchers. Total recall would mean

---

13 For a periodization of the *Xinxu gaoseng zhuan* biographies see Hasebe Yūkei who is among the very few researchers who made use of this work. Hasebe Yūkei 長谷部幽蹊, “Min-Shin jidai kōkai no tenbō – jisatuu no fukkō o megutte” 明清時代教界の展望 – 寺刹の復興をめぐって – [A view on Buddhism during Ming and Qing dynasties – With attention to temple reconstruction]. *Zen Kenkyūjo kiyō* 禪研究所紀要 6 (1976): 189–225.



that by consulting the Historical Network of Chinese Buddhism, one sees all knowable connections that are available to the historian. It is thus a measure of the (in)completeness of a search. For a historical network this is impossible to quantify exactly, but users of the data should be aware of the fact that the incompleteness increases over historic time. The closer we get to the present, the larger the difference is between what was modeled and what could have been modeled as a network from existing sources. This is mainly because there are many more sources for the more recent past.

In its current version, the network models a substantial amount of what is possible to know, in terms of SNA relevant information, about the earlier sections of the main component, from c. 250 CE to c. 500 CE. However, even for this period more links and a few more actors can be found, for instance, in prefaces and colophons to texts, many of which are collected in the catalog section of the *Chu sanzang jiji* 出三藏記集, in the *Essay on Buddhism and Daoism* that is part of the official History of the Wei Dynasty 魏書釋老志, in the *Shishuoxinyu* 世說新語, and in apologetic literature, e.g., as collected in the *Hongmingji* 弘明集. For these centuries, it is in principle possible to survey the existing sources, and perhaps even to arrive at a complete mapping of relationships for this period.

For later times, after say the Sui, there are simply too many sources that record connections between Chinese Buddhists. As with the knowledge of the individual researcher, the network can never hope to be complete. Though this seems obvious and historians always have to work from incomplete, biased sources, users often seem to expect total recall, perhaps conditioned by the use of dictionaries, where one can generally expect to find a word and most of its relevant meanings. Working with the Historical Network of Chinese Buddhism, it is important to be aware that incompleteness of the data increases over time, and there are more additional sources available for the second millennium than for the first.

### 4.3 Errors (Precision)

Errors in the data can occur on different levels. Firstly, errors introduced at the encoding level include: an encoder might misunderstand a line of classical Chinese and create a nexus point between two actors where there is none; misidentify a person and connect the wrong actors,<sup>14</sup> or they might miss a connection, even though a connection is implied in the text.<sup>15</sup>

---

14 Because of widespread homonymy and the use of abbreviations, this can easily occur. There are e.g. two Xuanzang 玄奘 (Pers. Authority ID: A000294 and A009306) in the Tang dynasty, and the two influential Huiyuan 慧遠 (A001204 and A002000) are only one hundred years apart.

15 This case would simply impact the recall rate (Sec. 3.2), but erroneously, not because data collection was not attempted.

Secondly, the source material might already contain errors. While the encoding accurately represents the meaning of the text, the information is still wrong. This is familiar to historians: just like ourselves, our predecessors sometimes got names and dates mixed up, or they misspell, misidentify, or misunderstand *their* sources. Source errors are problematic, because they are often difficult to detect, and are generally only noticed in the process of active research, not during markup.

Although our dataset probably contains errors of all the above types, there are reasons to be optimistic about the error rate. First, if two persons from different periods are connected erroneously, the network graph will be visibly “bent” or “deformed” by a single connection. In fact, during the course of the project, several encoding errors were found and corrected just by visual inspection of the network. Second, we have checked with the help of programs for “impossible connections” between persons whose life dates do not overlap. This caught both some encoding and some source mistakes.<sup>16</sup> Third, the correlation, mentioned above between network diameter and the expected number of generations for the time period, indicates that the network probably contains few, if any, mistakes that connect two persons of different generations. Fourth, because of the tight authority control, where one single ID is assigned per person, we are confident that each person is indeed represented as only one single node in this network.<sup>17</sup>

The main limitation of the current dataset is clearly its recall rather than its precision, i.e., scholars cannot expect the dataset to include all or even most information about links between actors. Especially for later periods, the network offers only a first overview over the social network of an actor, and more focused research will usually be able to uncover more connections.

## 5. Potential for future use

In spite of its limitations, the dataset is sufficiently large to amount to a completely new tool for the study of Buddhist history in China. It can be directly used in open-source SNA tools such as Gephi or Cytoscape, and the XML of the GEXF version of the data (available on GitHub) can be easily transformed into other graph-data formats.

The most obvious immediate contribution to the research process is that historians can now quickly look up the ego-network of a person of interest, and see

---

16 Nexus points that reflect source errors are flagged and excluded when the dataset is assembled.

17 Representing the same actor by different nodes in the network, e.g. because of an unrecognized alias, can lead to serious confusion at all levels of inquiry.

who else this person has interacted with. Based on the source references in the edge list, researchers can identify the relevant passages in the sources. In this sense the network functions like a dictionary look-up, querying connections instead of meanings.

The truly innovative power of HSNA, however, lies in its “zoomability.” Far from being restricted to the *micro-level* of ego-networks, researchers can zoom out to the *meso-level* of distinct time periods or cliques of actors, and even further out to contemplate the larger course of Chinese Buddhist history on a *macro-level*, e.g. by comparing network densities between dynasties.

It seems to me that this ability to zoom seamlessly between levels of historical inquiry while remaining within the same methodological framework is unprecedented for the study of Chinese Buddhism. There is no other dataset and methodology that allows the comparison of Buddhist networks in the Tang Dynasty with those of the Ming, and also, within the same tool-chain, the Buddhist network of the Sui emperor Wendi with that of the Liang emperor Wudi. Within the reductionist limits of HSNA as a method, the network perspective is universal in the sense that it can be applied to any period, indeed any tradition, as the papers in the special issue of this journal show.

There are several prospects for the future development of the dataset. One low-hanging fruit for the HSNA informed study of East Asian Buddhism would be to add Japanese and Joseon period (1392–1897) Korean actors. Many of these are already included in the Buddhist Studies Person Authority and can easily be merged in. Both the Korean and Japanese actors form distinct network regions in the second millennium, which are nonetheless connected to the main component. The problem is that the information for the Japanese and the later Korean actors in the Person Authority is not nearly as reliable as the data for Chinese Buddhists, and the network information for them is almost exclusively reliant on lineage charts, leading to relatively simple models. A “Historical Buddhist Network of East Asia”, though of course desirable, should ideally include Korean and Japanese sources beyond the lineage narrative.

Another (relatively) low-hanging fruit is to mine the corpus of Buddhist Temple Gazetteers for HSNA information by adding nexus point markup. This genre might “flesh out” the sparse network of the mid-Ming, as well as pulling the less well connected lineages of the Qing together, so that the Qing part of the network would visually come to more closely resemble the frayed band of the Tang and the Song periods than a blossoming flower.

Beyond simply expanding the Historical Network of Chinese Buddhism lies the challenge of how to merge other datasets into it. While researchers can easily merge smaller networks that result from their own research with the larger network in various ways, larger datasets need to be trimmed down before they can be

usefully included. The two main open source possibilities here are the China Biographical Database (CBDB) and Wikidata.

Connectivity with CBDB has been at least partially achieved by adding CBDB identifiers to the entries of the Buddhist Person Authority. As I have confirmed by testing, the networks can be connected via that crosswalk. In practice, however, only shorter time periods, or ego-networks, can be merged and meaningfully interpreted. Thus both datasets are better reduced first to a certain period or ego-network and then merged.<sup>18</sup>

Connectivity to Wikidata has also been achieved, as Wikidata has included the Dharma Drum Person Authority ID as a distinct property.<sup>19</sup> Querying Wikidata with SPARQL can yield interesting network data, which could then be merged with the Historical Network of Chinese Buddhism via the Person Authority ID.

## 6. References

- Bingenheimer, Marcus, Jen-Jou Hung, Simon Wiles, and Boyong Zhang. "Modeling East Asian Calendars in an Open Source Authority Database." *International Journal of Humanities and Arts Computing* 10, no. 2 (2016): 127–144.
- Bingenheimer, Marcus, Jen-Jou Hung, Simon Wiles, "Social Network Visualization from TEI Data." *Literary and Linguistic Computing* 26, no. 3 (2011): 271–278.
- Bingenheimer, Marcus, Jen-Jou Hung, Simon Wiles, and Boyong Zhang. "Modeling East Asian Calendars in an Open Source Authority Database." *International Journal of Humanities and Arts Computing* 10, no. 2 (2016): 127–144.
- Bingenheimer, Marcus. "Who was 'Central' for Chinese Buddhist History? – A Social Network Approach." *International Journal of Buddhist Thought and Culture* 28, no. 2 (Dec. 2018): 45–67.
- Brughmans, Tom, Anna Collar, and Fiona Coward eds. *The Connected Past – Challenges to Network Studies in Archaeology and History*. New York: Oxford University Press, 2016.
- Chen, Yuan 陳垣. *Shi shi yi nian lu* 釋氏疑年錄 [Record of doubtful dates of Chinese Buddhists]. Beijing: Zhonghua 中華書局, 1964.

---

18 Luling Huang and I have developed a script that queries an XML representation of CBDB and outputs SNA ready GEXF and GraphML files (<https://github.com/hlulingt/CBDBto-SNA>). Similar queries could be performed on the SQL data now distributed by the CBDB project.

19 Named "Dharma Drum Buddhist College Person ID" (<https://www.wikidata.org/wiki/Property:P1187>) after the previous name of the college (today called Dharma Drum Institute of Liberal Arts).

- Collar, Anna. *Religious Networks in the Roman Empire – The Spread of New Ideas*. Cambridge: Cambridge University Press, 2013.
- Düring, Marten, Ulrich Eumann, Martin Stark, and Linda von Keyserlingk eds. *Handbuch Historische Netzwerkforschung*. Münster: Lit Verlag, 2016.
- Eichman, Jennifer. *A Late Sixteenth-Century Chinese Buddhist Fellowship: Spiritual Ambitions, Intellectual Debates, and Epistolary Connection*. Leiden, Brill, 2016.
- Grandjean, Martin. “Introduction à la visualisation de données: l’analyse de réseau en histoire.” *Geschichte und Informatik*, no. 18/19 (2015): 109–128.
- Hasebe, Yūkei 長谷部幽蹊. “Min-Shin jidai kyōkai no tenbō – jisatsu no fukkō o megutte” 明清時代教界の展望 – 寺刹の復興をめぐって – [A view on Buddhism during Ming and Qing dynasties – With attention to temple reconstruction]. *Zen Kenkyūjo kiyō* 禪研究所紀要, no. 6 (1976): 189–225.
- Hasebe, Yūkei 長谷部幽蹊. *Min Shin bukkyō kenkyū shiryō (sōden no bu)* 明清佛教研究資料(僧傳之部) [Materials for the Study of Ming and Qing Buddhism (Monastic Biographies)]. Tokyo: Ōbaku bunka kenkyūsho 黄檗文化研究所, 2008.
- Knappett, Carl ed. *Network Analysis in Archaeology: New Approaches to Regional Interaction*. New York: Oxford University Press, 2013.
- Knappett, Carl. “Networks in Archaeology: Between Scientific Method and Humanistic Metaphor.” In *The Connected Past*, edited by Tom Brughmans et al. 21–33. New York: Oxford University Press, 2016.
- Marx, Christian. “Forschungsüberblick zur historischen Netzwerkforschung. Zwischen Analysekategorie und Metapher.” In *Handbuch Historische Netzwerkforschung*, edited by Marten Düring et al. 64–84. Münster: Lit Verlag, 2016.
- Tremblay, Marc, and Hélène Vézina. “New Estimates of Intergenerational Time Intervals for the Calculation of Age and Origins of Mutations.” *American Journal Human Genetics* 66 (2000): 651–658.

MICHAEL FULLER/HONGSU WANG

# Structuring, Recording, and Analyzing Historical Networks in the China Biographical Database

*Journal of Historical Network Research* 5 (2021) 248–270

**Keywords** China Biographical Database project, social network analysis, prosopography, relational database

**Abstract** The China Biographical Database (CBDB) is a relational database of over 470,000 individuals from pre-modern Chinese history. CBDB is distinctive as a prosopographical database in that it allows users to generate kinship and social networks for individuals – and groups of individuals – in the database. At the beginning of the project, to develop CBDB, we sought models among other digital prosopography projects but realized that, with CBDB's focus on analytic procedures and extracting data from the vast resources of the historical Chinese textual archive, we were developing a highly different model for digital prosopography. This paper presents an overview of the China Biographical Database, its capacities for exploring networks, and how we are extending those capacities through the ever-broadening extraction of data from the corpora of historical sources.

## 1. Introduction\*

Robert Hartwell's "Chinese Historical Studies" database system was the precursor to the current China Biographical Database (CBDB). Hartwell tracked both kinship and social relations among the Song dynasty officials he was studying, but he did not take the next step of exploring the networks defined by his data. When Michael Fuller rewrote Hartwell's database, he added recursive searches, starting with kinship data, to build social and kinship networks.

Once we realized the power of adding recursive searches to CBDB's structure, we understood the importance of collecting the data on kinship and social associations in carefully structured ways. Because the coders collecting the data from the texts wanted to stay as close as possible to the wording of the textual sources, their categories for coding associations expanded quickly, and we needed systematic ways to control the addition of new categories. Similarly, the variations in types of kinship relations that coders identified in the sources led to a large number of kinship categories, which complicated the analytic routines.

We solved the problem of the thicket of association codes through the aggregation of those codes through layers of higher-order classifications, so that scholars could decide how broadly or narrowly they wanted to construct the network of associations they sought to explore. We also added a halting-condition, restricting the node distance – the number of edges between the starting nodes and the final nodes – to be included in the network.

For kinship relations, we defined all relations in terms of generational (ancestor and descendent generations), lateral (sibling), and affinal metrics (in which a father's brother's wife shared the same metrics with a brother's wife's father). All combinations of kinship terms can be described within these metrics, and scholars can specify limits to their searches by setting upper bounds to these metrics.

Once scholars have results from the searches that build networks, they can export them to standard SNA software for further analysis. Because the CBDB is a relational database with significant additional information on the people in networks, scholars can look for additional patterns within their resultant networks. The first, immediate additional dimension is GIS data for the individuals, as the CBDB also outputs data on networks to standard GIS software.

---

\* **Acknowledgements:** We would like to thank Chen Song for the thorough discussion and editorial assistance.

**Corresponding author:** Hongsu Wang, hongsuwang@fas.harvard.edu

Because the CBDB is a relational database, scholars can constrain the groups of individuals whose networks they wish to explore: doctors in the Song dynasty, then Ming, then Qing, or degree-holders from Sichuan over time, etc. With these variations, the CBDB has developed into a powerful tool for using flexible criteria to generate networks from data on over 470,000 individuals in pre-modern Chinese history and to then explore them through many dimensions.

### 1.1 In the Beginning

Robert M. Hartwell (1932–1996) explained the goals of his ambitious Chinese Historical Studies (CHS) database project in a 1991 essay “A Computer-Based Comprehensive Analysis of Medieval Chinese Social and Economic History:”

Although research on Medieval China expanded at a rapid rate during the past two decades, these investigations suffer from the absence of a comprehensive understanding of the complex interrelationships among the social, economic, political and intellectual variables that, in the aggregate, constituted the structure or nature of Chinese Society. Such investigations as the many disparate and skillfully crafted monographs on separate localities, specific families or lineages, incumbents in one or another bureau of government, the functioning of one or another governmental institution, the growth of single industries and the nature of separate fiscal policies suggest many propositions about China from Han through early Ming. However, the collective information contained in these studies (Chinese, Japanese and Western) does not (and never will), provide the quantitative data necessary to test these propositions through investigating interregional and inter-institutional similarities and differences and changes in them over specified periods of time. The goal of the project described in this essay was to create the requisite database and use it to construct a framework for future research on Traditional China comparable to the ones available to students of European and American history.<sup>1</sup>

Upon Hartwell’s death in 1996, he left the *Chinese Historical Studies* database to the Harvard-Yenching Institute, and the database – revised in its structure and moved to a new software platform – became the *China Biographical Database*. Peter K. Bol, with the crucial support of Chen Song and his successor project managers, organized a collaboration with an expanding board of partners in China, Taiwan, and Japan to develop the CBDB.

Along the way, we broadened CBDB’s capacity to produce data on historical networks, a feature largely made possible by the accident of the language in which the database was written. Hartwell had created CHS in the dBase program-

---

1 Robert M. Hartwell, “A Computer-Based Comprehensive Analysis of Medieval Chinese Social and Economic History,” in *Characters and Computers*, ed. Victor H. Mair and Yongquan Liu (Amsterdam: IOS Press, 1991), 89.



ming language, and we ported the database to MS Access, where we used Visual Basic to create the database's analytic functions. We realized that we could use Visual Basic to loop through the records for kinship and associations to build networks. Once we understood the power of linking networks with the host of other historical "entities" in the CBDB (e.g., entry into office, office holding, and place of origin), we focused a great deal of effort both on extending the capacity to generate and contextualize social networks, and on expanding the data needed for it.

## 2. The China Biographical Database as a Model of Relations within Pre-Modern Chinese Society

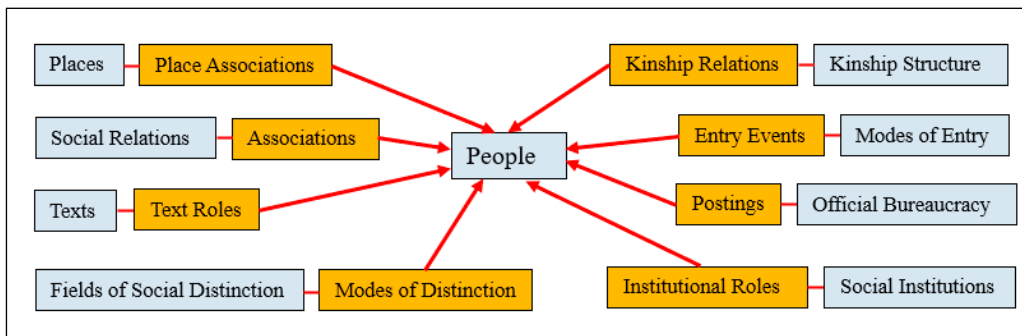
The design of the CBDB as a relational database derives from the ongoing effort, started by Hartwell, to model the core elements that shaped the lives of individuals in pre-modern Chinese society. Although the CBDB is based on the components of Hartwell's Chinese Historical Studies database, we expanded his initial design by applying an entity-relationship model to the historical sources. We identified nine basic components around which we built the CBDB. The current CBDB entity-relationship model thus includes these nine basic entities and the relations between them:

- 1) **People** (the central entity to which all other entities relate)
- 2) **Place** (not only people, but most other entities potentially have place attributes)
- 3) **Kinship Structure**
- 4) **Social Relations** in the society
- 5) The **Official Bureaucracy**
- 6) The **Modes of Entering** the official bureaucracy
- 7) The **Social Institutions** in which people participate (academies, temples, hospitals, etc.)
- 8) The **Fields of Social Distinction** through which individuals acquire status
- 9) **Texts** (texts often participate in the delineation of social networks, e.g., the exchange of writings and writings for specific occasions are expressions of social relationships)

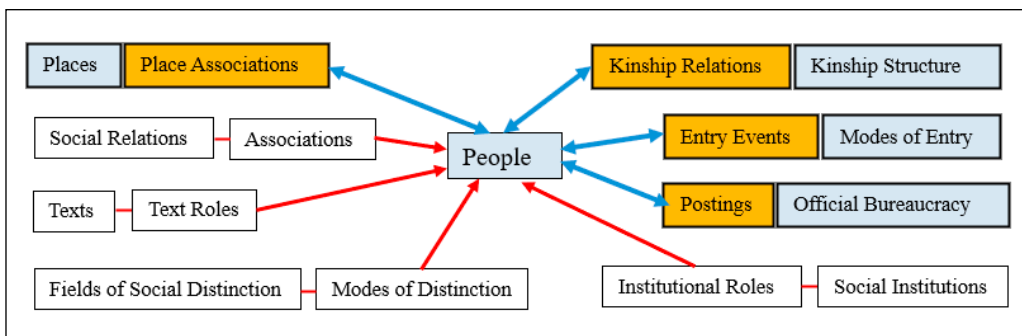
The relationships between people and the other entities are one-to-many.

The CBDB database converts this model into digital form by transforming each of the nine entities into code tables and transforming the relationships between entities into data tables linking the entities.

The power of a relational database is that one can easily explore the interaction of different factors. Consider, for example, the question of whether the families of officials who entered office through the "Presented Scholar" (*jinshi* 進士) examination made local marriage alliances and, more specifically, whether the pat-



**Fig. 1** The Basic Entities and Relationships in the CBDB.



**Fig. 2** The Interaction of Entities in Examinations, Office-holding and Social Networks.

tern differed from region to region and over time. One can trace the data required to answer the question by looking at the correlation of Place, Entry, Office, and Kinship.

Historical networks in pre-modern China interacted with – and were informed by – a host of structural factors that shaped social experience. One of the goals of the CBDB is to allow researchers to tease out these interactions.

### 3. Querying the China Biographical Database

The basic strength of the CBDB lies in its structure, as well as the ever-increasing data included in the database. Because the CBDB is complex, however, Michael Fuller has developed a series of forms to help users explore the contents of the database. Once users become familiar with the structure of the CBDB, they can then use the MS Access Query Builder to create customized SQL queries.

### 3.1 The Forms “Look at Entry” and “Look at Office Holding”

Two of the forms, “Looking at Entry” and “Looking at Office” report groups of individuals whose networks can then be explored. Qualifying for government service was one of the major life events for the sons of elite households in imperial China. From the Song dynasty (960–1279) onward, the most prestigious mode of qualifying was to pass the “Presented scholar” examination, which often took years or decades of preparation and many attempts. Passing the examination created networks with the cohort of successful examinees, with the officials overseeing the examination, and with the families of other examination graduates in one’s native community. However, there were other routes to qualifying for government service, such as “yin privilege”, which allowed officials to recommend relatives for admittance to the bureaucracy. Thus a key historical question is the differential patterns in marriage and social networks in which officials who followed different modes of entry participated. The forms “Look at Entry” and “Look at Office Holding” identify the men for these analyses of entry, office-holding, and participation in networks.

