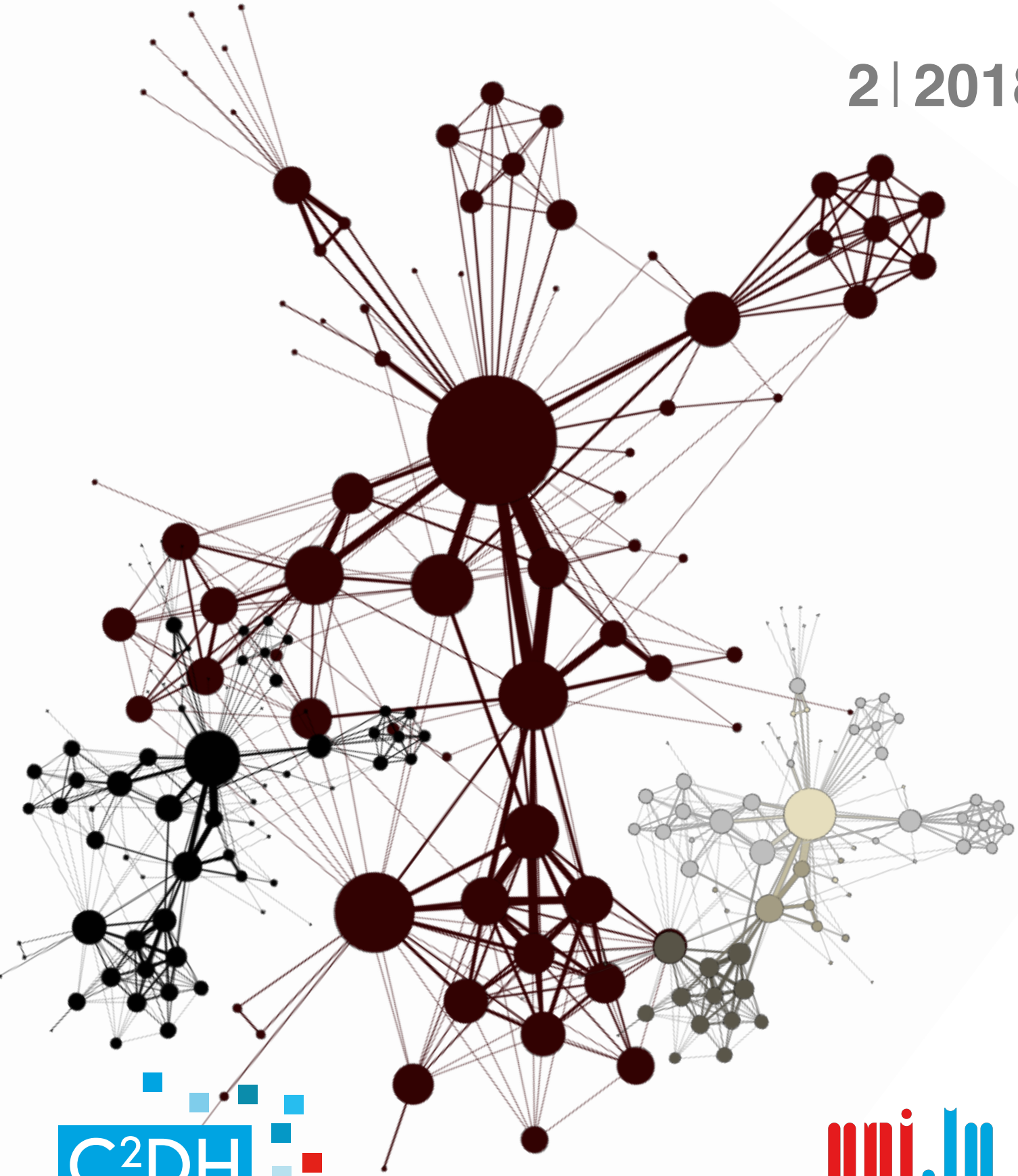
Journal of Historical Network Research

2 | 2018

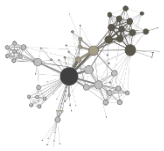


C²DH
LUXEMBOURG CENTRE FOR
CONTEMPORARY AND DIGITAL HISTORY

UNIVERSITÉ DU
LUXEMBOURG

UNIVERSITÉ DU LUXEMBOURG

JHNR
Journal of Historical Network Research



Issue #2 - Table of Contents

This article applies network analysis tools to letters written by and about English Benedictine nuns living in Brussels during the seventeenth century in order to demonstrate the ways in which such an approach expands our picture of early modern religious communities, makes visible the protagonists of religious controversy, and advances debates about enclosure and anonymity. The dataset for this network analysis is taken from the RECIRC project database (the project is entitled "The Reception and Circulation of Early Modern Women's Writing, 1550-1700" http://recirc.nuigalway.ie/). The RECIRC project is producing a large-scale quantitative analysis of the ways in which women's writing was received and circulated in the early modern English-speaking world; its database will be open-access from the project's close. The project has captured reception data on a range of female-authored sources, including texts produced in English convents established in Europe during the sixteenth and seventeenth centuries. The metadata extracted from the nuns' letters that form the basis of this study have generated 1,188 reception records, each tracing a connection between a female author and a receiver. Network analysis is shown here to illuminate debates about the nature and extent of enclosure imposed upon early modern nuns, as well as the sheer breadth and diversity of their epistolary relationships. Furthermore, it exposes otherwise invisible protagonists in religious controversy, and progresses methodological debates about the presentation of data relating to anonymity.

We propose that questions of long-standing interest in the study of *wayang kulit*, Indonesia's centuries-old shadow puppet theatre tradition, can potentially be posed in structural terms and investigated using the tools of network science. Here, we construct weighted character co-occurrence networks based on the Javanese *wayang kulit* incarnation of the Mahabharata epic, supplementing nodes with metadata specifying characters' tribal affiliations and historical origins in either Indian or Javanese traditions. In order to identify characters who play unique structural roles which other approaches may overlook, we generate null model ensembles of artificial networks that share the empirical networks' degree sequences, underlying episodic structures, and node metadata. By ranking nodes by the extent to which their betweenness centrality exceeds a null model's expectations, we reveal characters whose appearances in a story, while not necessarily large in number, tend to serve the specific topological function of bridging groups of other characters. Decomposing betweenness centrality values into their inter-faction components then clarifies how these bridge-like characters are situated among the epic's

various social factions. We observe that female characters, despite being few in number and appearing relatively infrequently, appear to dominate these rankings disproportionately. Analyses involving closeness centrality reveal low-closeness outliers whose appearances, although relatively frequent, keep them structurally isolated and distant from the rest of the Mahabharata universe; these include the epic's antagonists, the Korawa. Characters with historical origins in the Javanese tradition are found to be embedded just as closely within the network as are characters from the Indian canon when their degrees are taken into account using null models.

Family Networks of an Emerging Jewish Intelligentsia (Cracow, 1850-1918)

(Marek Jerzy Minakowski)

53-75

Mass-genealogical research of the Jewish community in Cracow in 19th and early 20th century provides extraordinary opportunity to investigate the process of the emergence of a new social class: society of traditional merchants and peddlers produced modern attorneys and doctors. Now we are able