“Look at Entry” is perhaps the simplest CBDB form. One simply specifies the target mode of entry or a more general category, a range of years (for either the date of the entry-event or the index year of the person), and a location (either the index place for the person or the place of the entry-event).

Two of these data for individuals – index year and index place – merit additional comment. Because the CBDB is a historical database, it needs to be able to link the people in it to dates (years) that are as specific as possible. We have done this linking through an approach Hartwell started with CHS; since Hartwell was particularly interested in senior officials in the Song dynasty, he most wanted to know when someone turned 60. We adopted this practice and developed an algorithm to deal with the fact that we often do not know the year of birth or death. When possible, we calculate the index year based on a hierarchy of other information to estimate the index year. In our most recent release, however, we have replaced “age 60” with the birth year, either known or calculated. Because geographic information is also centrally important in Chinese prosopography, we treat “index place” in a similar manner to index year. We then use a descending hierarchy of place information to assign the index place.<sup>2</sup>

---

2 Since “place” is an administrative unit in the imperial order – usually a county or a prefecture – and because these designations change over time, the CBDB has developed two ways to handle both the hierarchy of administrative units and the changes across time. “Look at Entry” implements these methods. There are check-boxes on the form to allow the user to include all the subordinate units under the jurisdiction of a target unit (the counties in a selected prefecture, etc.) and to use the coordinates of the administrative seat of the target unit (and its subordinate units) to identify other units in the same location across the specified time period.

Once one has selected the mode of entry and specified the additional parameters, one runs the query. The form in Figure 3 shows the 1402 men who passed “regular examinations” between 1130 and 1160. The attribute data provided by this form, such as each man’s place of origin and the exact year of his examination success, are potentially useful for constructing a two-mode network. Or one can examine the geographic distribution of these men on a map by exporting this result as a GIS file. To look at the other forms of networks in which these people participated, one can also save the IDs for reuse in other queries in the CBDB, such as queries on kinship and social networks, by simply clicking a button.

“Look at Office” is very similar to “Look at Entry.” Since the official bureaucracy is complex, the CBDB allows one to specify any level within that bureaucracy and choose either all the offices attached to the level or select particular offices. One can filter by index year, dynasty, index place, and the location of the office.

The example in Figure 4 is a search for all the people who served as magistrates at the county level in the Southern Song dynasty (1127–1279). The attribute data reported here for the magistrates, like those provided in the query output in “Look at Entry,” may be used to construct a two-mode network, but most of the interest will be in the other types of networks in which the office-holders participated. For instance, since the northern elite had fled to the south after the Jurchen conquest of the north that ended the Northern Song (960–1127), one question is

Name	姓名	Index Yr	Entry Yr	Entry	入仕法	From	地址	地址類別
Zha Yue	查岳	1181	1151	examination: jinshi (regular)	科舉: 正奏名進士	Jiangling	江陵	籍貫(基本地址)
Chao Gongwu	郭公武	1164	1132	examination: jinshi (general)	科舉: 進士(龍統)	Qianshan	鉛山	籍貫(基本地址)
Chen Zhiyuan	陳之淵	1163	1132	examination: jinshi (general)	科舉: 進士(龍統)	Wuxi	無錫	籍貫(基本地址)
Chen Jiong	陳局	1168	1138	examination: jinshi (general)	科舉: 進士(龍統)	Ouning	歐寧	籍貫(基本地址)
Chen Liangbi	陳良弼	1178	1148	examination: jinshi (general)	科舉: 進士(龍統)	Sha Xian	沙縣	籍貫(基本地址)
Chen Mizuo	陳彌作	1168	1138	examination: jinshi (general)	科舉: 進士(龍統)	Min Xian	閩縣	籍貫(基本地址)
Qin Changshi	秦昌時	1154	1142	examination: jinshi (general)	科舉: 進士(龍統)	Jiangning	江寧	籍貫(基本地址)
Zhang Zhen	張震	1181	1151	examination: jinshi (general)	科舉: 進士(龍統)	Mianzhu	綿竹	籍貫(基本地址)
Zhang Fu	章服	1165	1132	examination: jinshi (general)	科舉: 進士(龍統)	Yongkang	永康	籍貫(基本地址)
Zhao Gongcheng	趙公稱	1168	1138	examination: jinshi (general)	科舉: 進士(龍統)	Xingzi	星子	籍貫(基本地址)
Zhao Buyu	趙不愚	1170	1160	examination: jinshi (general)	科舉: 進士(龍統)	Haiyan	海鹽	籍貫(基本地址)
Zheng Zhongxiong	鄭仲熊	1162	1132	examination: jinshi (general)	科舉: 進士(龍統)	Xi'an	西安	籍貫(基本地址)
Jiang Can	蔣堪	1144	1148	examination: jinshi (general)	科舉: 進士(龍統)	Yixing	宜興	籍貫(基本地址)
Zhou Cao	周操	1165	1135	examination: jinshi (regular)	科舉: 正奏名進士	Gui'an	歸安	籍貫(基本地址)
Zhu Guanqing	朱冠卿	1165	1135	examination: jinshi (general)	科舉: 進士(龍統)	Huating	華亭	籍貫(基本地址)
Fan Guangyuan	樊光遠	1161	1135	examination: jinshi (general)	科舉: 進士(龍統)	Qiantang	錢塘	籍貫(基本地址)
Fang Shiyin	方師尹	1178	1148	examination: jinshi (general)	科舉: 進士(龍統)	Yiyang	弋陽	籍貫(基本地址)
Feng Fang	馮方	1175	1145	examination: jinshi (general)	科舉: 進士(龍統)	Anyue	安岳	籍貫(基本地址)
Han Yanzhi	韓彥直	1178	1148	examination: jinshi (general)	科舉: 進士(龍統)	Wu Xian	吳縣	籍貫(基本地址)
He Fengyuan	何逢原	1165	1135	examination: jinshi (general)	科舉: 進士(龍統)	Yongjia	永嘉	籍貫(基本地址)
He Fu	何備	1172	1142	examination: jinshi (general)	科舉: 進士(龍統)	Wu Xian	吳縣	籍貫(基本地址)

Fig. 3 Early Southern Song Entry by Examination in “Look at Entry.”

Name	姓名	Index Year	Gender	Address Type	Place (Person)	Place (Office)
Wang Zongzhe	王宗哲	1148	False	Basic Affiliation	籍貫(基本地址)	Changting
Guan Tingrui	關廷瑞	1270	False	Basic Affiliation	籍貫(基本地址)	Chengdu
You Dong	尤棟	1292	False	Basic Affiliation	籍貫(基本地址)	Wuxi
Kong Yuanzhong	孔元忠	1216	False	Basic Affiliation	籍貫(基本地址)	Changzhou
Fang Lei	方來	1196	False	Basic Affiliation	籍貫(基本地址)	Putian
Ding Mu	丁木	1241	False	Basic Affiliation	籍貫(基本地址)	Huangyan
Fang Fu	方符	1233	False	Basic Affiliation	籍貫(基本地址)	Putian
Wang Xuan	王選	1242	False	Basic Affiliation	籍貫(基本地址)	Jintan
Ding Zongwei	丁宗緯	1241	False	Basic Affiliation	籍貫(基本地址)	Jintan
Fang Zhen(4)	方軫	1130	False	Basic Affiliation	籍貫(基本地址)	Yin Xian
Wang Wei	王維	1187	False	Basic Affiliation	籍貫(基本地址)	Jintan
Wang Yuanshi	王元實	1216	False	Basic Affiliation	籍貫(基本地址)	Yixing
Wang Yuanshi	王元實	1216	False	Basic Affiliation	籍貫(基本地址)	Yixing
Ding Yi	丁倚	1161	False	Basic Affiliation	籍貫(基本地址)	Liling
Fang Xian(2)	方憲	1154	False	Basic Affiliation	籍貫(基本地址)	Putian
Wang Shu	王恕	1229	False	Basic Affiliation	籍貫(基本地址)	Wuyuan

Fig. 4 Magistrates in the Southern Song Dynasty in “Look at Office Holding.”

how the refugee elite reestablished itself in the south. Did the displaced northern magistrates in the early Southern Song, for example, establish kinship and social networks with the local elite families in their jurisdictions? “Look at Office” can provide a list of local officials in the late Northern Song which one can explore from the GIS perspective, and whose networks one can also explore through “Looking at Kinship” and “Looking at Social Networks.”

### 3.2 The Form “Look at Associations”

As of May 2020, the CBDB has records for 149,611 instances of associations for the 472,090 individuals in the database. The CBDB’s form for exploring the basic grouping of people through categories of associations is “Look at Associations.”

As Figure 5 shows, the CBDB records 2,824 associations between people defined through some form of teacher-student relationship. From these data, one can ask whether there were, for example, regional networks of teacher-student relations or whether the networks were empire-wide. When do the networks start appearing in the texts, and do they change over time in size and geographic distribution? The CBDB form allows one to look at large categories of associations or more narrowly at subcategories or specific forms of associations. One can also constrain the search to specific places and time periods. The results can be saved as spreadsheets or as files for further SNA and GIS analysis. Equally significantly,

The screenshot shows the 'Look at Associations' software interface. At the top, there are search filters for 'Select Association' (set to [All]), 'Type' (Teacher-Student), 'Index Years' (From -200 to 1911), 'Dynasties', 'Select Place', 'Import Places', 'All Places', 'Include XY References' (Broad/Narrow), and 'Include Subordinate Units' (checked). Below the filters is a table with the following columns: Name, 姓名, Index ye, Sex, Associate, 社會關係人姓, Assoc. Ind., Assoc. Si, and Association. The table lists various teacher-student relationships, such as Wei Xiang being a student of Zhu Xi in 1189, and Zhang Jun being a student of Shao Yong in 1070. At the bottom, there are options to 'Save to Pajek', 'Save to Gephi', 'Save to UCInet', 'Pinyin', 'UTF-8', 'Big-5', 'GB18030', 'Save to GIS', 'KML', 'Help', and 'Display Language' (Simplified/Traditional).

Name	姓名	Index ye	Sex	Associate	社會關係人姓	Assoc. Ind.	Assoc. Si	Association
Wei Xiang	未詳		M	Zhu Xi	朱熹	1189	M	Menren of
Wei Xiang	未詳		M	Jin Juan	金涓	1350	M	Student was
Wei Xiang	未詳		M	Guang Zhaoyi	晁昭裔	950	M	Menren of
Wei Xiang	未詳		M	Li Sui	李燧	1120	M	Student was
Wei Xiang	未詳		M	Wang Gong	王鏊		M	Student of
Wei Xiang	未詳		M	Xu Bochen	徐伯琛		M	Student was
Wei Xiang	未詳		M	Xiao Yi	蕭頤		M	Student was
Zhang Ju	張巨	1102	M	Ouyang Xiu	歐陽修	1068	M	Student of
Zhang Ju	張巨	1102	M	Hu Yuan	胡瑗	1052	M	Student of
Zhang Xun(4)	張詢	1071	M	Shao Yong	邵雍	1070	M	Student of
Zhang Mian	張沔	1042	M	Yang Yi	楊億	1020	M	followed
Zhang Dong	張洞	1067	M	Sun Fu	孫復	1051	M	Menren of
Zhang Dun	章惇	1094	M	Shao Yong	邵雍	1070	M	Student of
Zhao Ji	趙濟	1090	M	Shao Yong	邵雍	1070	M	Student of
Zhao Ji	趙濟	1090	M	Shao Yong	邵雍	1070	M	followed

**Fig. 5** Teacher Student Associations in Pre-Modern China (“Look at Associations”).

one can save the IDs of the people participating in the associations for other forms of analysis. For example, what did it mean to be a student of someone and to be a member of a cohort of students? Did these students, teachers, and their families form other sorts of kinship and social alliances? Thus, for example, one can save the IDs of the people in teacher-student relations and explore them in the form “Looking at Kinship.”

### 3.3 The Form “Looking at Kinship”

At present, the CBDB has data on 482,979 instances of kinship relations for the 472,090 people in the database. Given the interconnectedness of elites in pre-modern China, the challenge has been how to control the scope of the kinship networks that “Looking at Kinship” dynamically generates from the data.

Because of both the structure of kinship terms and the level of detail recorded for kinship relations, the CBDB has accumulated 479 codes for kinship. To control the size of the kinship networks defined through these many terms, in “Looking at Kinship” we adopted a simple four-value metric for inclusion in networks. While scholars who edit the results of searches for their own use can apply more specific criteria for kinship distance (for which our metric is not a historically meaningful substitute), the metric we apply simply tells the routine when to stop searching. The routine counts the accumulation of ancestor generations



(F [Father], M [Mother]), descendent generations (S [Son], D [Daughter]), sibling links (B [Brother], Z [Sister]), and affinal relations (H [Husband], W [Wife]) as it adds nodes to the network. The texts that preserve kinship relations for pre-modern Chinese individuals use many terms, some of which have no unambiguous translation into English. A *biaodi* 表弟, for example, can be either the younger son of one's father's or mother's sister or the younger son of one's mother's brother. Where possible, the CBDB uses the available data to distinguish between these three options. Still, it turns out that all three (FZS-, MZS-, and MBS-) share the same metric: ancestral generation = 1, descendant generation = 1, sibling = 1, marriage = 0. In constructing kinship networks, CBDB treats adopted siblings, bastard siblings, and half-siblings alike, even though it preserves the particularity of the edge relationship and therefore allows the user to winnow the data as necessary. Disambiguating kinship terms in natural language with a string of basic kinship symbols that can be concatenated in iterative searches lets us determine the four-value metric for kinship distance and produce kinship networks from the CBDB data.

Figure 6 shows the results for Huang Tingjian 黃庭堅 (1045–1105), a prominent Northern Song dynasty cultural figure, when the query metrics are set to 3 ancestral generations, 3 descendent generations, 1 sibling link, and 1 affinal link. If one looks at Huang's kinship data in the basic browser for individuals, the CBDB has only 32 records for instances of kinship for him, while “Looking at Kinship” identifies 295 that meet the defined parameters. Thus iteratively searching through the CBDB's kinship data allows users to discover kinship relations that would otherwise be difficult to discover. A user can quickly experiment with the search results by changing the parameters. We assume that it is better to get too large a network and cull, than to create a network that misses important connections.

The CBDB uses an automatic routine to simplify eight relations produced by the concatenation of edge relations in the construction of the network: BZ > Z, BB > B, ZZ > Z, ZB > B; SB > S, SZ > D, DB > S, DZ > D. That is, in the first iteration the CBDB identifies Li Bu 李布 as Huang Tingjian's maternal uncle (MB, i.e., mother's brother). A search of Li Bu's kinship data then reveals that Li Chang 李常 was Li Bu's older brother (B+) and therefore, with the concatenation, Huang's mother's brother's older brother (MBB+), but this relationship can be reduced to simply another of Huang's mother's brothers (MB). Because each of these simplifications decreases the sibling-link count by 1, it broadens who meets the sibling-link search metric, and the additional nodes will then generate additional nodes and edges to add to the network. Consider the impact of this simplification on the results for Huang Tingjian:

Lady Li (MBB+Z- > MZ), or Li Shi as she is called in Chinese, is Huang Tingjian's mother's sister, but the network arrived at Lady Li through Li Bu (MB) and Li Bu's older brother Li Chang (MBB+ > MB). The routine then added Hong Dan 洪亶 as Lady Li's husband (MBB+Z-H > MZH). We believe that further improving

Looking at Kinship

To save the data to the Clipboard, click on the square in the upper left corner to select all the records, then right-click on it and copy.

Kinship Network Ego-Relative Kinship Network

Name	姓名	Kin Name	親戚姓名	Index Ye	Female	Kinship	Address	地名
Li Bu	李布	Li Dong	李東	1065		F	Jianchang	建昌
Li Bu	李布	Li Chang	李常	1065		B+	Jianchang	建昌
Li Chang	李常	Li Cui	李萃	1086		B+	Jianchang	建昌
Li Chang	李常	Li Dong	李東	1086		F	Jianchang	建昌
Li Chang	李常	Wei Yan	魏琰	1086		W3F	Jianchang	建昌
Li Chang	李常	Wei Yan	魏琰	1086		W2F	Jianchang	建昌
Li Chang	李常	Di Zundu	狄遵度	1086		W1F	Jianchang	建昌
Li Chang	李常	Li Meng(2)	李蒙	1086		D	Jianchang	建昌
Li Chang	李常	Huang Shuao	黃叔敖	1086		DH	Jianchang	建昌
Li Chang	李常	Li Shi(Wife of	李氏(洪喜妻)	1086		Z-	Jianchang	建昌
Li Chang	李常	Sun Duan	孫端	1086		DH	Jianchang	建昌
Li Dong	李東	Li Zhizhi	李知至	1045		F	Jianchang	建昌
Li Dong	李東	Li Cui	李萃	1045		S	Jianchang	建昌
Li Shi(Wife of	李氏(洪喜妻)	Hong Dan	洪璽	1072		H		
Li Shi(Wife of	李氏(洪喜妻)	Li P	李P	1079		F	[Unknown]	[未詳]
Li Cui	李萃	Huang Shuao	黃叔敖	1070		DH	Jianchang	建昌
Du Shenlao	杜莘老	Du Fushi	杜輔世	1164		F	Jiangjin	江津
Du Fushi	杜輔世	Du Zemin	杜澤民	1138		FF	Qingshen	青神
Liang C	梁C	Liang Feng	梁璽	1042		SR	Xuchang	須城

Record: 156 of 295

Save to Gephi Save to UCINET Save to Pajek Pinyin GB UTF-8 Big5 UTF-8 GB18030 UTF-8 KML

Exit

Fig. 6 Huang Tingjian's Kinship Network ("Looking at Kinship").

Looking at Kinship

To save the data to the Clipboard, click on the square in the upper left corner to select all the records, then right-click on it and copy.

Kinship Network Ego-Relative Kinship Network

Name	姓名	Kin	親戚	KinRel to Ego	KinRel calc	Notes
Huang Tingjian	黃庭堅	Li Bu	李布	MB	MB	Notes:
Huang Tingjian	黃庭堅	Li Chang	李常	MBB+	M(BB>B)	Notes: (BB>B)
Huang Tingjian	黃庭堅	Li Shi(Wife of	李氏(洪喜妻)	MBB+Z-	M(BZ>Z)	Notes: (BB>B)(BZ>Z)
Huang Tingjian	黃庭堅	Wei Guan	魏琬	W3FB+	WFB	Notes:
Huang Tingjian	黃庭堅	Zhang Shi(Wife of	章氏(黃璽妻)	FB-W	FBW	Notes:
Huang Tingjian	黃庭堅	Sun Lan	孫覽	WFB-	WFB	Notes:
Huang Tingjian	黃庭堅	Sun Zhiyan	孫直言	AF	AF	Notes:
Huang Tingjian	黃庭堅	Hong Dan	洪璽	MBB+Z-H	M(BZ>Z)H	Notes: (BB>B)(BZ>Z)
Huang Tingjian	黃庭堅	Huang Dalin	黃大臨	FS1	FS	Notes:
Huang Tingjian	黃庭堅	Xie Yin	謝愐	WFS3	WFS	Notes:
Huang Tingjian	黃庭堅	Meng Mou	孟某(孟揚父)	FDH	FDH	Notes:
Huang Tingjian	黃庭堅	Xu Xi	徐禧	FDH	FDH	Notes:
Huang Tingjian	黃庭堅	Chen Shuo	陳瓘	FDH	FDH	Notes:
Huang Tingjian	黃庭堅	Wei Shu	魏紆	W3FS	WFS	Notes:
Huang Tingjian	黃庭堅	Sun Duan	孫端	WFS	WFS	Notes:
Huang Tingjian	黃庭堅	Huang Shuao	黃叔敖	BSF	BSF	Notes:
Huang Tingjian	黃庭堅	Huang Shuxia	黃叔夏	FBS	FBS	Notes:
Huang Tingjian	黃庭堅	Huang Jinggui	黃景珪	FBS	FBS	Notes:
Huang Tingjian	黃庭堅	Xu You	徐祐	FDHB-	FDHB	Notes:

Record: 132 of 245

Save to Gephi Save to UCINET Save to Pajek Pinyin GB UTF-8 Big5 UTF-8 GB18030 UTF-8 KML

Exit

Fig. 7 Huang Tingjian's Ego-Relative Kinship Relations ("Looking at Kinship").



the CBDB's procedures for dynamically handling the simplification of kinship relations that extend beyond sibling relations is one place where future collaboration may prove very fruitful.

### 3.4 The Form “Looking at Social Networks”

“Looking at Social Networks” is the main form for exploring historical networks in the CBDB. As Figure 8 shows, in appearance, the form is far more complex than “Looking at Kinship” because it must account for place and time constraints in constructing networks. The form also allows users to delimit the categories of association around which the network is to be built, and also allows the user to intermix social and kinship networks. Yet the algorithm for finding a stopping point for the search is far simpler than for kinship: one merely specifies the maximum allowed node-distance from the initial target individual(s).

The search for networks produces three types of data: the individual edges, the individual nodes, shown in the tab “People in the Social Network”, and a version of the relationship that merges parallel edges, shown in the tab “Aggregated Social Networks.” The GIS export function saves the place information for the individuals in the network, while exporting to UCInet and Gephi uses the individual edges in “Social Network Relationship,” and files for Pajek use the aggregated edges.

Fig. 8 “Looking at Social Networks,” the Form for Querying Networks.

To make “Looking at Social Networks” easier and more convenient to use, we added additional features. As discussed above, while records for many individuals lack adequate data to determine an index year, we at least know the dynasty, as one can now filter by dynasty. Similarly, because administrative units change name and size in ways that most users lack the expertise to master, we have added the option of using the coordinates for a location (or group of locations) to identify all relevant place codes when searching over long time periods. When including kinship relations in the construction of networks, we optionally use the approach to kinship distance developed to deal with the complexity of kinship relations in “Looking at Kinship.” And users can recall the IDs of people produced in other forms of searches (status, office-holding, mode of entry, place association, etc.) to explore the social networks among those people and, conversely, one can store the IDs of people participating in a network to make the list available for other queries.

### 3.5 The Forms “Look at Pair-Wise Associations” and “Look at Place”

The CBDB offers two additional forms to allow users to explore networks that are otherwise difficult to approach within a large dataset. The first is “Look at Pair-wise Associations.” This form allows one to enter two people (or a list of people) and see the network of associations linking them either directly or through intermediaries. The example in Figure 9 shows the intermediaries between the great Northern Song dynasty literary figure Su Shi and his contemporary, the Neo-Confucian philosopher Cheng Yi, with whom he had a contentious relationship. This search is restricted to older and younger contemporaries and shows 114 instances of relationships among 15 people. That is to say, a substantial community provided lines of communication between these two cultural rivals.

The form allows one to search for links mediated by both one and two persons. The form has wide application for exploring interactions between groups. For example, one can identify Southern Song dynasty Buddhist monks and the Neo-Confucian advocates of the period – who, in theory, were strongly opposed to Buddhism – and see the extent of their intersecting social networks.

The final form, “Look at Place,” provides the user with a list of people who participated in any forms of connection to a specified place: people who served in office there, taught in or were students in an academy there, had family or associates from there, or passed an examination there. Looking at the results, the user may discover overlapping networks that brought people together in a place in unanticipated ways that would otherwise be difficult to sort out from the raw data.

Figure 10 provides the 3,353 records in the CBDB that connect people to Jinhua county in the late Southern Song dynasty. These records are raw data for further analysis, but they conveniently draw all the data together.

Look at Pair-Wise Associations

Recall Person IDs:  Import List of People:  Select First Person:  蘇軾 Su Shi Index Years: From 1050 To 1150 Run Query

Clear List of People:  Select Second Person:  程頤 Cheng Yi

☐ Include Kinship relations ☐ Allow 2-node Intermediaries

☐ No Dates ☒ Use Index Years ☐ Use Dynasties All Dynasties

Name	姓名	Linked to	社會關係人姓	Kin/N	Link	聯
Sima Guang	司馬光	Fan Zuyu	范祖禹	N	Sacrificial prayer written by	祭文由Y所作
Sima Guang	司馬光	Fan Zuyu	范祖禹	N	Sacrificial prayer written by	祭文由Y所作
Sima Guang	司馬光	Fan Zuyu	范祖禹	N	Sacrificial prayer written by	祭文由Y所作
Sima Guang	司馬光	Fan Zuyu	范祖禹	N	Sent letter to	致書Y
Sima Guang	司馬光	Chao Yuezhi	晁說之	N	Recognized the virtue of	節行為Y所稱道
Sima Guang	司馬光	Chao Yuezhi	晁說之	N	Postface of book written by	書跋由Y所作
Sima Guang	司馬光	Chao Yuezhi	晁說之	N	Praised or admired by	被Y欣賞/器重
Sima Guang	司馬光	Wen Yanbo	文彥博	N	Composed Building inscription for	為Y之建築物題詠、記
Sima Guang	司馬光	Wen Yanbo	文彥博	N	Sent letter to	致書Y
Sima Guang	司馬光	Wen Yanbo	文彥博	N	Member of same club (hui, she, et	同會
Sima Guang	司馬光	Wen Yanbo	文彥博	N	Ancestral stele or records written f	為Y作世系碑記
Sima Guang	司馬光	Wen Yanbo	文彥博	N	Prefaced book by	為Y所著書作序
Sima Guang	司馬光	Fan Zuyu	范祖禹	N	Recommended	推薦
Sima Guang	司馬光	Fan Zuyu	范祖禹	N	Preface of book by	書序由Y所作
Su Shi	蘇軾	Li Zhichun	李之純	N	Supported by	得到Y的支持
Su Shi	蘇軾	Li Qingchen	李清臣	N	Sent letter to	致書Y
Su Shi	蘇軾	Xie Jingwen	謝景溫	N	Impeached by	被Y彈劾
Su Shi	蘇軾	Xie Jingwen	謝景溫	N	Impeached by	被Y彈劾

Record: 1 of 114 No Filter Search

Store Person IDs:  Save to UCINET  Save to Gephi  Save to Pajek ☒ UTF-8 ☐ GB18030  Save to GIS ☐ KML  Help  Display Language:  簡體  繁體

☐ Remove 0-degree ☐ Include Person ID

Fig. 9 “Look at Pair-Wise Associations.”

Look at Place 查詢地區關係

Select Place:  Jinhua ☐ Use XY References ☒ Include Subordinate Units Index years: From 1200 To 1320 繁體

Import Places:  金華 ☒ Use Index Years ☐ Use Dynasties All Dynasties To 簡體

Name	姓名	Index Yea	Place Name	地名	Assoc. Name	有關名	Category
Xu Gong	許翬	1277	Jinhua	金華	Han Shi(Wife o	韓氏(許翬妻)	Kinship
You Kui	游夔	1202	Jinhua	金華	Chen Shi(Moth	陳氏(游夔母)	Kinship
Tang Zhongyou	唐仲友	1188	Jinhua	金華	He Song	何松	Kinship
Zhou Yanzhao	周彥昭	1119	Jinhua	金華	Zhou Shi(Wife	周氏(戚揚之妻)	Kinship
Murong Yanfeng	慕容彥逢	0	Jinhua	金華	Shan Zhao	單照	Kinship
Pan Jingliang	潘景良	1200	Jinhua	金華	Lv Huanian	呂華年	Kinship
Wu Zi	吳詒	1148	Jinhua	金華	Wang Yang	王洋	Kinship
Wang Hao	汪浩	1182	Jinhua	金華	Wang Shi(Wife	王氏(汪浩妻)	Kinship
Zheng Yuzeng	鄭與曾		Jinhua	金華	Zheng Gangzho	鄭剛中	Kinship
Wang Pu	王溥	981	Jinhua	金華	Wang Conghao	王從浩	Kinship
Su Taigu	蘇太古	1283	Jinhua	金華			Office Place
Dong Shu	董銖	1211	Jinhua	金華			Office Place
Dong Shu	董銖	1211	Jinhua	金華			Office Place
Zhao Gongsheng	趙公升	1202	Jinhua	金華			Office Place
Dai Ji	戴機	1201	Jinhua	金華			Office Place
Wang Bi(2)	王泌	1250	Jinhua	金華			Office Place
Chen Tianrui	陳天瑞	1299	Jinhua	金華			Office Place

Record: 1 of 3353 No Filter Search

Run Query ☒ Individual ☒ Entry ☒ Association ☒ Office Posting  Store Person IDs  Save to Pajek  Save to UCINET  Save to Gephi

☒ Institutional ☒ Kinship ☒ Associate ☒ UTF-8 ☐ Big-5 ☐ GB18030 ☒ UTF-8 ☐ GB18030

Fig. 10 People Associated with Jinhua County in the Southern Song Dynasty.

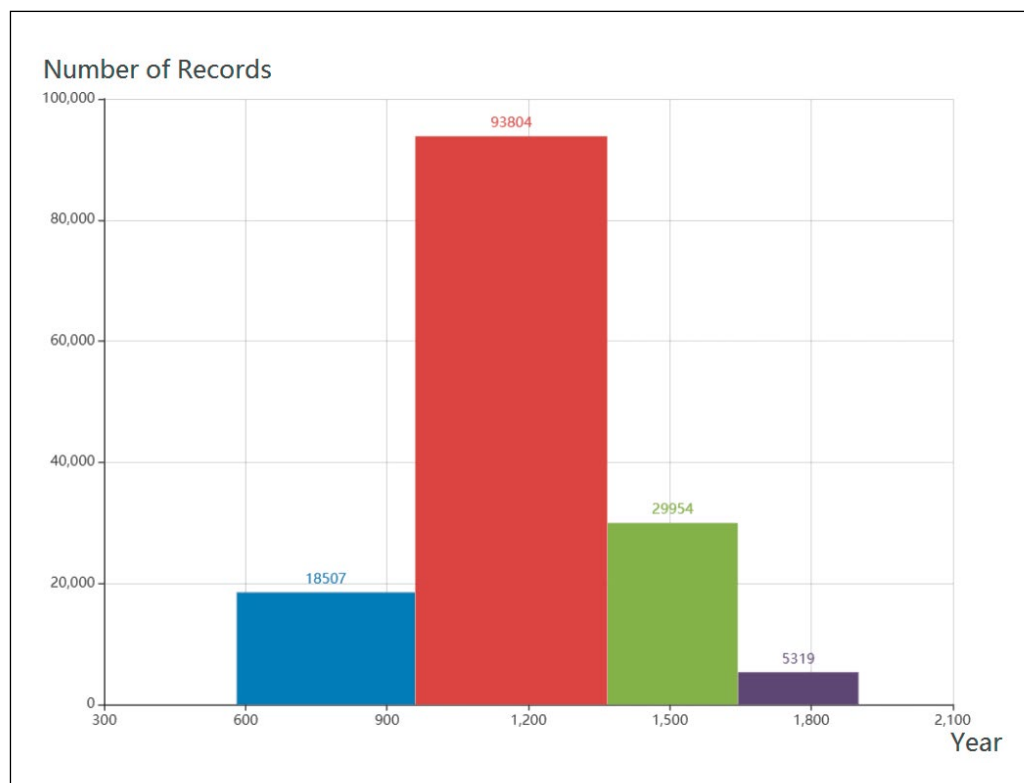
## 4. A Summary of Network Data in the China Biographical Database

### 4.1 Data on Non-Kin Social Associations

To support prosopographical research, the China Biographical Database had collected 482,979 instances of kinship relations and another 149,610 instances of non-kin social associations by May 2020. Most of our data concern the period between 600 and 1912 CE (Figure II).

Reflecting the history of the CBDB, whose earliest developers, including Robert Hartwell, are predominantly scholars specializing in the Tang (618–907) and Song (960–1279) dynasties, until recently more than 75% of its data on social association pertained to this period. It is our plan, however, for the next few years to significantly expand its coverage by collecting large amounts of data on social associations between the fourteenth and nineteenth centuries.

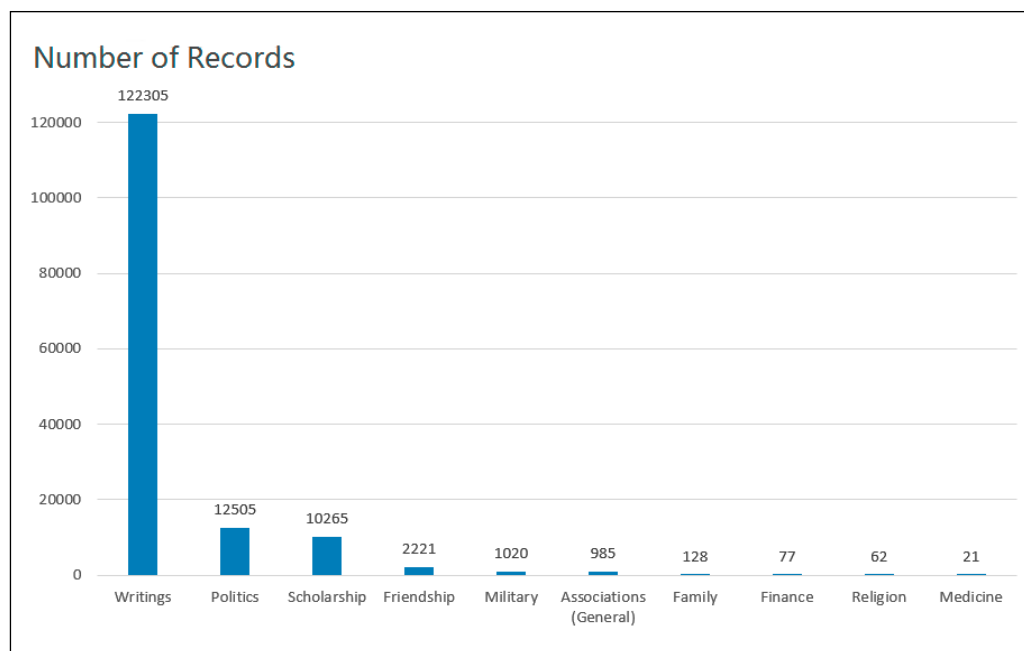
To capture the nuances of different types of relationships found in the historical record, but also provide convenience to general users, the CBDB adopts a two-



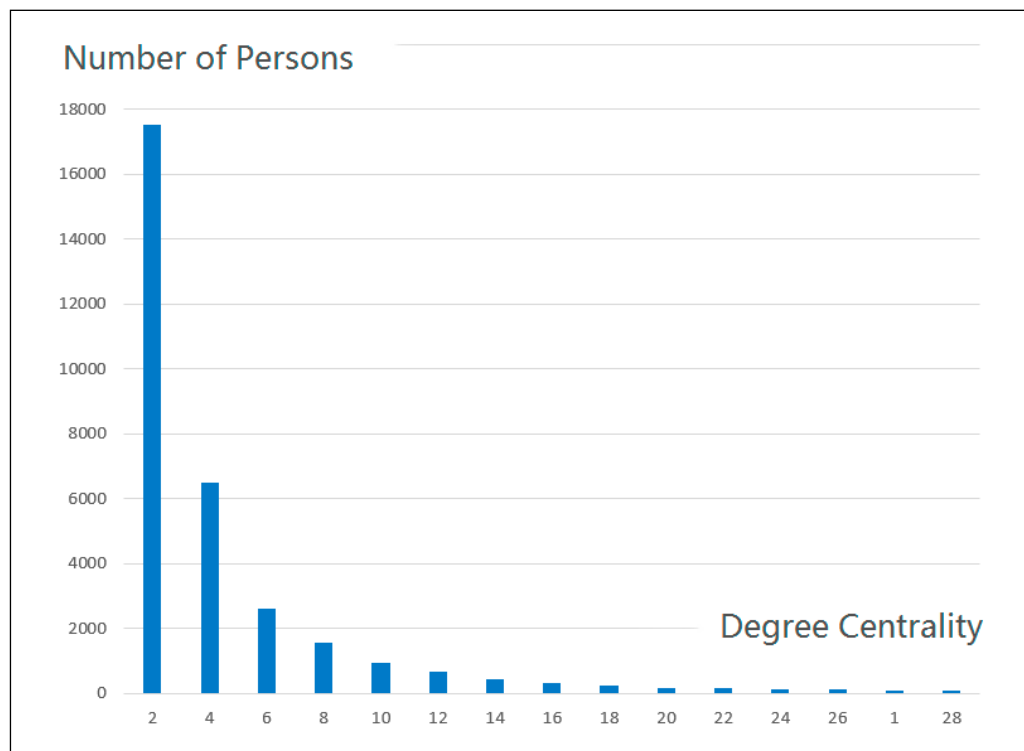
**Fig. II** The temporal distribution of CBDB data on social associations.

tier approach in recording social connections. First, we define 480 types of social association, every two of which form a pair, such as “epitaph written by” and “epitaph written for,” “purged” and “purged by,” etc. We then classify these different types of associations into ten broad categories, among which writings, politics, and scholarship have the most records (Figure 12). This is a result of the nature of the CBDB’s data sources. In sharp contrast to Europe, where church records and bank records abound, the Chinese historical record before the twentieth century is rich in political, scholarly, and literary interactions. The CBDB systematically harvests data from biographical indexes, literary collections, and local gazetteers, which document these social interactions using formulaic expressions conducive to semi-automated data extraction.

A visualization of all social association data in the CBDB generates a massive network with 33,433 persons (nodes) and 149,610 connections (edges). Where multiple instances of associations are documented for a given pair of nodes, duplicate edges are counted and removed and the count is added to the remaining edge as its weight. This reduces the total number of edges in the network to 55,799. Analyzed as an undirected network, it has an average degree of 3.35, and a network diameter of 19. The average path length in the network is 5.509, suggesting that on average a node is 5 to 6 steps away from other nodes in the network.



**Fig. 12** The number of records for different social association categories in the CBDB.



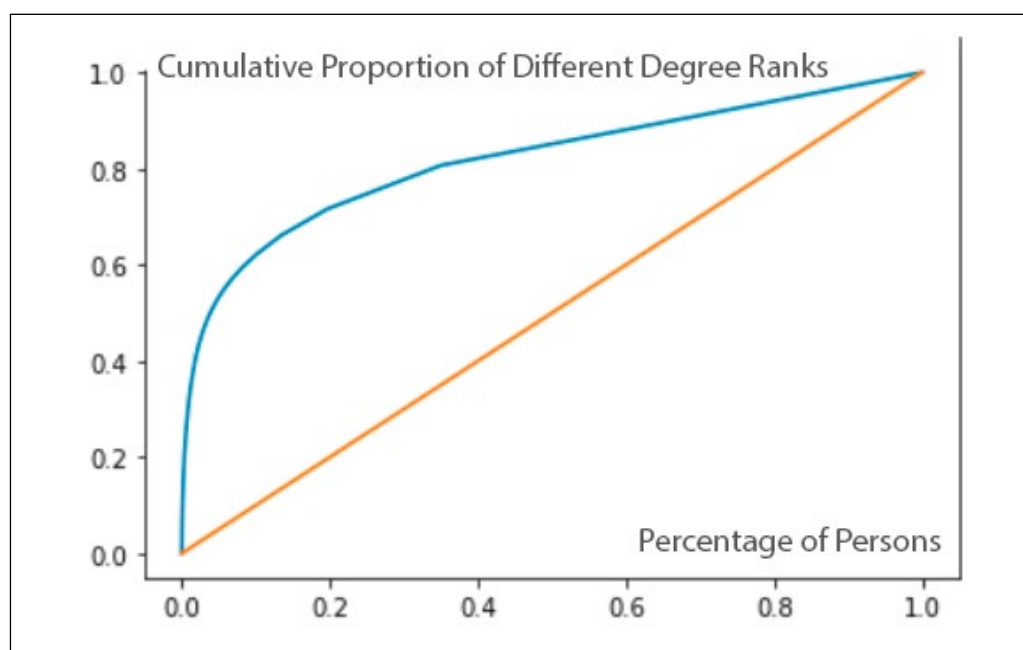
**Fig. 13** The distribution of weighted degree centrality values in the CBDB social association network (number of persons  $\geq 100$ ).

The distribution of the weighted degree centrality of each node follows a power-law curve: 90.20% of persons have a degree centrality of only four or less, whereas a tiny percentage (3.69%) of persons have a degree centrality above 10 (Figure 13).

Likewise, the network's Gini coefficient (0.63), computed on the basis of different degree ranks, indicates a high degree of inequality in how social connections are distributed among the persons in the CBDB (Figure 14).

One may also examine the network's clustering tendency by comparing it to the Erdős-Rényi random graph, where  $n = 33,433$  (nodes),  $m = 55,799$  (edges),<sup>3</sup> there are on average 119 3-cliques but only  $1.91 \times 10^{-5}$  4-cliques. In other words, in a random graph, the possibility of having a  $k$ -clique (i.e., a subgroup composed of  $k$  nodes and where each node in the subgroup has a tie to every other node) in the network where the  $k$  value is greater than or equal to 4 is very low. In contrast, the observed network in CBDB reports a total of 313 3-cliques, 154 4-cliques,

3 The value  $p(\text{prob.})$  of Erdős-Rényi model for CBDB data is  $p = m/\binom{n}{k} = 0.000267$ .



**Fig. 14** The Lorenz curve of the degree centrality ranks for CBDB social association data.

54 5-cliques, and 27 6-cliques. This implies a strong tendency towards local clustering in the CBDB that deserves the attention of humanities scholars. It is also noteworthy that while the observed network in the CBDB includes 1099 components, 92% of the persons are members of the giant component.

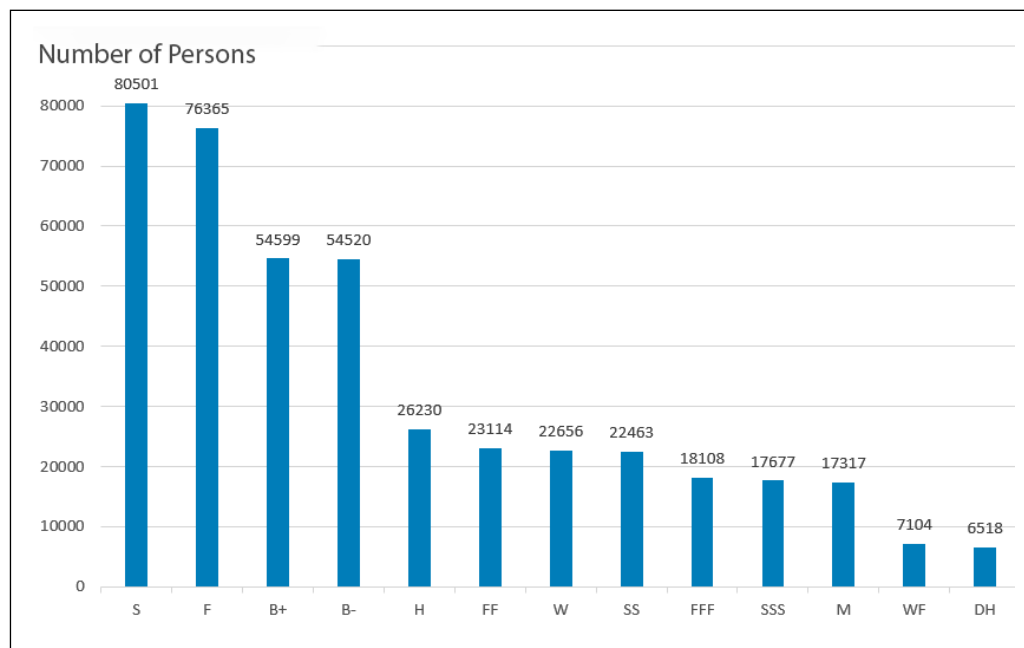
## 4.2 Data on Kinship Ties

The CBDB has recorded 481,476 kinship ties that link together 244,658 persons. These ties are classified into 479 types, of which thirteen have more than 5,000 records: son (S), father (F), elder brother (B+), younger brother (B-), husband (H), grandfather (FF), wife (W), grandson (SS), great-grandfather (FFF), great-grandson (SSS), mother (M), wife's father (WF) and daughter's husband (DH) (Figure 15).

Analyzed as undirected and unweighted, this kinship network has an average degree of 1.97, an average path length of 24.235, and a diameter of 79.<sup>4</sup> This network

4 The CBDB records only kinship ties explicitly mentioned in our data sources. As illustrated by the example of Huang Tingjian in the preceding section, for example, one of a person's maternal uncles may be directly registered under that person, while his other





**Fig. 15** The number of records for different types of kinship ties in CBDB (number of records > 5000).

is more fragmented than the non-kin social network. It contains 23,532 components, but only 51,738 persons (i.e., 21.15% of the total) are part of the giant component, and even the second largest component has only 357 nodes.

15,573 persons in the kinship network are also found in the non-kin social network. Therefore,

$$\frac{\text{Number of overlapping nodes between CBDB kinship and social networks}}{\text{Number of nodes in the CBDB social network}}$$

$$= \frac{15,573}{33,433} = 46.58\%$$

---

maternal uncles and aunts may not, despite their presumably equally close ties to that person. Therefore, some network metrics – such as network density, diameter, average degree, and average path length – are not accurate measures of the structural properties of a network constructed from unprocessed CBDB kinship data. It is more meaningful to interpret these metrics as indicating how kinship data are registered in the CBDB and its data sources.



On the other side,

$$\frac{\text{Number of overlapping nodes between CBDB kinship and social networks}}{\text{Number of nodes in the CBDB kinship network}}$$

$$= \frac{15,573}{244,658} = 6.37\%$$

These numbers suggest that a person's relatives are more likely to enter the historical record if that person's social connections are recorded in the historical sources. This implies that family members of a socially prominent individual tended to have a greater chance of being influential.

## 5. Extracting Data from Texts

### 5.1 Identifying Suitable Historical Sources

The CBDB prioritizes the collection of data from digitized texts that contain large amounts of reliable biographical information and present them in a clear and consistent format, which makes it possible for us to extract data from these texts in a systematic manner using computational methods. The systematic harvesting of data has not only increased the efficiency of expanding our data coverage, but it also ensures that research using CBDB data is statistically meaningful. For example, *Quan Song wen*, a collection of all preserved prose style literature from the Song dynasty (960–1279), is an ideal source for data collection. Its table of contents, in essence a nearly exhaustive list of more than 178,000 unique pieces of writing from the three centuries, includes letters, funerary biographies, elegies, commemorative inscriptions for private and public constructions, etc. that signify different types of social relationships.

### 5.2 Acquiring and Preparing Texts for Data Mining

Whenever possible, the CBDB prefers to work with texts for which searchable digitized editions already exist, and that the copyright owner agrees to share these editions with us for data mining. In some cases, where working with non-digitized texts is inevitable, we have to assess whether it is more efficient to create a digital edition by ourselves or simply have someone type the data directly into the computer. We have developed an open-source online inputting system for data entry. But for some texts, we have editors first enter data into spreadsheets, then code and upload them to the database in batches. We find simple spreadsheets with a limited number of value fields most effective for this purpose.

When the need arises for us to create a digital edition by ourselves, we scan the documents and use Abbyy, an optical character recognition (OCR) package,

to transform them into searchable text files. The precision of OCR depends on a number of factors, such as the layout and quality of the scanned pages. We typically start off by checking the precision rate of a few sample pages and proceed only if it exceeds 95%.

### 5.3 Data Mining

Patterns of language use vary from text to text. To efficiently extract data, we have employed different data mining techniques for different texts. The most common technique relies on regular expressions. The basic idea here is string searching based on patterns of language use. Therefore, it works best only on texts that use a finite number of formulaic expressions, which has the benefit of not requiring any training data. Take teacher-disciple relations, for example. A common expression of this relationship in ancient Chinese texts describes the disciple as “following” (*cong* 從) the teacher in his “travels” (*you* 遊). One typically encounters the teacher’s name sandwiched between these two Chinese characters meaning “follow” (*cong*) and “travel” (*you*) respectively, and therefore to find the teachers named in a person’s biography, we can ask the computer to scan through the document for all text strings that have this pattern.

For texts where patterns of language use are less obvious and more variegated, regular expressions have to give way to more sophisticated methods. We have tried several methods of machine learning for different genres of texts, with varying degrees of success.

Random forest is a supervised learning algorithm that we have been experimenting with to extract social association data from the biographies. First, we collect from this corpus all the sentences that contain two person names and use a small number of these sentences as the training data. Next, we use a feature vector to represent each sentence in the training data based on the frequency of each word in the sentence. We then read each sentence in the training data and assign it to a specific type of social association that CBDB has coded. This allows us to gain knowledge of how the mathematical properties of each feature vector, which represents a sentence in the training data, correlate with the type of social association it expresses. In the final step, we apply this knowledge to those sentences that are not yet in the training set, trying to predict what type of social association they express based on their mathematical properties. Thus, we in effect resolve the problem of detecting and extracting network data in a corpus by transforming it into a problem of identifying sentences with name collocations and then classifying them into a finite group of social association types.

Unlike random forests that handle the extraction of named entities as a classification problem, BiLSTM-CRF approaches it as a sequential tagging problem. Compared to random forests, we find the combination of BERT and BiLSTM-

CRF particularly effective. BERT is an unsupervised language representation technique that transforms texts into vectors, but BERT surpasses all previous techniques for being deeply bidirectional and taking into account the context for each occurrence of a given word (here, a given Chinese character). After vectorizing our texts with BERT, we apply the state-of-art sequence tagging model, called BiLSTM-CRF (Bidirectional-Long Short Term Memory-Conditional Random Field), to predict whether the occurrence of a given character in our texts is part of a person's name, official title, place name, kinship term, or something else. This prediction is based on both the immediate context of each occurrence of the character (the Bidirectional LSTM model), but also on more general knowledge of how the character is used in the entire training set (the CRF model). This method is particularly useful for handling Chinese texts, where the names of entities (e.g., persons, offices, and places) have no obvious stylistic features, such as the use of uppercase letters, and are not separated from other words in the sentence by white spaces. This proved successful when we applied it to local gazetteers, from which we extracted biographical information, including the officials' degrees, careers, and kinship relations.

#### 5.4 Data Standardization

Our data mining algorithms typically export the results in the format of a collection of tagged XML documents or a large spreadsheet that includes information on a person's relationship to many different entities (e.g., offices held, social associations, kinship ties). Since the CBDB stores each entity in a separate code table and the relationship between each entity and a person in a separate data table, we convert these tagged texts and large spreadsheets into a group of linked tables that fit the CBDB data model.

The principal challenge in this step is disambiguation. The CBDB uses a unique ID for each person, as it does for other entities, too. Yet a person may appear in historical sources under different names, and two persons may have also had exactly the same name. Before uploading newly harvested data to the CBDB, we disambiguate each occurrence of a person's name within the new dataset and against existing data in the CBDB. Although we occasionally seek advice from historians specializing in the relevant topic or period, the sheer size of our data requires that we automate this process as much as possible. To disambiguate person names, we draw upon our knowledge of the data source, attribute data to each person, and also his or her networks. We assume, for example, a name that repeatedly appears in different chapters of the same book most likely refers to the same person. We assume that persons who had the same name but hailed from different places, lived in different centuries, or passed the civil service examination in different years were merely homonymic by chance. These biographical details, however, are not always available. It is worth noting here that we have also found network data harvested by the CBDB useful for disambiguation purposes. In recent years, we have had success in developing disambiguation algorithms on the assumption

that two “persons” with the same name whose kinship and non-kin social networks overlapped significantly were most likely the same person.

## 6. Appendix: Publications Using CBDB Social Association Data

- Bol, Peter. “Changing Literati Networks: Kinship and Collegiality, 1100–1400.” *Journal of Historical Network Research* 5 (2021): 87–113.
- Chen, Song 陳松. “Governing a Multicentered Empire: Prefects and Their Networks in the 1040s and 1210s.” In *State Power in China, 900–325*, edited by Patricia Ebrey and Paul J. Smith, 101–152. Seattle: University of Washington Press, 2016.
- De Weerd, Hilde, Brent Ho, Allon Wagner, Qiao Jiyan, and Chu Mingkin. “Is There a Faction in This List?” *Journal of Chinese History* 4, no. 2 (2020): 347–389.
- Hsu, Ya-hwei 許雅惠. “Bei Song wanqi jinshi shoucang de shehui wangluo fenxi” 北宋晚期金石收藏的社會網絡分析 [The Social Networks of Antiquities Collectors in the Late Northern Song]. *Xinshixue* 新史學 29, no. 4 (2018): 71–124.
- Liu, Feiyan 劉飛燕, and Gao Jianbo 高劍波. “Sui Tang zhi Song shiqi jingying shehui wangluo donglixue de yanhua yanjiu” 隋唐至宋時期精英社會網絡動力學的演化研究 [Dynamical Evolution of Social Networks of Elites from Sui-Tang to Song Dynasty]. *Shuzi renwen* 數字人文 1 (2020): 118–127.
- Tackett, Nicolas. “The Evolution of the Tang Political Elite and its Marriage Network.” *Journal of Chinese History* 4, no. 2 (2020): 277–304.
- Yan, Chengxi 嚴承希, and Wang Jun 王軍. “Shuzi renwen shijiao: jiyu fuhao fenxifa de Songdai zhengzhi wangluo keshihua yanjiu” 數字人文視角：基於符號分析法的宋代政治網絡可視化研究 [Digital Humanistic Perspective: A Study on the Visualization of Political Network in Song Dynasty Based on Symbolic Analysis]. *Zhongguo tushuguan xuebao* 中國圖書館學報 44, no. 5 (2018): 87–103.



The Journal of  
**HISTORICAL  
NETWORK  
RESEARCH**

MARILYN LEVINE

# Biography for Historical Analysis: A Chinese Biographical Database

*Journal of Historical Network Research* 5 (2021) 271–287

**Keywords** Chinese Biographical Database, Chinese Communist Party, Chinese Nationalist Party, quantitative history, network analysis

**Abstract** Biography is a common approach to examining history as event and process. Whether leaders or followers, extraordinary or average individuals, together they are a part of the historical process. The Chinese Biographical Database (CBD), originally created in 1997–98, was made available online through 2006 and was intended as a research tool to aid in understanding biography and history. The CBD originated in the study of Chinese politics in Europe during the 1920s and included information on more than 1,100 Chinese who went to France, Germany, and Belgium, and later expanded to include other Chinese individuals, creating a diverse array of people in politics, academics, military, and other spheres. Data attributes included basic biographical information such as names, gender, birthplace, birth and death dates, main career, affiliations, education, positions, youth activities, historical events, family members, alternate names, sources, locator data, and multimedia. Since 2018, the CBD has been transformed into an analytical resource with updated information for 2,109 individuals and 840 attributes. This article will introduce the CBD and then discuss the historical context, documentation, composition, and future of the CBD, as well as presenting two examples of its utilization for historical network analysis.

## 1. Introduction\*

The Chinese Biographical Database (CBD) is a resource that was created over thirty years ago.<sup>1</sup> It emerged out of two studies on Chinese political party development in Europe during the 1920s, including international archival research and a series of transcribed Chinese interviews in 1985 and 1990 with CBD individuals, their relatives, and the CCP (308,000 characters). The CBD had as its core philosophy the intent to develop multiple methods for highlighting information on Chinese individuals who lived during one of the most creative decades of the twentieth century. The CBD was a pioneer in forwarding the idea of a shared scholarly resource and was online from 1998 through 2006. The online CBD displayed over two dozen reports and queries, received numerous accolades, more than a million visits, and the basic report form was linked on the Chinese Worldwide Web (WWW) Virtual Library site.<sup>2</sup>

Although the CBD was utilized, it did not achieve its original intention of on-line collaboration, with scholarly moderation of new data and figures. Several challenges to establishing a collaborative online effort became apparent. Some of these challenges were slow, less reliable connection speeds (via dialup), and the WWW only gained its first browsers in 1993 (Mosaic, Netscape). In 1994 there were just 2,738 websites, and when the CBD was being developed there were 257,601 websites (1996), while in 2020 there are over 1.8 billion websites.<sup>3</sup> Be-

---

\* **Acknowledgements:** I would like to acknowledge two grants that allowed the formation of the CBD from the Chiang Ching-kuo Foundation and the Idaho Higher Education Major Research Grant program; I valued from collaborations and help from Chen San-ching, Yves Chevrier, Christian Henriot, Jean-Louis Bouilly, Liu Guisheng, and Zhu Yuhe. For original entry in the CBD, the help of Zhou Baodi and Eric Barnes was excellent and in updating the database I am indebted to Howie X. Lan, John Bowen, Tan Hongxing, and Wu Yiwei.

**Corresponding author:** Marilyn.Levine@cwu.edu

- 1 Abbreviations used in this article include Chinese Biographical Database (CBD), Soviet Returned Leaders data subset (SRL), Chinese Students in Europe data subset (CSE), Sino-French Institute at Lyon (SFI), European Branches of the Chinese Communist Organizations (ECCO), European Branch of the Chinese Nationalist Party (EGMD), Chinese Communist Party (CCP), Chinese Nationalist Party – Guomindang (GMD). The Levine CBD should not be confused with the Harvard CBDB (China Biographical Database), discussed elsewhere in this issue by Michael Fuller and Wang Hongsu, which emerged after the CBD and focuses on premodern China.
- 2 For more information on the collaborative goals and impact of the CBD see: Marilyn Levine, “The Chinese Biographical Database Project: A Model for Collaborative Scholarship,” in *Proceedings of the EBTI, SEER, ECAI, and PNC Meeting* (Taipei: Pacific Neighborhood Consortium, 1999), 514–20; Marilyn Levine, “Modern China and the Chinese Biographical Database” (First International Workshop on Biographical Databases for China’s History, Harvard University, 2010). In addition, numerous speeches and workshops were given about the CBD in Europe, Asia, Australia, and the United States.
- 3 “Total Number of Websites – Internet Live Stats,” <https://www.internetlivestats.com/total-number-of-websites/#trend>, accessed August 17, 2020.

sides technical challenges of the time, there were utilization and attitudinal challenges. The actual audience of scholars for historical biographical data was rather small, and web-based collaborations were more focused on creating digital resources – digitalization, not utilization. In terms of the usage of the CBD, more important for viewers was the capacity to extract information on CBD biographies through more than a dozen queries and reports.

After a ten-year hiatus, the CBD was reprised, with different aspirations, including the idea of utilizing the data for statistical and network analysis. Two simple primary questions were asked: (1) Would quantitative and network analyses mirror actual history? (2) Would there be any new discoveries in such analyses? Some secondary questions were: What were the impacts of shared regional origin, political, educational, career, political, or cultural affiliations? Could one ascertain common generational trends? How does a network view of political leadership alter our understanding of leadership patterns in modern China?

To utilize analytical programs, the data were first exported to a single workbook, with the original tables as separate worksheets, followed by a macro that merged the worksheets into a single spreadsheet for seven of the tables (see Table 1). Latitude and longitude values were added in the basic biographical table. This allowed the usage of statistical, network, and geospatial programs. The primary spreadsheets exist in 2-mode (individuals vs all attributes) and are used for multivariate and centralization analysis, but 1-mode (individual vs individual) datasets were also created from 2-mode datasets through sums of cross-products for most network analyses. One must be aware that the attributes are only *potential affiliations* of ties, and the binary attributes are undirected.

Some confusion may exist surrounding 2-mode (rectangular person x attributes) conversion to 1-mode (symmetric person x person) datasets. One common example is the sum of cross-products method (Bonacich, 1972) where both the sums of cross-products as well as their normalization are produced. An example from a CBD data subset of three individuals and five attributes can be seen in Table 2, with resulting sums of a cross-products matrix in Table 3 that demonstrates the conversion to 1-mode.<sup>4</sup>

---

4 For more in-depth information, see: Stephen P. Borgatti, “Two-Mode Concepts in Social Network Analysis,” *Encyclopedia of Complexity and System Science* 6 (2009): 8279–91; Stephen P. Borgatti and Daniel S. Halgin, “Analyzing Affiliation Networks,” in *The SAGE Handbook of Social Network Analysis*, by John Scott and Peter Carrington (London, UK: SAGE Publications Ltd, 2014), 417–33, <https://doi.org/10.4135/9781446294413.n28>. Also see: P. Bonacich, “Factoring and Weighting Approaches to Status Scores and Clique Identification,” *Journal of Mathematical Sociology* 92 (1972): 1170–82.

Table Name*	N. Individuals	Attributes	Range of Attributes
1. Basic Biodata	2,109	46	Full name, Chinese characters (traditional and simplified) birth year, birth date, death year, lifespan, birth city, birth province, latitudes/longitudes for birth cities, gender, provinces (binarized columns), web links, and 2,093 comments.
2. Career	1,252	40	Careers range from academic categories to veterinarian.
3. Affiliations	2,242	50	Focus on affiliations in the 20 <sup>th</sup> century with the majority political.
4. Education	1,601	570	Most of the institutions are post-secondary, many in Europe.
5. Positions	2,954	45	Wide range of positions in politics, culture, and military areas.
6. Youth	331	17	Emphasis on May Fourth groupings.
7. Historical Events	1,696	72	Focus on 20 <sup>th</sup> century events in China.
<b>Total:</b>	<b>12,185</b>	<b>840</b>	

**Tab. 1** Chinese Biographical Database structure.

Attribute						
No.	Individual	May 4 <sup>th</sup>	ECCO	Central Comm.	1927 Nanchang	CPPCC
1	Zhou Enlai	1	1	1	1	1
2	Luo Yinong	1	0	1	1	0
3	Dai Kunzhong	1	1	0	0	0

**Tab. 2** Three example individuals versus five binarized attributes (2-mode).

Example sums of cross-products calculated for Zhou Enlai versus Luo Yinong as:  
 $\text{sum} = 1 \times 1 + 1 \times 0 + 1 \times 1 + 1 \times 1 + 1 \times 0 = 3.$

	1	2	3
1	5	3	2
2	3	3	1
3	2	1	2

**Tab. 3** Resulting sums of cross-products matrix (1-mode) for three individuals.



Some of the advantages of conversion to 1-mode with sums of cross-products if the data are binary (1, 0) are: the final sum per individual is the degree centrality; the sums are a multivariate measure frequently spanning over 200 attributes yielding an order of magnitude or more statistical advantage; and the sums can be calculated with normalization, allowing different dataset comparisons.

The most cited earliest reference of a 2-mode study that also was converted to a 1-mode matrix for network analysis is one on Southern women (1941)<sup>5</sup> which was a dataset of 18 individuals and 14 attributes, where the attributes were: “a day’s work behind the counter of a store, a meeting of a women’s club, a church supper, a card party, a supper party, a meeting of the Parent-Teacher Association, etc.”<sup>6</sup> These attributes were an example of biographical data for these Southern women, albeit very abbreviated, with only 14 attributes, although it was successfully used in clique analysis, as demonstrated in a meta-analysis by Linton Freeman of 21 major studies based on the Southern women dataset.<sup>7</sup> Like the Southern women, the CBD allows either 2-mode or 1-mode utilization.

## 2. “To Save China!” A Historical Context of the CBD Individuals

The CBD concentrates on twentieth-century Chinese leadership. It emerged out of research from two studies of the development of five Chinese political parties in Europe during the early 1920s.<sup>8</sup> China was undergoing a generational revolution after the Republican Revolution of 1911, when the Qing dynasty was overthrown. With the abolition of many traditions, change was led by teachers and their students in the New Culture Movement that began in 1915. The West was looked upon as a model for national salvation, both intellectually and technologically. Several giants of Chinese education who were oriented towards Europe, such as Li Shizeng (李石曾 1881–1973), Cai Yuanpei (蔡元培 1868–1940), Wu Zhihui (吳稚暉 1865–1953), and Wang Jingwei (汪精衛 1883–1944) organized the Diligent-Work Frugal-Study Movement (勤工儉學運動), whereby over 1,800 Chinese youths went to French-language preparatory schools and then traveled to

---

5 Allison Davis et al., *Deep South: A Social Anthropological Study of Caste and Class* (Chicago: University of Chicago Press, 1941).

6 G. C. Homans, *The Human Group* (New York: Harcourt, Brace and Company, 1950), 82.

7 Linton Freeman, “Finding Social Groups: A Meta-Analysis of the Southern Women Data,” in *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, by Ronald L. Breiger, Kathleen M. Carley, and Philippa Pattison (Washington, D.C.: National Academies Press, 2003), 1–39.

8 Marilyn A. Levine, *The Found Generation: Chinese Communists in Europe during the Twenties* (Seattle: University of Washington Press, 1993); Marilyn A. Levine and San-ching Chen, *The Guomindang in Europe: A Sourcebook of Documents* (Berkeley, Calif.: Institute of East Asian Studies, University of California, Berkeley, Center for Chinese Studies, 2000).

France between 1919 and 1921. The concept was based on a calculation that with losses of young French men in WWI, French factories could employ young Chinese, who would save their earnings and attend the French colleges to learn about new ideas and technology to save their homeland.

Unfortunately for the worker-students, there were not enough factory jobs, and the movement was halted in January 1921, when economic support for these youths was withdrawn. Three political struggles followed in the Chinese community in France, including: an unsuccessful demonstration led by a faction of Chinese youths who were working and studying in Montargis and requested government support; a successful attempt during the summer to block a French bank loan to the Chinese warlord government, known as the loan struggle; and a culmination of frustration in what became known as the Lyon struggle, which involved the occupation of a dormitory at the University of Lyon to demand matriculation at the newly opened Sino-French Institute (SFI). Many of the worker-students had relied on the SFI opening as the educational path forward they had been promised by the promoters of Work-Study. Yet these same promoters selected all new students, who were then being escorted from China to Lyon, to be the first entering class at the SFI. Over 104 worker-students were arrested for occupying the dormitory and within a couple of weeks were deported back to China. The remaining Chinese had diverse reactions to this defeat. Some accepted free passage and were repatriated; some continued work they could find in the factories; others stayed in Europe and attended universities, obtained degrees and became celebrated scholars back in China; and more than a thousand Chinese (over 50 percent of the 1,800 worker-students) who had been exposed to the various ideologies that emerged after WWI decided to form political parties that were based on nationalism and international communism, among other ideologies.

The activities of these Chinese political parties were robust, including membership recruitment, numerous publications, meetings and demonstrations, even occupying the Chinese legation during the May 30<sup>th</sup> incident in 1925, which resulted in more deportations. Some of the most famous Chinese revolutionaries and politicians emerged from these activities, including Zhou Enlai (周恩來 1898–1976), Zhu De (朱德 1886–1976), Deng Xiaoping (鄧小平 1904–1997), Nie Rongzhen (聶榮臻 1899–1992), Cai Chang (蔡暢 1900–1990) and Li Fuchun (李富春 1900–1975). Other well-known Chinese party leaders developed strong ties in Europe, including all three factions of the GMD.<sup>9</sup> It is the breadth

---

9 Marilyn Levine, "The Diligent-Work, Frugal-Study Movement and the New Culture Movement," *Republican China* 12, no. 1 (1987): 67–88; Paul Bailey, "The Chinese Work-Study Movement in France," *The China Quarterly*, no. 115 (1988): 441–61, <http://www.jstor.org/stable/654865>; Levine, *The Found Generation*; Marilyn A. Levine and San-ching Chen, "Communist-Leftist Control of the European Branch of the Guomindang, 1923–1927," *Modern China* 22, no. 1 (1996): 62–92, <http://www.jstor.org/stable/189290>; Tsin-

of these types of individuals, who went to Europe and developed diverse affiliations and careers upon returning to China, as well as the additional biographies of other prominent leaders in society, culture and the economy of twentieth-century China, that are the basis of the CBD.

### 3. Documentation and Composition of the CBD

The CBD was developed utilizing books, journals, interviews, and archival materials. The archives were particularly important and provided a treasure trove of information (see Table 4). An extensive analysis of these archives is given, along with 72 document translations in *The Guomindang in Europe: A Sourcebook of Documents*, in an article on conducting research in the French archives, along with four shared online databases of archival notes.<sup>10</sup>

- 
- |   |   |
|---|---|
| 1 | Archives Nationales (AN, Paris)   |
| 2 | Archives Nationales, Section d'Outre-Mer (AOM, Aix-en-Provence)   |
| 3 | Archives du Ministère des Affaires Étrangères (AAE, Archives Diplomatiques, Paris)  |
| 4 | Écoles des Hautes Études en Sciences Sociales, Centre de Recherches et de Documentation sur la Chine Contemporaine (EHESS, "Centre Chine," Paris) |
| 5 | Archives de l'Association Universitaire Franco-Chinoise (AAUFC, Lyon)   |
| 6 | Bibliothèque Municipale de Lyon (Lyon)  |
| 7 | Shanghai Guomindang Archives (Yangmingshan, Taipei, Taiwan)   |
| 8 | Chinese Communist Party Archives at Tsinghua University (Beijing)   |
| 9 | Public Record Office (PRO, London)  |
- 

**Tab. 4** List of archives consulted for CBD (N = 9).

- 
- |    |  |
|----|--|
| 10 | <p>ghua University Faculty Research Unit on the History of the Communist Party, <i>Fu Fa Qingong jianxue yundong shiliao</i> 赴法勤工儉學運動史料 [<i>Documents on the Travel to France Work-Study Movement</i>] (Beijing: Beijing chubanshe, 1979); San-ching Chen, <i>Qingong jianxue yundong</i> 勤工儉學運動 [<i>The Diligent-Work Frugal-Study Movement</i>] (Taipei: Zhengzhong shuju, 1981).</p> <p>Levine and Chen, <i>The Guomindang in Europe</i> (see the introduction); Marilyn Levine, "Conducting Research in the French Archives on Chinese Radicalism," <i>Republican China</i> 22, no. 2 (April 1997): 93–102; "Marilyn Levine   History," accessed June 25, 2020, <a href="http://www.cwu.edu/history/marilyn-levine-archival-research">http://www.cwu.edu/history/marilyn-levine-archival-research</a>.</p> |
|----|--|

G.- M. TRẦN QUI PHƯƠNG (陈桂芳) représentant des  
ouvriers chinois de la Hollande.

II.- OUVERTURE de la SEANCE

A.- Siège de la réunion: N° 96, Rue Monge, Paris (58)

B.- Date de la séance: Le 16 Août 1925 de 8 h. à 12 h.  
et de 14 h. à 19 h.

C.- Président: ĐĂNG HY HIÊN (邓希贤); Vice Prési-  
dent: QUÁCH THANH CHÁNH (郭清正); Secrétaires: LIÊU PHO  
THANH (柳圃青) et PHƯƠNG CHI CƯỜNG (方至刚).

D.- Ordre du jour:

- 1°.- Discours d'ouverture
- 2°.- Lecture du testament de Sun Yat Sen par le Pré-  
sident
- 3°.- Lecture des 3 derniers décrets Nos 10, 12, 13,  
du Comité exécutif central
- 4°.- Rapport du Comité Général et critique de ses  
travaux
- 5°.- Rapport des délégués des différents sous-comités  
et critiques de leur action
- 6°.- Invitation faite au partisan TRẦN QUI PHƯƠNG  
(陈桂芳) représentant des ouvriers chinois de Hollande, à  
rendre compte de la situation du Parti en Hollande
- 7°.- Délibération sur le Sous Comité de Lyon
- 8°.- Délibération sur les affaires d'organisation et  
d'émancipation
- 9°.- Délibération sur les affaires de propagande
- 10°.- Délibération sur les frais du parti
- 11°.- Délibération sur le journal " Kuo Min " (國民)
- 12°.- Délibération sur l'affaire de la trahison des  
partisans - affaire soumise par le Comité Général
- 13°.- Election.....

Fig. 1 An example of a captured report of an EGMD meeting in 1925 led by Deng Xiaoping, whose Chinese alias is Deng Xixian with his name listed in Vietnamese. This document was seized by the French from a Chinese who landed in Saigon. This may be one of the earliest records of Deng Xiaoping's leadership. Archives Nationales – Section d'outre-mer, SLOTFOM VIII, 6.

From lists of Catholic aid to Chinese youths in 1921 that includes 264 names, to matriculation in French educational institutions, to captured materials of party minutes, handbooks, and reports, and of course Sûreté surveillance and political reports, these archives provide a unique framework for capturing multiple historical actors who might not otherwise have made it into the historical accounts, as well as the more famous individuals (see Fig. 1 for an example of a captured document).

Although some document listings, such as educational matriculation or factory lists, might give one datum per individual, the idea was to see if there was more information for individuals listed in other sources. Furthermore, standard biographical encyclopedias were also used, resulting in a good range of twentieth-century individuals being added. The first step in organizing the CBD was to understand the functional goal in constructing the appropriate analytical foundation. As articulated in his article about prosopography, Lawrence Stone remarked, “The various types of information about individuals in the universe are then juxtaposed and combined and are examined for significant variables. They are used both for internal correlations and for correlations with other forms of behavior or action.”<sup>11</sup>

The CBD was designed as a relational database. Every single individual entry was supported by discrete sources, which altogether numbered 5,127 and were later expanded by 1,323 web linkages in the updated 2018 revision. Some of the most common attributes of the CBD are: the CCP (N = 580), ECCO (N = 273), educational position on campus (N = 249), EGMD (N = 241), military official (N = 214), Hunan (N = 202), Sichuan (N = 180), Guangdong (N = 179), and editor position (N = 165).

Two organizational issues deserve some discussion: historical name categorization, and the individual questionnaire. One of the most important issues regarding historical database information is the concept of names, as examined in the excellent chapter “Biography, Identity, and Names: Understanding the Pursuit of the Individual in Prosopography” by K. S. B. Keats-Rohan, who stated the foundational role that categorizing and cross-referencing individual names serves in database creation: “A name-bearer will be named, that is, have had a personal name given to him, be known by a name in a formal social situation, be identified or described by some form of name in an official document, and be ‘called’ by a name by his intimates...Names once bestowed have a power of their own...Personal names are inextricably bound up with a sense of identity.”<sup>12</sup> Iden-

11 Lawrence Stone, “Prosopography,” *Daedalus* 100, no. 1 (1971): 46, <https://www.jstor.org/stable/20023990>.

12 K. S. B. Keats-Rohan, *Prosopography Approaches and Applications: A Handbook* (Occasional Publications UPR, 2007), 154–58.

1	Original Name	10	Western Name
2	Alternate Name	11	Western Transliteration Name
3	Courtesy Name	12	Dharma Name
4	Childhood Name	13	Tonsure Name
5	School Name	14	Ordained Name
6	Alias Name	15	Precept Name
7	Pen Name	16	Secular Name
8	Literary Name	17	Posthumous Name
9	Chamber Name		

**Tab. 5** CBD Alternative names – Types of names.

tity can be influenced by the type of name category for alternative names for an individual. In the CBD there are 17 types of names, ranging from pen names to childhood names to Buddhist religious names (see Table 5). Name(s) plus birth date is the most reliable way to search the CBD.

Keats-Rohan also makes the salient point that although a biographical database is analytical, the project also needs a biographical questionnaire for each individual that provides a field, in order to give descriptive comments that capture a short, but holistic view of the individual.<sup>13</sup> A report form is separate from a group analysis as it gives a synopsis of the individual. When the CBD was online from 1998 through 2006, there were dozens of queries and report forms (see Table 6), forms for new entries, and suggested attributes.

The most common usages of the CBD were queries on alternative names and a report form that presented a synopsis of the individual and sources (see Table 6). The CBD was discussed on H-NET, at conferences and through personal correspondence by scholars around the world. On the H-Net discussion network the introduction of the database served as a basis for several discussions, such as truth and the nature of historical evidence for sources.<sup>14</sup> Correspondence to the author ranged from the nature of multilingual input to queries about what had become of Robert Hartwell's biographical database. Participation in the Electronic Cultural Atlas Initiative, the Pacific Neighborhood Consortium, and the

13 Keats-Rohan, 146–51.

14 Marilyn Levine, "A Query on Evidence and Historical Fact," H-ASIA, April 21, 1998, <https://lists.h-net.org/cgi-bin/logbrowse.pl?trx>.

**CBD Report Form****Wang Jingqi** 王景岐 1895–1925 Male Origin: Shengzhou, Zhejiang**Education:** Zhejiang University**Political Affiliations:** EGMD, GMD, Left faction**Political Activities and Career:** Party Branch Head, Propaganda, Labor Agitation, Editor Position, EGMD Executive Committee**Documents & Sources:** AN, SLOTFOM, Sûreté, *Found Generation* appendix, 國民報紙, Yangmingshan *Guomindang* archives: D5032 D5038a D7485 D6515 D7613 D7605a D7817a D7636 D7800 D8704 D7817b D8703 D8705 D8708 D6646a D6520 D7604 D6648**Comments:** Chairman and founder of the European Branch of the Guomindang. Other spelling of name: 王景岐 Expelled in the Lyons Incident, Wang fervently desired revolution, and was appointed by Dr. Sun Yatsen to form the European Branch of the Guomindang, which he did in cooperation with Zhou Enlai. Wang was a leftist and supported several right-wing expulsions. Expelled from France in 1925 for a second time, he died aboard ship and was buried at sea. Wang Jingqi is in many archives, including many of his communications with GMD headquarters (Yangmingshan), and in the AN there are confiscated correspondence, translated materials, and police reports. Party affiliation per 1929 Captured EGMD Report – SLOTFOM VIII, 6.**Tab. 6** CBD online report form, a synopsis with basic information, comments, and sources.

National Initiative for a Networked Cultural Heritage were some of the opportunities to share the CBD model.

#### 4. Using CBD data subsets for statistical, geospatial, and network analysis

Several data subsets of the CBD have now been created and analyzed, including: Soviet returned leaders (SRL, N = 133); a SRL and a three province comparison data subset (Hunan, Sichuan, Zhejiang, N = 588); the ECCO (N = 190); and a group of Chinese Students in Europe (CSE, N = 200), who matriculated at four universities in France, Germany, and Belgium. These data network analyses mirror the historical record, as well as revealing unknown features of well-known individuals within the network and those individuals that have not been recognized but who warrant significant new exploration, as suggested in the fundamental research objectives.<sup>15</sup> The development of these data subsets allows for a

15 These findings have been delivered in several presentations during the past two years. For example, Marilyn Levine, “Revolutionary Roads: An Integrative Analysis Utilizing a Chinese Biographical Database” (Workshop, ERC ENP-China Project, “Elites, Knowledge, and Power in Modern China,” Aix-en-Provence, France, October 7, 2019), <https://enepchina.hypotheses.org/>.

demonstration of the utilization of the CBD, due to the smaller scale. This also includes the geospatial information that has been used to analyze regional origins of birth cities and provinces.

Two brief examples, one a statistical and one a network analysis, will demonstrate the possibilities of the utilization of the CBD for analysis. The CSE data subset of Chinese students shows that the region appears to have played a role in which universities were attended by the New Culture youth (see Table 7). The presentation of statistical evidence allows one to ascertain that in terms of this CSE subset, Guangdong is dominant at the Sino-French Institute in Lyon (SFI)

Province	Charleroi	SFI-Lyon	Univ Berlin	Univ Paris	Total
Guangdong	17	29	0	1	47
Hunan	8	17	3	2	30
Sichuan	17	6	2	2	26
Jiangsu	4	8	5	2	19
Hebei	3	7	1	2	13
Henan	4	3	1	1	9
Zhejiang	3	2	1	3	9
Anhui	0	4	1	1	6
Fujian	2	3	0	0	5
Guangxi	1	2	0	1	4
Hainan	0	3	0	1	4
Jiangxi	0	2	1	1	4
Guizhou	2	1	0	0	3
Hubei	1	1	0	0	2
Liaoning	0	0	0	1	1
Shaanxi	0	0	1	0	1
Shandong	0	1	0	0	1
<b>Total</b>	<b>62</b>	<b>89</b>	<b>16</b>	<b>18</b>	<b>184</b>

**Tab. 7** Provincial origin and educational attendance (N = 184) for the CSE group. There are thirteen missing province origins and six individuals who attended other institutions (not shown).\*

\* This table originally was in the presentation: Marilyn Levine and Heidi Yu Huang, "Chinese Political and Cultural Elites: Twentieth Century Transformations" (Association for Asian Studies, Boston, April 2000).



and also has a large presence at Charleroi. Meanwhile, Hunan had more than half its members represented at the SFI, in contrast to Sichuan, which was well represented in the University of Charleroi group. Jiangsu has almost a third of the students at the University of Berlin. A multivariate dendrogram also supports that there were important regional clusters in these institutions (data not shown).

Another example focuses on network analysis based on the more political ECCO data subset, which is presented in Figure 2. The graph is pruned to those with more than six ties ( $N = 82$ ). The ECCO network has four Louvain subgroups (colors) and displays attributes such as degree centrality (no. of external ties) are shown by symbol size, while circles and orange rims represent provincial origins (circle = Sichuan, rims = Hunan).

Figure 2 is from the ECCO data subset of individuals who had a common history while they were in Europe, so it is not surprising that there are many potential ties. The graph shows subgroup 4 ( $N = 17$ , red) on the left of the graph, which includes the most powerful ECCO individuals such as Zhou Enlai, Deng Xiaoping, Nie Rongzhen, and Zhu De. When looking at degree centrality, shown by the symbol size, it is clear that subgroup 4 has the most external ties. The other group with many external ties, and which composes the largest set of individuals, is subgroup 2 ( $N = 40$ , yellow). What is exciting in terms of the results is that it contains individuals who were historically important in the formation and activities of the ECCO. Subgroup 2 includes prominent early martyrs such as the first General Secretary of the ECCO, Zhao Shiyan (趙世炎 1901–1927), powerful early party members who dropped out or were expelled from the CCP, as well as long-term stalwart leaders such as wife-husband Cai Chang and Li Fuchun. This allows one to consider the various tiers of CCP leadership in terms of historical affiliations and individual relations. These types of analyses may bring into focus the key questions of who the important people were and how they interacted with others in relation to ECCO and later CCP activities.

Subgroup 1 ( $N = 19$ , green) includes largely peripheral individuals, most of whom participated in ECCO activities while they were attending educational institutions in Europe, but did not later become prominent in the CCP. They made it to the greater than six ties level, but they do not have central places or high centrality scores. Likewise, subgroup 3 ( $N = 5$ , blue) has four out of five members who are academics. This group has higher centralities than subgroup 1, and historically had more in-depth early period political activities and long-term relationships with ECCO individuals.

Figure 2 also displays the importance of regional affiliations. Interestingly, in the two most powerful subgroups there is a large presence of Hunan and Sichuan individuals. In the 17 individuals in subgroup 4, there are 6 individuals from Sichuan and 4 from Hunan, while Hunan dominates the center of the second most powerful subgroup 2, which numbers 40 individuals. Subgroup 2 is also domi-



nated by Hunan individuals, while the Sichuan individuals are located in more peripheral areas and are of smaller degree centrality, with the two important exceptions of Zhao Shiyuan and Liu Bojian (劉伯堅 1895–1935). Neither subgroups 1 nor 3 has many individuals from Hunan or Sichuan.<sup>16</sup>

In the future development of the CBD project, a time series analysis (start and end dates for the attributes), a more refined categorization of the attributes, textual analysis, and additional geospatial information will be added.<sup>17</sup> Ultimately, it is planned to pursue the excellent advice of Claire Lemerrier, who advocated three issues for the attention of historians: documentation of the interactions, relational patterns, and the temporality of ties.<sup>18</sup> In 2020, the author joined the Elites, Networks and Power in Modern China Project (ENP) at Aix-Marseilles as a Research Associate, and will be examining ways to develop and share these plans with other interested scholars.

In conclusion, with this very brief introduction, one can see many possibilities for quantitative approaches to biographical information. Due to its scope and scale, the CBD is a research tool that was adapted to model the idea that large-scale studies can be developed to understand potential individual and group patterns of biography. It is not just the role of the individuals, or those who knew each other, but also what the affiliations were based upon, and how individuals interacted within the larger scale web of connections. This is hard to visualize in a standard narrative, which relies on source materials that are usually limited to one or two individuals (or groups), but can be conveyed with the converted I-mode data in a network analysis. Both in current and future developments, the goal is to expand the historical narrative to better understand these individual and group patterns, an area that will hopefully provide fruitful opportunities for China scholars.

---

16 In a regression analysis that had been done on the entire data subset, it was shown that Hunan was an outlier with 39 individuals, which is much higher than their relative population. Sichuan, because of its high population, was still a province of leverage with 55 individuals (data not shown).

17 A good template for best practices is suggested by the work of Victor Shih on the CCP Central Committee. See Victor Shih, Shan Wei, and Mingxing Liu, “The Central Committee, Past and Present: A Method of Quantifying Elite Biographies,” in *Contemporary Chinese Politics New Sources, Methods, and Field Strategies*, by Allen Carlson et al. (Cambridge: Cambridge University Press, 2010), 51–68.

18 Claire Lemerrier, “Formal Network Methods in History: Why and How?,” in *Social Networks, Political Institutions, and Rural Societies*, ed. Georg Fertig, vol. II (Turnhout: Brepols Publishers, 2015), 281–310, <https://doi.org/10.1484/M.RURHE-EB.4.00198>.

## 5. References

- Bailey, Paul. "The Chinese Work-Study Movement in France." *The China Quarterly*, no. 115 (1988): 441–61. <http://www.jstor.org/stable/654865>.
- Bonacich, P. "Factoring and Weighting Approaches to Status Scores and Clique Identification." *Journal of Mathematical Sociology* 92 (1972): 1170–82.
- Borgatti, Stephen P. "Two-Mode Concepts in Social Network Analysis." *Encyclopedia of Complexity and System Science* 6 (2009): 8279–91.
- Borgatti, Stephen P., and Daniel S. Halgin. "Analyzing Affiliation Networks." In *The SAGE Handbook of Social Network Analysis*, by John Scott and Peter Carrington, 417–33. London, UK: SAGE Publications Ltd, 2014. <https://doi.org/10.4135/9781446294413.n28>.
- Chen, San-ching 陳三井. *Qingong jianxue yundong* 勤工儉學運動 [*The Diligent-Work Frugal-Study Movement*]. Taipei: Zhengzhong shuju, 1981.
- Davis, Allison, Burleigh B. Gardner, Mary R. Gardner, and Lloyd, Warner W. *Deep South: A Social Anthropological Study of Caste and Class*. Chicago: University of Chicago Press, 1941.
- Freeman, Linton. "Finding Social Groups: A Meta-Analysis of the Southern Women Data." In *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, by Ronald L. Breiger, Kathleen M. Carley, and Philippa Pattison, 1–39. Washington, D.C.: National Academies Press, 2003.
- Homans, G. C. *The Human Group*. New York: Harcourt, Brace and Company, 1950.
- Keats-Rohan, K. S. B. *Prosopography Approaches and Applications: A Handbook*. Occasional Publications UPR, 2007.
- Lemercier, Claire. "Formal Network Methods in History: Why and How?" In *Social Networks, Political Institutions, and Rural Societies*, edited by Georg Fertig, 11: 281–310. Turnhout: Brepols Publishers, 2015. <https://doi.org/10.1484/M.RURHE-EB.4.00198>.
- Levine, Marilyn. "A Query on Evidence and Historical Fact." H-ASIA, April 21, 1998. <https://lists.h-net.org/cgi-bin/logbrowse.pl?trx>.
- "Conducting Research in the French Archives on Chinese Radicalism." *Republican China* 22, no. 2 (April 1997): 93–102.
- "Modern China and the Chinese Biographical Database." Presented at the First International Workshop on Biographical Databases for China's History, Harvard University, 2010.
- "Revolutionary Roads: An Integrative Analysis Utilizing a Chinese Biographical Database." Workshop presentation for the ERC ENP-China Project "Elites, Knowledge, and Power in Modern China," Aix-en-Provence, France, October 7, 2019. <https://enepchina.hypotheses.org/>.
- "The Chinese Biographical Database Project: A Model for Collaborative Scholarship." In *Proceedings of the EBTI, SEER, ECAI, and PNC Meeting*, 514–20. Taipei: Pacific Neighborhood Consortium, 1999.

- “The Diligent-Work, Frugal-Study Movement and the New Culture Movement.” *Republican China* 12, no. 1 (1987): 67–88.
- Levine, Marilyn A. *The Found Generation: Chinese Communists in Europe during the Twenties*. Seattle: University of Washington Press, 1993.
- Levine, Marilyn A., and San-ching Chen. “Communist-Leftist Control of the European Branch of the Guomindang, 1923–1927.” *Modern China* 22, no. 1 (1996): 62–92. <http://www.jstor.org/stable/189290>.
- *The Guomindang in Europe: A Sourcebook of Documents*. Berkeley, Calif.: Institute of East Asian Studies, University of California, Berkeley, Center for Chinese Studies, 2000.
- Levine, Marilyn, and Heidi Yu Huang. “Chinese Political and Cultural Elites: Twentieth Century Transformations.” Presented at the Association for Asian Studies, Boston, April 2000.
- “Marilyn Levine | History.” Accessed June 25, 2020. <http://www.cwu.edu/history/marilyn-levine-archival-research>.
- Shih, Victor, Shan Wei, and Mingxing Liu. “The Central Committee, Past and Present: A Method of Quantifying Elite Biographies.” In *Contemporary Chinese Politics New Sources, Methods, and Field Strategies*, by Allen Carlson, Mary E. Gallagher, Kenneth Lieberthal, and Melanie Manion, 51–68. Cambridge: Cambridge University Press, 2010.
- Stone, Lawrence. “Prosopography.” *Daedalus* 100, no. 1 (1971): 46–79. <https://www.jstor.org/stable/20023990>.
- “Total Number of Websites – Internet Live Stats.” Accessed August 17, 2020. <https://www.internetlivestats.com/total-number-of-websites/#trend>.
- Tsinghua University Faculty Research Unit on the History of the Communist Party. *Fu Fa Qingong jianxue yundong shiliao* 赴法勤工儉學運動史料 [Documents on the Travel to France Work-Study Movement]. Beijing: Beijing chubanshe, 1979.

MATTHIAS ARNOLD/HENRIKE RUDOLPH

# Network Data in the Early Chinese Periodicals Online Database (ECPO)

*Journal of Historical Network Research* 5 (2021) 288–302

**Abstract** This paper introduces the Early Chinese Periodicals Online database (ECPO), based at the University of Heidelberg. It offers a short overview of the development of this research-driven database as it grew to incorporate additional projects. Today, the ECPO hosts a variety of textual and visual sources from the Republican period (1912–1949), as well as biographical data of twentieth-century academic and political elites. We aim to show how this database can be used and extended to pursue historical network research, even for scholars previously unfamiliar with digital infrastructures. The concluding part discusses privacy concerns in data curation and management, which are increasingly relevant for studies of twentieth-century actors and limit the potential to release and reuse these datasets, especially if they pertain to the People's Republic of China's political history.

## 1. Introduction\*

The late Qing and Republican era (mid-nineteenth to mid-twentieth century) is considered one of the most productive periods of Chinese print culture. Books, journals, newspapers, leaflets, pamphlets, and posters printed in the urban centers circulated across the country in vast numbers. With millions of pages, the sheer quantity of Republican sources is beyond what researchers can read and digest during a lifetime. Therefore, historians were only able to subject a fraction of these publications to close reading. Digitization efforts in the People's Republic of China, the Republic of China (Taiwan), and beyond have created databases that enable historians to re-engage with these sources and pursue qualitative and quantitative inquiries that are unfathomable to previous generations of scholars. This article introduces the Early Chinese Periodicals Online (ECPO) infrastructure and its data subsets, focusing on its potential use for historical network analysis.

## 2. From Image Scans to Full Text

In the past two decades, even though Chinese full-text databases grew at an impressive speed, most digitization projects only produced image scans, even in the People's Republic of China. Specialist Greenapple Changsha's endeavor to double-key the *Shun Pao* (*Shenbao* 申報, 1872–1949) database to produce fully searchable texts and complete author-title indexes is an exception in this regard. This picture started to gradually change when the “Global Press Archive” from East View Information Services, Inc. and the “Chinese Newspaper and Periodical Index” (*Quanguo baokan suoyin* 全國報刊索引) from Shanghai Library recently began to offer individual newspapers in full text.

The “Early Chinese Periodicals Online” (ECPO)<sup>1</sup> database is no exception. The ECPO has two major precursor projects. The first introduced Chinese entertainment newspapers (*xiaobao* 小報) as an important primary source for research on late Qing and Republican period cultural and social history in an online database. Inspired by Western entertainment tabloids, *xiaobao* became one of the most popular media for the daily publication of entertainment news, serialized fiction, short stories, social and political commentary, and cartoon drawings in China. The “Chinese Entertainment Newspapers” database<sup>2</sup> was designed to “present

---

\* **Corresponding author:** Matthias Arnold, [arnold@hcts.uni-heidelberg.de](mailto:arnold@hcts.uni-heidelberg.de)  
 1 Early Chinese Periodicals Online (ECPO), <https://uni-heidelberg.de/ecpo>.  
 2 *Chinese Entertainment Newspapers Database* (Heidelberg University, 2006), <http://xiaobao.uni-hd.de>.

information about the particular profile of a given entertainment paper, and to highlight the kinds of information that might be found there.”<sup>3</sup>

The second predecessor for the ECPO grew out of an international network of scholars in the project “A New Approach to the Popular Press in China: Gender and Cultural Production, 1904–1937,” which resulted in the database “Chinese Women’s Magazines in the Late Qing and Early Republican Period” (WoMag).<sup>4</sup> The digitization of four of the most important Chinese women’s journals from the Republican period allowed the researchers to trace common tropes, reconstruct discourses, and compare editorial styles across publications, as well as relating them to other sources and the broader historical context. The project established a web-based repository to provide access to the image scans in a readable size and compiled a detailed author-title index enhanced with analytic metadata and cross-references.

Both approaches, the “extensive” analysis of publications in general for the Xiaobao database, and the “intensive” recording of all individual items for the WoMag database, became the basis of the ECPO project where they are combined and elevated to a larger scale. As shown in Fig. 1, the WoMag database with all scans and metadata is still available online, as is the Xiaobao database with its selected contents. The full scans of the Xiaobao periodicals have recently become accessible in the ECPO.

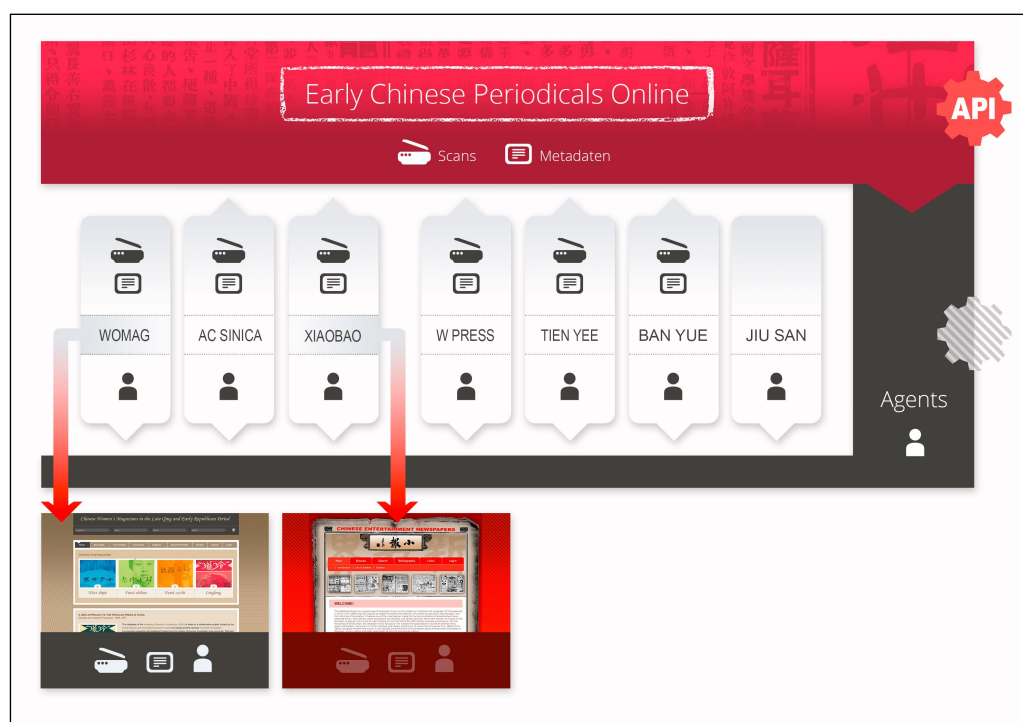
As a result of the editing workflow, the project recorded about 50,000 names of “agents,” which not only contain the names of authors or photographers, but also of those that are mentioned in the texts or depicted in images. Originally a mere by-product, these names are now organized in a dedicated database that provides an “agent service” to all ECPO modules. We use the generic term “agent” to refer to individuals, groups (such as photo studios or sports teams), institutions, or corporations that are tied to an “item,” meaning any text (“article”), image, or advertisement within the journals. We are currently preparing a user frontend for the agent service to allow easy browsing of the agents’ data and provide an interface for basic export functionalities (API).

---

3 Catherine V. Yeh, “Entertainment Newspapers and the Rise of Mass Culture,” *Chinese Entertainment Newspapers*, 2006, <http://projects.zo.uni-heidelberg.de/xiaobao/index.php?p=intro>.

4 *Chinese Women’s Magazines in the Late Qing and Early Republican Period (WoMag)* (Heidelberg University, 2009–2020), <https://uni-heidelberg.de/womag>. See also Doris Song, Liying Sun, and Matthias Arnold, “The Birth of a Database of Historical Periodicals: Chinese Women’s Magazines in the Late Qing and Early Republican Period,” *Tulsa Studies in Women’s Literature* 33, no. 2 (2014): 227–37; Michel Hockx, Joan Judge, and Barbara Mittler, eds., *Women and the Periodical Press in China’s Long Twentieth Century: A Space of Their Own?* (Cambridge: Cambridge University Press, 2018), <https://doi.org/10.1017/9781108304085>.





**Fig. 1** ECPO – data structure with agent service and sub-projects.

This significant database development was part of a collaborative project with the Institute of Modern Chinese History at the Academia Sinica, for which we received generous funding from the Chiang Ching-kuo Foundation for International Scholarly Exchange. The cooperation led to a substantial expansion of ECPO's content and metadata. Since then, data from several individual projects has been added to the platform. For example, the research outcome of a project on the anarchist magazine *Tien Yee* (*Tianyi* 天義, 1907–08),<sup>5</sup> a complete set of the literary magazine *The Half Moon Journal* (*Banyue* 半月, 1921–25), or the collection of “Western Publications Printed in China” by the late Professor Rudolf G. Wagner, including *The Canton Register* (1827–43) and *The Canton Press* (1835–44). The database is an open-access resource with a user-friendly frontend that has been online since 2016.

Today, the ECPO comprises entertainment journals, women's magazines, periodicals from the Academia Sinica collection in Taipei, and a collection of West-

5 Matthias Arnold and Lena Hessel, “Transforming Data Silos into Knowledge: Early Chinese Periodicals Online (ECPO),” in *E-Science-Tage 2019: Data to Knowledge*, ed. Vincent Heuveline, Fabian Gebhart, and Nina Mohammadianbisheh (Heidelberg: heiBOOKS, 2020), 106 and note 28.

ern periodicals printed in China. As of September 2020, it hosts 308 publications comprising over 300,000 image scans. For 134 of these publications we manually compiled additional content information about items, including articles, images, and advertisements.

### 3. The Agent Service

In order to trace appearances across journals, the mere recording of names as they appear in the sources proved to be insufficient, for the following reasons: first, a single Chinese author might have published under a variety of names, ranging from his or her personal name (without surname), well-known pen names, or even pseudonyms intended to mask their identity. Second, Republican Chinese publications embraced cosmopolitanism. They covered Hollywood gossip, reprinted the speeches of Japanese intellectuals, or reported on social events in the foreign concessions of Shanghai. The journals thus frequently feature Chinese transcriptions of foreign individuals, institutions, or place names.

Starting with Western names, i.e., Europeans and Americans, which Chinese sources transcribed in Chinese characters, the project has begun to assign the different names to the records of their respective individuals or institutional bodies. Since the transcription of Western names was not standardized in the Republican era, certain individuals may occur with several variant names in Chinese, not to mention the various misspellings when the Latin script was used. This is why, for example, the actress Constance Bennett has 26 different Chinese name variants recorded from resources in ECPO, most of which were hitherto unknown. From the user's perspective, whatever name variant is used as a search term, every mention of the agent that this name variant is assigned to will show up in the search results.

The agent service allows us to distinguish between names (name records) and agents (agent records) and provides functionality to assign different names and occurrences to their respective agents. ECPO records all names as they appear in the publications, including pen names, studio names, or possible misspellings. We developed a module to merge names, occurrences, and additional metadata into one agent across the databases. It is also possible to separate names or extract individual names from one agent to assign them to another. In some cases, the disambiguation of individuals is very tricky, sometimes impossible. We collect these names in name records. For example, we assigned the 340 occurrences of items signed with "editor" (*bianzhe* 编者) to one single name record. We did the same with acronyms, such as Ping 平, or "Y.D." Identifying the agents "behind" names may require quite an amount of academic research, but we are steadily making progress.

As of today, the project assigned 54,932 name records to 48,124 agents. Lists still require manual editing and cleaning before they can be used for substantial historical analysis. Yet, recording occurrences and co-occurrences of individuals and other entities in texts, pictures, and advertisements allows us to investigate previously neglected historical actors and their networks beyond the well-researched elites.<sup>6</sup>

#### 4. Relational Data in ECPO

The most basic unit of relational data in the women's magazine dataset was the connections between various agents and texts or pictures. Agents could be either actively connected to texts or photographs as authors or creators, or passively by being mentioned in an article or appearing in an image, coded, for example, as "author of" or "mentioned in." The same logic of either active or passive appearance applies to institutions, companies, or organizations. This distinction allows us to not only create two-modal datasets, but also to make meaningful conversions into one-mode datasets to analyze co-authorship and co-occurrence networks. Person-to-person datasets can map, for example, who wrote about whom, who appeared in the same picture or was mentioned in the same set of articles.

As an example, we can examine the two-mode dataset of Song Qingling, a leading women's activist and wife of Sun Yat-sen, the founding father of the Chinese Republic. Apart from the articles she authored, Song Qingling currently appears in the publications collected in ECPO in eight instances: Four articles, one image, and three advertisements. The latter advertised the newest books of the Life Publishing House (*Shenghuo shudian* 生活書店). Thus, the books of Song Qingling were promoted alongside those, for example, of the president Chiang Kai-shek (Jiang Jieshi). As this example shows, the distinction between "mentioned in advertisement" and "depicted in image" points to very different analytical categories. A co-occurrence in a publisher's advertisement alongside other prominent figures speaks to a good reputation, but does not reveal anything about personal ties between the individuals. A personal encounter is, however, a prerequisite for being in the same photograph. The clear distinction between these categories within the dataset thus allows for a nuanced analysis of network structures.

#### 5. The Expansion of ECPO: Beyond Republican Journals

With its potential to give structure to complex historical datasets, the agent service has been further enhanced since 2018. Now, the database also allows the mapping of connections between persons, persons and events, as well as between

---

6 Arnold and Hessel, "Transforming Data Silos into Knowledge."

persons and institutions or organizations without a direct reference to a digitized source. As each individual, each event, and each institution/organization is treated like an independent “agent,” the data structure still allows us to assign a variety of names to a single agent and to merge data entries across different datasets.

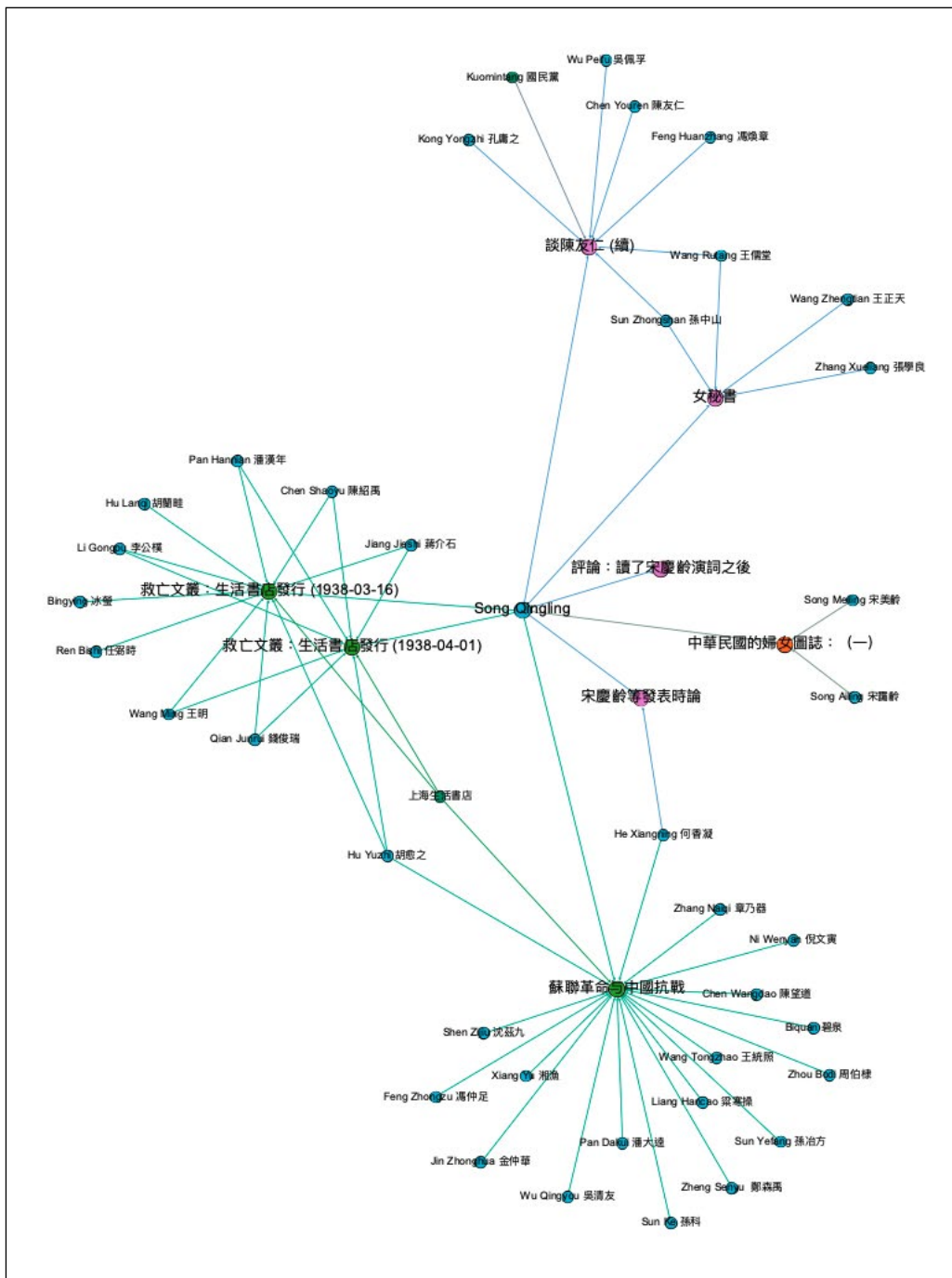
This move was driven by the need to incorporate research data of a project exploring the development of the networks of the Jiusan Study Society (*Jiusan xueshe* 九三學社) from the 1940s to the early 2000s. The Jiusan is one of the so-called ‘democratic parties and groups’ with roots in the Republican period, which were incorporated in the new Communist regime after 1949. This project uses previously neglected archival and published party documents as well as secondary sources and explores the channels of communication and control that the Jiusan established between party members and CCP cadres. A network approach allows us to trace the temporal and spatial dimensions of the Jiusan’s development, and offers new insights into its functioning throughout the turbulent years of political campaigns and economic reforms since the 1940s.

Most of the Jiusan members are, to this day, highly qualified medical or technical experts, scientists, or university lecturers. Therefore, each member is acting in different roles and is embedded in various social and political contexts. For example, each Jiusan member is tied to relatives, a place of work, a field of expertise, a local branch of the Jiusan Study Society, and possibly even to a political sub-group within this local branch or as a representative of the Jiusan to a regional or national representative body. We drafted a data structure that can host both, attribute data (e.g., birthplace or gender) and complex relational data. The resulting dataset records one-modal data (person-person) as well as bi-modal data (person-institution/organization) in its temporal and spatial dimensions. Additionally, relationship types such as “teacher at,” “head of,” “member of,” and so forth enable us to design queries that help to explore hierarchies, as well as the social and cultural capital at work within the Jiusan network.<sup>7</sup>

In projected one-mode networks, tie weights can offer additional insights. They can, for example, help to quantify if an appearance in the same picture was a coincidence or if there is strong evidence that two or more people socialized regularly. Simultaneously, the database offers possibilities to investigate social divisions within the elites. Did bankers and movie stars, for example, mingle, or did they form distinct circles? Because we expanded the agent service to store additional information, including the age, the date and place of birth, and further information from a person’s CV, we can address such questions. However, to date,

---

7 Generous support from the Friedrich-Alexander University Erlangen-Nürnberg and the Wikimedia Foundation, the Volkswagen Foundation, and the Stifterverband enabled the process of data coding and the expansion of the agent service.



**Fig. 2** Two-mode ego-network of Song Qingling in ECPO. Persons in blue, articles in purple, advertisements in light green, organizations in dark green, and images in orange. Visualization created in Gephi.

only a small number of agents' entries possess such additional information as it still has to be entered manually.

## 6. The Agent Service Data Structure

Following the structural logic of network data, the agent service is divided into attribute data ("name data" and "base data") and relational data ("ECPO assignments" and "agent relations"). For each agent, we can add an infinite number of name variants, as many as are found in the historical sources. However, for the rest of the information stored in the base data, we have to make decisions, such as in cases where sources give conflicting information on the date and place of birth.<sup>8</sup> Yet, in most cases, the birth dates deviate by only one or two years, and locations only because place names have changed. We also create ontologies that are not necessarily explicit in the sources, such as nationalities or institutional natures. The initial aim of the agent service was to trace agents across publications. In the Jiusan project, however, the sources recede into the background, and the agents' relations and interactions are of particular interest. This not only includes relationships between agents, for example family ties (person-person) or institutional affiliations (person-institution), but also personal encounters and cooperation. We thus created a special category: "events." Events can be a variety of social, professional, or political gatherings, ranging from participation in political councils to academic conferences. Furthermore, the publication of a book or article is also coded as an event. Each event is described with a start and end date, a location, and a host. In the case of an academic event, this could be, for example, the university or research institute hosting a conference. A second host institution can be added with its respective location, in case of a cooperatively organized event. For publications, we added the location of the publishing house.

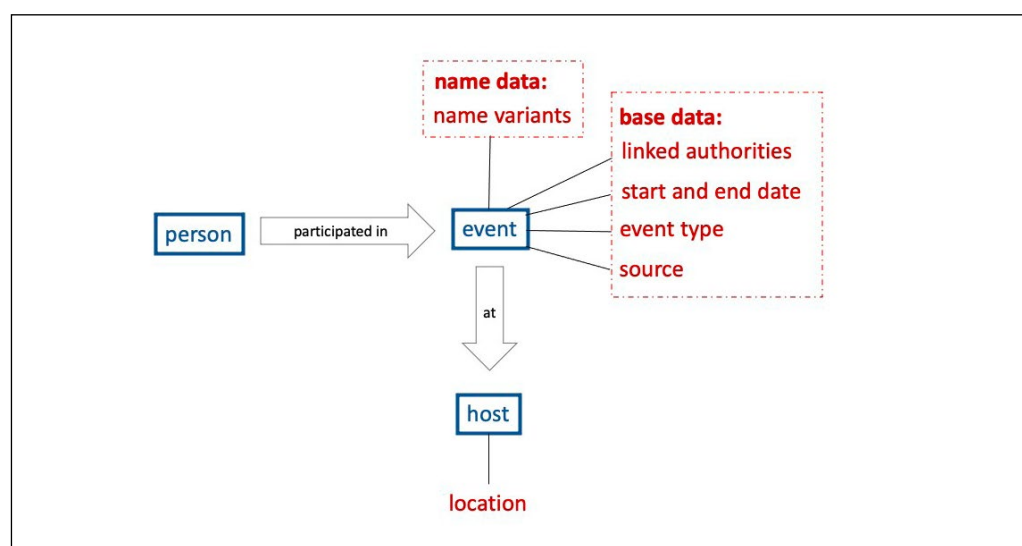
Below you can see how we code information, such as a person's participation in an academic conference, in three consecutive steps (Fig. 3): First, an "agent" of the type "person" is created (if they are not already part of the dataset) with a unique ID and all relevant name and base information (see Tab. 1). In the second step, in a separate table, an event entry is added with a unique ID, and again the relevant name and base data for this particular event. In a third table, the institution or organization, which was hosting the event (or, if the event is a publication, published it) is added.

---

8 Because conflicting source information is not a frequent problem here, we decided against a more work-intensive factoid approach that records all information as provided by different sources. Michele Pasin and John Bradley, "Factoid-Based Prosopography and Computer Ontologies: Towards an Integrated Approach," *Digital Scholarship in the Humanities* 30, no. 1 (April 2015): 86–97.

Attribute data		
Name data		Base data
1	Surname	Type (person, organization, institution)
2	First name	Institution nature (academic, political)
3	Surname in pinyin	Nationality (or national affiliation for organizations or institutions)
4	First name in pinyin	Field of expertise (biology, chemistry, etc. Only for persons)
5	Language (Chinese, English, Japanese, etc.)	Gender (only for persons)
6	Script (Chinese characters, Latin, Cyrillic script, etc.)	Date birth/founded
7	Name type (pen name, given name, other name)	Location birth/founded
8		Date death/dissolved
9		Location death/dissolved
10		Linked authorities ID (GND, VIAF, Wikidata, etc.)
11		Projects (ECPO, WOMAG, etc.)
12		Notes

**Tab. 1** Attribute data in the agent service. Note: For each additional name variant, name data fields 1 to 7 are added.

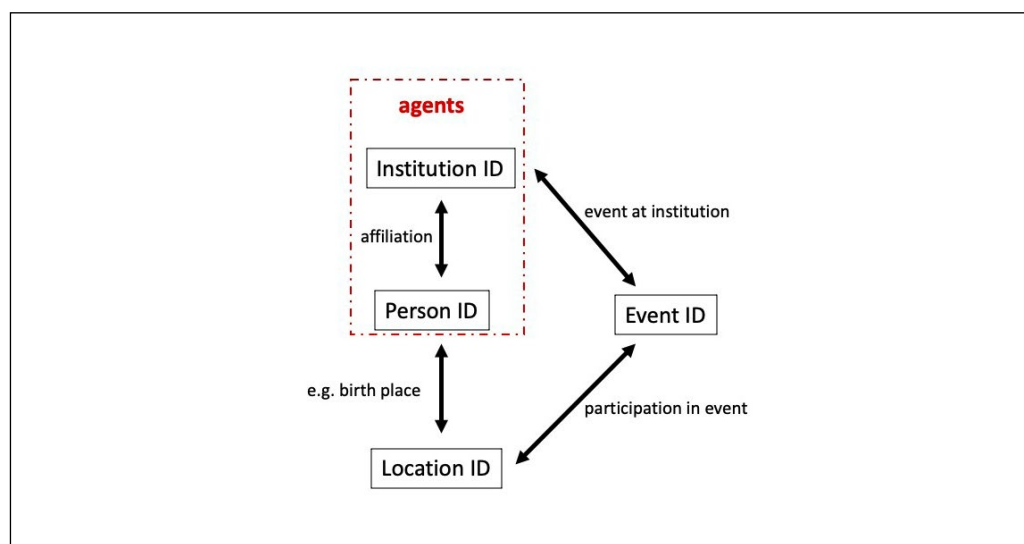


**Fig. 3** Event data structure.

The data structure thus maps both one-modal and bi-modal ties, namely direct links between people as well as links between people via institutions or events as intermediaries. This means, for example, that we would code that “Zhang > is the father of > Li” (direct tie), but *not* “Hu > participated in the same conference as > Gao.” Instead, this would be “Hu > participated in > conference XY” as well as “Gao > participated in > conference XY.” As the database grows, we can add participants to events or assign new events (such as publications) to people.

The one-modal and bi-modal ties have very different explanatory potentials. Direct social connections (one-modal), such as being the child or the Ph.D. student of someone, give us a reasonably good idea of how persons were socially connected to each other. It is much harder to make similar claims, however, merely based on congruent participation in a huge national event. Nonetheless, in analyzing the data, we can still infer networks by making assumptions about probabilities. When two people participated several times in events with a small number of participants or worked at the same institute for an extended period, it is certain that they maintained social relations, regardless of whether their interaction was friendly, strictly professional, or antagonistic. Using the person’s and event’s base data, we can even try to study patterns such as possible interdependencies, like a birthplace or participation in high political bodies.

Apart from the agent service (which includes the persons and institutions or organizations) and the events table, we also created a location table. The China Historical Geographic Information System (CHGIS) laid the groundwork for historical spatial analysis by generating a reference database for place names. The CHGIS tracks changes in place names and administrative units from the third to



**Fig. 4** Links between data tables.



the early twentieth century.<sup>9</sup> Unfortunately, no reference database exists for the period after 1911, although Christian Henriot and his team have already begun working on a Modern China Geospatial Database.<sup>10</sup>

The biographical data on Jiusan members was abstracted from archival documents in the People's Republic of China and contains information on the place of birth, date of party admission, institutional affiliations, past educational and professional experiences, and the place of residence. As many of the Jiusan members were known scientists or medical or technical experts, we can enrich archival data with information abstracted from Chinese secondary sources. To trace the origin of information within the Jiusan dataset, we collect additional primary and secondary sources in a Zotero library, which links to the database entries.

## 7. Privacy Concerns

The archival membership lists of the Jiusan store a variety of personal information of everyday people. The extraction of such information poses questions regarding privacy concerns and the protection of individual data. Biographical information can also be found in published sources for some members, but it may be spread across online biographical sketches, party histories, and local biographical reference works. The aggregation of data from published sources in one database produces new knowledge, and the publication of this knowledge might already affect individuals. This is especially true if we cannot identify a person, for example, if they have a very common name, as we also have no way of knowing if the person is still alive and consents to the publication of their personal data. Taking these sources and further analyzing them through statistical network tools allows us to map connections and draw conclusions that might have real-life implications on a personal or even political level for the individuals involved. Therefore, even though ECPO was initially designed as an open-access database, biographical data from Jiusan members is still not openly retrievable.

The solution could be to anonymize data for publication purposes by replacing the person's name, for example, with a numerical identifier; however, additional information such as institutional affiliations and birth dates would still make it possible to identify the individuals. As research concerning the anonymization or de-anonymization of graph data has shown, even inferring a one-mode network (where, for example, a tie between two anonymized persons expresses a joint affiliation) from the bimodal dataset would not be sufficient. To name only one exemplary work, Backstrom, Dwork, and Kleinberg demonstrated that even

---

9 *China Historical GIS*, 2001–2020, <http://chgis.fas.harvard.edu/>.

10 ENP-Project, "Databases," *Elites, Networks and Power in Modern China (ENP China)*, 2020, <https://www.enpchina.eu/bases-de-donnees/>.

in a fully anonymized graph, as soon as one is able to identify the components of a sub-graph (one component of the network), the anonymization of individuals unravels.<sup>11</sup> This means that even the publication of aggregated data abstracted from published sources might undermine the protection of more confidential information contained in archival sources. Therefore, the highly sensitive nature of the Jiusan dataset poses particular challenges to the integration of data within the rest of the ECPO dataset and the sharing and publication of data. We will nonetheless continue to make at least the biographical base data available, and to merge overlaps in the datasets wherever possible while protecting individual rights and interests.

## 8. Outlook

Despite these concerns, the expansion of the data structure and the agent service's implementation has significantly increased the analytical potential of the ECPO. With the addition of locations, events, and inter-agent relations, the agent service gained several new dimensions and now makes it easier to add agent-only data. Its central position within the ECPO infrastructure facilitates name control across database modules and systematic name management. To ensure our identification of agents is transparent and verifiable, we provide links to international authority records wherever possible. At the moment, we provide references to the German Integrated Authority File (GND),<sup>12</sup> OCLC's Virtual International Authority File (VIAF),<sup>13</sup> Wikidata,<sup>14</sup> DBpedia,<sup>15</sup> and the Chinese online encyclopedia *Baidu Baike* (百度百科).<sup>16</sup> Even though we do not import data from Wikidata or Baidu Baike (as these sites do not adhere to scientific standards), tracking the information on persons, organizations, and institutions stored in these external datasets is nonetheless of value. It enables us to cross-check ECPO entries to assess their accuracy, or to probe which biographical information might be missing from the Baidu Baike entries because they are subject to government censorship. In other words, linking to external datasets opens new paths for ex-

- 
- 11 Lars Backstrom, Cynthia Dwork, and Jon Kleinberg, "Wherefore Art Thou R3579x?: Anonymized Social Networks, Hidden Patterns, and Structural Steganography," in *Proceedings of the 16th International Conference on World Wide Web – WWW '07* (New York: ACM Press, 2007), 181–90.
  - 12 The Integrated Authority File (*Gemeinsame Normdatei*, GND), is the central German authority service hosted by the German National Library, <https://www.dnb.de/EN/gnd>.
  - 13 The Virtual International Authority File combines multiple national and international name authority files in a single OCLC-hosted name authority record service.
  - 14 Wikidata is a collaboratively edited knowledge base hosted by the Wikimedia Foundation, <https://www.wikidata.org>.
  - 15 DBpedia aims to allow users to semantically query structured information from Wikipedia resources, <http://dbpedia.org/>.
  - 16 Baidu Baike is a Chinese-language encyclopaedia by the Chinese search engine Baidu; <http://baike.baidu.com/>.

ploring biases and gaps in popular and official historical narratives and lays the foundation for later exchanges of linked open data.

In the future, references to external authorities can certainly be expanded, especially towards more domain-specific services, for example to the China Biographical Database (CBDB, Harvard), the “Authorship of Chinese Women’s Periodicals” database (*Funü qikan zuozhe yanjiu pingtai* 婦女期刊作者研究平臺) at the Academia Sinica, or the Modern China Biographical Database (MCBD, Lyon).<sup>17</sup> We also plan to develop an API to the agent service to provide machine-readable data for external reuse. We hope that the ECPO will continue to grow in the coming years and that more projects will contribute to the database, whether they are analyzed scans, “just” scans, or agent-related material.<sup>18</sup>

## 9. References

- Arnold, Matthias, and Lena Hessel. “Transforming Data Silos into Knowledge: Early Chinese Periodicals Online (ECPO).” In *E-Science-Tage 2019: Data to Knowledge*, edited by Vincent Heuveline, Fabian Gebhart, and Nina Mohammadianbisheh, 95–109. Heidelberg: heiBOOKS, 2020.
- Backstrom, Lars, Cynthia Dwork, and Jon Kleinberg. “Wherefore Art Thou R3579x?: Anonymized Social Networks, Hidden Patterns, and Structural Steganography.” In *Proceedings of the 16th International Conference on World Wide Web – WWW ’07*, 181–90. New York: ACM Press, 2007.
- China Historical GIS, 2001–2020. <http://chgis.fas.harvard.edu/>.
- Chinese Entertainment Newspapers. Heidelberg University, 2006. <http://xiaobao.uni-hd.de>.
- Chinese Women’s Magazines in the Late Qing and Early Republican Period (*WoMag*). Heidelberg University, 2009–2020. <https://uni-heidelberg.de/womag>.
- ENP-Project. “Databases.” Elites, Networks and Power in Modern China (ENP China), 2020. <https://www.enpchina.eu/bases-de-donnees/>.
- Harvard University. “Introduction.” China Biographical Database Project (CBDB), 2020. <https://projects.iq.harvard.edu/cbdb/home>.
- Hockx, Michel, Joan Judge, and Barbara Mittler, eds. *Women and the Periodical Press in China’s Long Twentieth Century: A Space of Their Own?* Cambridge: Cambridge University Press, 2018. <https://doi.org/10.1017/9781108304085>.

---

17 Harvard University, “Introduction,” China Biographical Database Project (CBDB), 2020, <https://projects.iq.harvard.edu/cbdb/home>; Institute of Modern History, Academia Sinica, “The Introduction of the Database,” Authorship of Chinese Women’s Periodicals, 2018, <http://mhdb.mh.sinica.edu.tw/ACWP/about.php>; ENP-Project, “Databases.”

18 Templates of data structures for an import in ECPO are available upon request.

- Institute of Modern History, Academia Sinica. "The Introduction of the Database." *Authorship of Chinese Women's Periodicals*, 2018. <http://mhdb.mh.sinica.edu.tw/ACWP/about.php>.
- Pasin, Michele, and John Bradley. "Factoid-Based Prosopography and Computer Ontologies: Towards an Integrated Approach." *Digital Scholarship in the Humanities* 30, no. 1 (April 2015): 86–97.
- Song, Doris, Liying Sun, and Matthias Arnold. "The Birth of a Database of Historical Periodicals: Chinese Women's Magazines in the Late Qing and Early Republican Period." *Tulsa Studies in Women's Literature* 33, no. 2 (2014): 227–37.
- Yeh, Catherine V. "Entertainment Newspapers and the Rise of Mass Culture." *Chinese Entertainment Newspapers*, 2006. <http://projects.zo.uni-heidelberg.de/xiaobao/index.php?p=intro>.

BAPTISTE BLOUIN/NORA VAN DEN  
BOSCH/PIERRE MAGISTRY

# Creating Biographical Networks from Chinese and English Wikipedia

*Journal of Historical Network Research 5 (2021) 303–317*

**Keywords** Wikipedia, biography, deep learning, historical network analysis, Wikidata, BERT, NER

**Abstract** With the rise of digital humanities, historians are exploring how to intellectually engage with textual sources given the computational tools available today. The ENP-China project employs Natural Language Processing methods to tap into sources on an unprecedented scale, with the goal of studying the transformation of elites in Modern China (1830–1949).<sup>1</sup> One of the subprojects is extracting various kinds of data from biographies, for which we created a large corpus of biographies automatically collected from the Chinese and English versions of Wikipedia. The dataset contains 228,144 biographical articles from the offline Chinese Wikipedia copy and is supplemented with 110,713 English biographies that are linked to a Chinese page. We also enriched this bilingual corpus with metadata that records every mentioned person, organization, geopolitical entity and location per Wikipedia biography and links the names to their counterpart in the other language. This data structure allows the researcher to analyze the relationships between biographies via shared contents and compare networks in different language settings. In this paper, we will describe our methodology for building this new dataset. The first step was to use automatic text classification for extracting Chinese biographies. We trained a binary classifier to detect biographies on manually classified examples and used a subset of unseen texts to assess its accuracy. The second step used Named Entity Recognition to gener-

---

1 <https://enepchina.hypotheses.org/>

ate metadata and extract relations from the links in Wikipedia. Furthermore, we will delve into the method for building networks from this dataset. We argue that depending on the specific research question, different networks may be built. Using the metadata, researchers can create various kinds of networks to suit their needs. As well as releasing this dataset as an enriched bilingual corpus, we will provide an online interface to query and explore it. Our interface benefits from a bipartite graph structure (which can be seen as a network of documents and entities) and applies the same exploration and clustering strategy found in Cillex.<sup>2</sup>

---

2 <https://www.istex.fr/cillex/>

## 1. Introduction\*

The project “Elites, Networks and Power in Modern China” (ENP – China) aims to study the history of elites in Modern China (1830–1949) using computational techniques, and one of the subprojects is to extract various kinds of data from the elites’ biographies. Biographies are an important source of historical research, as they not only contextualize past lives and embed them within larger narratives and various social networks, but they are also a time capsule for how people’s pasts were evaluated. For this subproject, we turn to one of the largest sources of information on the web, Wikipedia, to create a collection of biographies. We employed automatic text classification to obtain biographies from the Chinese Wikipedia, and supplemented the collection with English Wikipedia biographies from another project. We then added an extensive index that helps us navigate through the corpus.

Although the biographies were collected as part of the subproject, the corpus presented in this article is not limited to articles about Chinese elites. We started by selecting all articles from the Chinese and English Wikipedia that met the criteria of being a biography, and will at a later stage proceed to filter out the relevant texts on Chinese elites from the pool of documents. As a result, other researchers (Chinese- or English-speaking) can use our preliminary corpus to explore the biographies of historical figures from different time periods. With the corpus, they can examine how the lives of specific people are presented in the online public space of today and, by combining the articles with the index, can map and analyze the relationships between biographical texts. Since the documents originate from two different language and cultural communities, one can also use the same index to gain insight into how biographies are linked differently depending on the language/cultural context in which they were created. In other words, Scholars can exploit the structure of this dataset to produce and study various kinds of biographical networks.

The paper is divided into three parts. The first part describes the content, structure, and size of the dataset. The second part focuses on the compilation of the corpus and metadata. It will discuss the extraction and updating of Chinese biographies, the selection of English biographies, and the named entity recognition. The last part briefly presents a use case of the dataset.

---

\* **Acknowledgements:** This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement No 788,476).  
**Corresponding author:** Baptiste Blouin, [baptiste.blouin@univ-amu.fr](mailto:baptiste.blouin@univ-amu.fr)

## 2. The Dataset

### 2.1 Content

The dataset consists of two major components: the English and Chinese language biographies, and the metadata. We started by building the corpus from the Chinese Wikipedia. Although English is the dominant language of Wikipedia (with over six million articles), this does not mean that the Chinese Wikipedia is merely a translated version of the English entries. The two sites are edited independently from each other, which results in a number of Chinese articles not connected to an English page. Therefore, we used Machine Learning techniques to extract Chinese articles that are likely to be a biography and added their English counterparts at a later stage. The collection of English biographies was made in a previous project.<sup>3</sup> However, we selected only those English articles that have a link to the Chinese page in the text. Although the English articles from that project were collected in 2016, we used their page identifiers to obtain the most recent versions of the pages.

Gathering the biographical entries was only the first step in creating the dataset. We also enriched them with additional information that creates interlinks within the whole corpus. Two types of metadata connect the articles. One is the named entities. Named entities refer to all things with a proper noun that are mentioned in the text. In this dataset, we recorded every mention of a person, an organization, a geopolitical entity (GPE) or location in an article. As a result, we have an index of the corpus that allows us to group biographies based on where they intersect in terms of content, and to observe the presence or absence of relationships between the documents' content. The other metadata is inter-language links, i.e., links to an article on the same subject but in another language. We used these links as a bridge between the Chinese and English biographies, as well as to establish connections between the names mentioned, given that a URL is provided in the biographical text for that name. The latter helps track the name in both languages, as well as to differentiate instances with the same name.

The dataset released on Zenodo<sup>4</sup> consists of a very large set of files (see below). To provide a more convenient way to browse it and obtain an overview of what can be found within, we decided to design an online exploration tool.<sup>5</sup> This tool was

---

3 Lebret, Remi, David Grangier, and Michael Auli. "Neural Text Generation from Structured Data with Application to the Biography Domain." In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin Texas, USA, November, 2016. <http://arxiv.org/abs/1603.07771>.

4 <https://zenodo.org/record/4059194>

5 <https://pdg.enpchina.eu/wiki-cillex>



inspired by our previous work on Cillex<sup>6</sup> and combines a full-text search index (Solr) with a network exploration interface (based on Padagraph<sup>7</sup>). It is available as a web application that operates in three stages. The first stage allows the user to run a query against the Solr index to retrieve a set of Wikipedia biographies with associated named entities. The query can be written in English or Chinese, and it is possible to expand the results by also retrieving the corresponding pages in the other language (whenever an inter-wiki link was found). The second stage helps the user to build a table that displays results in a format that is ready for visualization in Padagraph. The user can freely edit the table or proceed directly to the third step: graph exploration. Our interface dynamically creates networks, where the nodes represent Wikipedia pages and entity mentions. The edges between two pages are drawn when there is an inter-language link between the two pages. There are also edges that connect entities to the pages in which they are mentioned. The exploration tool allows the user to specify a query as a starting point for a random walk in the graph (to explore the neighborhood of a node), or to view a “global” graph made of the most central nodes in the result set. When a node is selected, the interface displays its properties, such as a link to the actual wiki page or a picture (if we could find one).

## 2.2 Structure

The data of each article is saved in a folder that carries the Wikipedia page identifier – which can be found in the html content – as its name. In every folder, there is an .xml file and a .csv file, both with the same name. The .xml file contains all the information from the article: the raw text, the Wikipedia ID, the URL, the article title, the identifier of the corresponding page in the other language, as well as the URL of the other page. The .csv file is the location where all named entities are stored. Every row in the file records a name mentioned in the article, the position of that name in the sentence, and the type of mention (i.e., ‘person,’ ‘geopolitical entity,’ ‘location,’ or ‘organization’).

When one applies named entity recognition to documents, there is some degree of information loss, in the sense that it does not distinguish different entities with the same name. For example, all places with the name ‘Paris’ (whether in the USA or France) would be treated as one single instance. However, we reduced this problem by means of entity linking. We found that sometimes, a name of a person, an organization, a GPE or location is followed by a URL link in the Wikipedia text, a link that leads to the article about the subject mentioned. This extra data could help distinguish various entities that share the same name. Another reason to incorporate this information is that this data provides an indirect way to connect entities that have both a Chinese and an English name. This is why some

---

6 <https://www.istex.fr/cillex/>

7 <https://www.padagraph.io/>

Id	entity	type	start_pos	end_pos	link_zh	id_zh	link_en	id_en
2278	周恩來	PER	0	5	None	None	None	None
...	...	...	...	...	...	...	...	...
2278	中國工農紅軍地一方面軍	ORG	69	90	<a href="https://zh.wikipedia/wiki/中國工農紅軍第一方面軍">https://zh.wikipedia/wiki/中國工農紅軍第一方面軍</a>	361551	None	None
2278	中共中央革命軍事委員會	ORG	101	122	<a href="https://zh.wikipedia/wiki/中共中央革命軍事委員會">https://zh.wikipedia/wiki/中共中央革命軍事委員會</a>	9620	<a href="https://en.wikipedia.org/wiki/Central_Military_Commission_(China)">https://en.wikipedia.org/wiki/Central_Military_Commission_(China)</a>	214345
...	...	...	...	...	...	...	...	...
2278	中國	GPE	1	4	None	None	None	None
...	...	...	...	...	...	...	...	...
2278	江蘇維安城內附馬港	LOC	60	77	None	None	None	None

**Fig. 1** Sample of different named entities retrieved from the Chinese biography of Zhou Enlai. This shows the document ID, name of entity, type of entity, the numbers indicating its place in the sentence, the URL and ID of its Wikipedia page and, finally, the link and page identifier of the page in English.

named entities in the file are accompanied by a URL of the article, the Wikipedia ID, as well as the link and identifier of the corresponding page in the other language. However, we did not proceed any further in entity linking. We only relied on the links proposed in Wikipedia pages, so it is possible that in some cases, the same entity points to two different pages or, conversely, that two different entities point to the same page. It is also possible that the same entity is linked in a biography in one language, but not in another.

## 2.3 Size

### *Documents*

We retrieved 228,601 biographical articles from the Chinese Wikipedia dump using a text classifier, and collected 728,321 English articles from a project by Lebre et al.<sup>8</sup> In the end, we only kept the Chinese pages that still existed in the current version of the Chinese Wikipedia and the English pages that had a corresponding page in Chinese. This left a total of 338,857 documents, 228,144 of which are Chinese and 110,713 English. Among them, 110,958 pages from the Chinese corpus have an English inter-language link and 110,713 pages from the English corpus have a Chinese one. The difference in the number of links between the two languages is either due to the fact that several Chinese pages are linked to the same English page, or that some pages are linked to pages that no longer existed on the day the corpus was created. For instance, Lady Guan (關氏 or 關羽女,

8 Lebre, Remi, David Grangier, and Michael Auli. "Neural Text Generation from Structured Data with Application to the Biography Domain." In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin Texas, USA, November, 2016. <http://arxiv.org/abs/1603.07771>.

daughter of Guan Yu) and Guan Yu (關羽) are both linked to a single page in English, that on Guan Yu.

### *Named Entities*

This section presents statistics on the entities named in the corpus. The results are divided into three tables: table 1 for the Chinese corpus; table 2 for the English corpus; and table 3 for the above two corpora combined.

These three tables each contain five columns. The second column gives the sum of occurrences of named entities for each category, and the third column gives the number of distinct annotated named entities. The second part of the table (columns 5 and 4) provides the number of distinct links associated with these entities and their occurrences, respectively. As with columns 2 and 3, column 4 contains all the values, and column 5 only the distinct values. The links are from Wikipedia, each of which connects a named entity to its corresponding page. More information on how we obtained these links can be found in section 3.1.

Type	Count	Distinct count	Link	Distinct Link
Persons	3,600,226	802,035	378,904	104,142
Organizations	1,319,972	407,601	146,179	29,972
GPEs	1,782,456	139,939	27,396	8,313
Locations	224,140	63,773	21,398	10,838
TOTAL	6,926,794	1,413,348	573,877	153,265

**Tab. 1** Chinese NER statistics.

Type	Count	Distinct count	Link	Distinct Link
Persons	4,801,680	1,504,807	331,838	159,727
Organizations	1,896,158	744,388	103,938	48,177
GPEs	1,717,476	260,366	13,347	7,714
Locations	198,509	81,119	53,918	23,244
TOTAL	8,613,823	2,590,680	503,041	238,862

**Tab. 2** English NER statistics.

Type	Count	Distinct count	Link	Distinct Link
Persons	8,401,906	2,306,842	710,742	263,869
Organizations	3,216,130	1,151,989	250,117	78,149
GPEs	3,499,932	400,305	40,743	16,027
Locations	422,649	144,892	75,316	34,082
TOTAL	15,540,617	4,004,028	1,076,918	392,127

**Tab. 3** Total NER statistics.

### 3. Creation Process

In this section, we will discuss the creation process of the dataset, which involves two steps. In the first part, we will dive into the composition of the corpus. We will outline the method for extracting biographies and talk briefly about the selection of English articles. The second part will focus on named entities and language linking.

#### 3.1 Compilation of the Chinese and English Subcorpus

In the first step, we downloaded an offline copy of the Chinese Wikipedia on [dumps.wikimedia.org](https://dumps.wikimedia.org/), in an .xml format.<sup>9</sup> The offline copy, also called a *wiki-dump*, has around two million pages (one-third the size of the English Wikipedia). However, not all pages in the wikidump are articles describing a subject. Some pages are meant to redirect visitors to the relevant page, some are lists of subjects with similar names, and others are lists of subjects within the same category, etc. So, after removing these non-articles with the python tool *WikiExtractor*, we reduced the size of the corpus to 1,046,744 pages.<sup>10</sup>

By inspecting the .xml files, we concluded that there was no metadata that identifies the biographies, and we thus had to rely on the unstructured textual data of the pages. We did some experiments on what method to use for classifying articles into biographies and non-biographies. At first, we tried to select articles by detecting predetermined keywords in the text, such as ‘born in’ ((*chu*)*sheng zai*... (出)生在) combined with ‘family background’ (*chushen* 出身), etc. Such a method is called rule-based classification. We experimented to assess the performance of this method. However, after a few iterations, we found that the sheer number and variety of articles complicated the process of finding the “right” keywords.

9 <https://dumps.wikimedia.org/zhwiki/>

10 <https://github.com/attardi/wikiextractor>

This is why we decided to rely on deep learning for text classification. Text classification is an important problem in natural language processing (NLP). The task is to assign a document to one or more predefined categories, in our case, “biography” or “non-biography.” This has been used in a wide range of applications, such as sentiment analysis,<sup>11</sup> topic categorization,<sup>12</sup> and email filtering,<sup>13</sup> and the methods for this task have changed significantly over the years. Early machine learning approaches for text classification were based on the extraction of bag-of-words features followed by a supervised classifier such as naïve Bayes<sup>14</sup> or a linear Support Vector Machine.<sup>15</sup> Later, better word representations were introduced, such as latent semantic analysis,<sup>16</sup> skipgram,<sup>17</sup> fastText,<sup>18</sup> and today contextualized word embeddings,<sup>19</sup> which improved classification accuracy. For our extraction, we used one of the most widely used contextualized word representations to date, BERT<sup>19</sup>, combined with the neural network’s architecture, BiLSTM. BiLSTM is state of the art for many NLP tasks, including text classification. In our case, we trained a model<sup>20</sup> with examples of Chinese biographies and non-biographies so that it relies on specific semantic features of each type of entry in order to predict its category. Therefore, once trained, given a new entry never seen before by the model, based on the representations of the words of this biography and the weights learned by our neural networks, the model will predict the most probable class to which this entry belongs. Below we will delve deeper into the process of selecting our training data.

- 
- 11 Pang, Bo, and Lillian Lee. “Opinion Mining and Sentiment Analysis,” n.d., 94.
  - 12 Lewis, David D., Yiming Yang, Tony G. Rose, and Fan Li. “RCV1: A New Benchmark Collection for Text Categorization Research,” n.d., 37.
  - 13 Sahami, Mehran, Susan Dumais, David Heckerman, and Eric Horvitz. “A Bayesian Approach to Filtering Junk E-Mail,” n.d., 8.
  - 14 McCallum, Andrew, and Kamal Nigam. “A Comparison of Event Models for Naive Bayes Text Classification,” s. d., 8.
  - 15 Joachims, Thorsten. “Text Categorization with Support Vector Machines: Learning with Many Relevant Features.” In *Machine Learning: ECML-98*, edited by Claire Nédellec and Céline Rouveirol, 1398: 137–42. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998. doi:10.1007/BFb0026683.
  - 16 Deerwester, Scott, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. “Indexing by Latent Semantic Analysis.” *Journal of the American Society for Information Science* 41, no 6 (1990): 391–407. doi:10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASII>3.0.CO;2-9.
  - 17 Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. “Distributed Representations of Words and Phrases and Their Compositionality,” s. d., 9.
  - 18 Joulin, Armand, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. “Bag of Tricks for Efficient Text Classification”. *arXiv:1607.01759 [cs]*, 9 August 2016. <http://arxiv.org/abs/1607.01759>.
  - 19 Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. “BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding”. *ArXiv:1810.04805 [Cs]*, 24 May 2019. <http://arxiv.org/abs/1810.04805>.
  - 20 <https://github.com/flairNLP/flair>

In order to train and test a binary classifier, we needed to have a collection of examples and counterexamples for the model to process. Not only does the number of articles have to be high enough for the algorithm to “understand the differences,” but the types of content need to be varied to minimize potential bias. So, instead of creating a list of articles on our own, we turned to Wikidata to generate it. Wikidata is a central repository that holds data of all kinds of subjects from various sources, including Wikipedia. Each subject is represented as an item, with a unique identifier, a label, and a description, and is further described by triple statements, each consisting of the item identifier, the property, and the value (which is usually the identifier of another item). This way of storing and linking data is highly structured, in the sense that the types of properties are standardized according to community guidelines, and computers are able to infer other statements from triple statements based on a schema that maps relations between properties. For example, ‘A is the daughter of B’ can be interpreted as ‘B is a parent of A.’<sup>21</sup> Such structuring allows for very powerful and specific queries. Using the Wikidata SPARQL endpoint, we obtained lists containing the titles, the Wikidata ID, and the Wikipedia links of articles.

While obtaining a list of potential biographies was unproblematic, it was a challenge to create an effective sample of non-biographies. Technically, every non-person page could serve as a counterexample. But after carrying out an error analysis, we found that a completely random sample could not prepare the model to reject pages of fictional characters, movies, films, manga, bands, etc. As a result, we recomposed the collection of non-biographies, which consists of:

- 2,860 *fake* persons (items categorized in Wikidata as “fictional characters,” “fictional humans,” “literary characters,” “comics characters,” “video game characters,” etc.)
- 3,040 media examples (categorized as “films,” “television series,” “literary works,” etc.)

This collection was further supplemented with a list of 2,984 random examples generated by the Wikipedia API. We removed person pages from the latter by accessing the Wikidata page of every list item using its Wikidata ID and rejecting it based on the presence of the attribute ‘human’ on that page.

To obtain the full texts of example and counterexample articles, we inserted the page titles from the lists on the Wikipedia Special:export page and downloaded the articles in .xml files. After that, we filtered out potential non-articles from the files. The algorithm was given the first three sentences of each (counter) example page and was tasked with distinguishing between the language of a biography and that of a non-biography. To test its performance during the train-

---

21 <https://www.wikidata.org/wiki/Wikidata:Introduction>

ing phase, we used a randomized 10 percent of the training sample. In the end, we did a final test. We presented the algorithm with a set of 415 manually labeled unknown articles. As an outcome, ten articles of the test samples were wrongly classified as biographies, one biography was skipped, and one Chinese article was excluded because it was written entirely in English. The model thus attained an accuracy level of 97.5%.

With the text classifier, we obtained a list of 228,601 articles that were likely to be biographies. We made a short survey of the list, after which the detection of biographies appeared to be successful. However, there were still a few false positives, which were mostly pages containing lists, such as awards. This collection was then supplemented with 4,502 extra Chinese Wikipedia articles. These extra articles are Chinese versions of pages present in the English Wikipedia biography dataset<sup>22</sup>. Some of these English articles contained links to Chinese pages that were missing in the Chinese subcorpus created from our automatic extraction. This is a relatively low number, considering that in all the biographies of the Lebre et al. project that contained a link to a Chinese page, 61,229 were already detected during our extraction.

We ran the extraction on the February 2020 version of the Chinese wikidump (zhwiki-2020-02-01). We used the collected page identifiers to retrieve the html content from the website around June 17, 2020, in order to have the most up-to-date version of the Wikipedia content, since the identifiers are invariant. During this phase, 37 Chinese pages were lost in the process due to the fact that the pages no longer existed in the latest version. After this, we repeated the process of extracting named entities and inter-language links.

To obtain English biographies, we kept only the English articles that were directly linked to a Chinese counterpart. As a result, based on our 228,144 articles in Chinese, we extracted 110,713 English biographies.

### 3.2 Named Entities

Named Entity Recognition (NER) is a typical sequence labeling task in the natural language processing field. The objective of this task is to determine entity boundaries and classify them into predefined categories such as persons, organizations, and location names. Named Entity Recognition forms a core subtask to build knowledge from semi-structured and unstructured text sources. Because we were processing data from Wikipedia, which provides hyperlinks to other

---

22 Lebre, Remi, David Grangier, and Michael Auli. "Neural Text Generation from Structured Data with Application to the Biography Domain." In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin Texas, USA, November, 2016. <http://arxiv.org/abs/1603.07771>.

pages, we broke our entity extraction into two steps. First, by using a pre-trained model, we extracted all the named entities present in a page; then, as a second step, we linked the entities to their Wikipedia pages in Chinese and/or English, if these pages exist. To do so, after extracting an entity we looked in the .html source to check if it had a Wikipedia page. If a Wikipedia page identifier was found, we attached this link to the corresponding entity in the current language and then opened the page to check if there was an inter-language link in the other language. We used two models, one for the Chinese page and one for the English page. Both models were trained on OntoNotes corpora<sup>23</sup> in their own language. For both of the models, we used a bidirectional recurrent neural network with a subsequent conditional random field decoding layer proposed by Flair<sup>24</sup> and trained them on OntoNotes.<sup>25</sup> For the Chinese part, we trained our own model, which we have discussed in a different paper,<sup>26</sup> and for the English part we used the pre-trained model proposed by Flair. Extracting named entities from each biography and linking them to their own Wikipedia pages, whenever possible, establishes links between biographies by way of the entities present in their content. Moreover, since most of the biographies, both in our corpus and that of the entities, are linked with their counterparts in the other language, it is also possible to link, for example, two Chinese biographies by means of the same entities present in their corresponding English pages.

#### 4. Potential Use

As mentioned before, the data is the basis for building a complex network incorporating two types of nodes: biographical texts and named entities. The edges either indicate an instance of a mention between page and named entity, or represent an inter-language correspondence between a biographical text (or named entity) in one language and its counterpart in the other language (although the

- 
- 23 Hovy, Eduard, Mitchell Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. "OntoNotes: The 90% Solution." In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*, 57–60. New York City, USA: Association for Computational Linguistics, 2006. <https://www.aclweb.org/anthology/N06-2015>.
  - 24 Akbik, Alan, Duncan Blythe, and Roland Vollgraf. "Contextual String Embeddings for Sequence Labeling." In *Proceedings of the 27th International Conference on Computational Linguistics*, Santa Fe, New Mexico, USA, August, 2018.
  - 25 Weischedel, Ralph, Martha Palmer, Mitchell Marcus, Eduard Hovy, Sameer Pradhan, Lance Ramshaw, Nianwen Xue, Ann Taylor, Jeff Kaufman, Michelle Franchini, Mohammed El-Bachouti, Robert Belvin, and Ann Houston. "Ontonotes release 5.0." Linguistic Data Consortium, October 2013. <https://catalog.ldc.upenn.edu/LDC2013T19>.
  - 26 Blouin, Baptiste, and Pierre Magistry. "Contextual characters with segmentation representation for named entity recognition in Chinese." In *Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation*, Hanoi, Vietnam (held online), October, 2020.



inter-language link of an entity node depends on whether a URL is provided in the biographical text).

To give a use case example for this network data, one may ask whether different language communities in Wikipedia emphasize different types of personal relationships in biographies. As an example, let us consider a prominent CCP figure such as Zhou Enlai (1898–1976). One tackles the question by drawing comparisons between the egocentric network of Zhou Enlai in English and Chinese. For each language, one may construct the network in which the page node, the biography of Zhou Enlai, is linked to different entity nodes, specifically the named entities of the type ‘person.’ One may then assess the importance of each person’s name using its edge weight (the number of times it occurs in the text). Finally, one may check whether there is an inter-language link for every entity node and use these links to detect whether some persons occur in both biographical networks. This allows one to discover which nodes are mentioned most frequently in English and Chinese, respectively, and which nodes receive special attention in both linguo-cultural communities. Of course, this method may be relatively crude, as not every mention of a person indicates a significant relationship to the subject of the biography. In some ways, one has to add another layer of data to evaluate and categorize the type of relationships in the biographical texts. One can also flip the approach by examining which biographies mention Zhou Enlai, and what relationship Zhou Enlai has to the subjects of those biographies.

Although this is an example for analyzing the links between the biographies and names mentioned, one can also exploit the mapped relations to easily create a subset of biographies and use the subcorpus for other modes of inquiry, such as discourse analysis, or, in the case of ENP-China, for data extraction. One is not confined by the given data structure, but can also repurpose the dataset in accordance with their particular needs and goals.

## 5. Conclusion

We compiled a large pool of biographies from Chinese and English Wikipedia. In order to make this bilingual corpus accessible, we enriched it with an extensive index that lists all mentioned persons, organizations, geopolitical entities and locations per article, and also collected inter-language links between the pages and between the mentioned names (given that the latter is accompanied by a link in the text).

Although the ENP – China project uses this corpus to extract data on Chinese elites in the Republican era, this dataset could be repurposed for building a network, in which biographical texts are indirectly connected to each other via the names mentioned within them. One could make use of the index to study the relationships of historical figures presented in popular digital sources, like Wiki-

pedia, and could go on to compare networks of biographies written in different languages by using the inter-language links.

As mentioned above, such a dataset can be used for various purposes, which is why we decided to make the data available for those interested in analyzing the relationships between online biographies. We hope that this dataset can contribute to the historical network research community and provide scholars with an opportunity to engage with biographical texts in novel ways.

## 6. References

- Akbik, Alan, Duncan Blythe, and Roland Vollgraf. "Contextual String Embeddings for Sequence Labeling." In *Proceedings of the 27th International Conference on Computational Linguistics*, Santa Fe, New Mexico, USA, August, 2018.
- Blouin, Baptiste, and Pierre Magistry. "Contextual characters with segmentation representation for named entity recognition in Chinese." In *Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation*, Hanoi, Vietnam (held online), October, 2020.
- Deerwester, Scott, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. "Indexing by Latent Semantic Analysis." *Journal of the American Society for Information Science* 41, no. 6 (1990): 391–407. doi:10.1002/(SICI)1097-4571(199,009)41:6<391::AID-ASII>3.0.CO;2-9.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding." In *Proceedings of NAACL-HLT 2019*, 4171–4186, Minneapolis, USA, May 24, 2019. <http://arxiv.org/abs/1810.04805>.
- Hovy, Eduard, Mitchell Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. "OntoNotes: The 90% Solution." In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*, 57–60. New York City, USA: Association for Computational Linguistics, 2006. <https://www.aclweb.org/anthology/N06-2015>.
- Joachims, Thorsten. "Text Categorization with Support Vector Machines: Learning with Many Relevant Features." In *Machine Learning: ECML-98*, edited by Claire Nédellec and Céline Rouveirol, 1398: 137–42. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 1998. doi:10.1007/BFb0026683.
- Joulin, Armand, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. "Bag of Tricks for Efficient Text Classification." In *EACL (2)*, edited by Mirella Lapata, Phil Blunsom, and Alexander Koller, 427–31. Association for Computational Linguistics, 2017. <http://dblp.uni-trier.de/db/conf/eacl/eacl2017-2.html#GraveMJB17>.

- Lebret, Rémi, David Grangier, and Michael Auli. "Neural Text Generation from Structured Data with Application to the Biography Domain." In *EMNLP*, edited by Jian Su, Xavier Carreras, and Kevin Duh, 1203–13. The Association for Computational Linguistics, 2016. <http://dblp.uni-trier.de/db/conf/emnlp/emnlp2016.html#LebretGA16>.
- Lewis, David D., Yiming Yang, Tony G. Rose, and Fan Li. "RCV1: A New Benchmark Collection for Text Categorization Research." *J. Mach. Learn. Res.* 5 (2004): 361–397. <http://portal.acm.org/citation.cfm?id=1005332.1005345>.
- McCallum, Andrew, and Kamal Nigam. "A Comparison of Event Models for Naive Bayes Text Classification." In *Learning for Text Categorization: Papers from the 1998 AAAI Workshop*, 41–48. Menlo Park, Calif.: The AAAI Press, 1998. <http://www.kamalnigam.com/papers/multinomial-aaaiws98.pdf>.
- Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. "Distributed Representations of Words and Phrases and Their Compositionality." In *Advances in Neural Information Processing Systems 26*, edited by C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, 3111–3119. NeurIPS Proceedings, 2013. <http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality>.
- Pang, Bo, and Lillian Lee. "Opinion Mining and Sentiment Analysis." *Foundations and Trends in Information Retrieval* 2, no. 1–2 (2008): 1–135. doi: 10.1561/15000000011.
- Sahami, Mehran, Susan Dumais, David Heckerman, and Eric Horvitz. "A Bayesian Approach to Filtering Junk E-Mail." In *Learning for Text Categorization: Papers from the 1998 Workshop*. Madison, Wisconsin: AAAI Technical Report WS-98-05, 1998. [citeseer.ist.psu.edu/sahami98bayesian.html](http://citeseer.ist.psu.edu/sahami98bayesian.html).
- Weischedel, Ralph, Martha Palmer, Mitchell Marcus, Eduard Hovy, Sameer Pradhan, Lance Ramshaw, Nianwen Xue, Ann Taylor, Jeff Kaufman, Michelle Franchini, Mohammed El-Bachouti, Robert Belvin, and Ann Houston. "Ontonotes release 5.0." *Linguistic Data Consortium*, October 2013. <https://catalog.ldc.upenn.edu/LDC2013T19>.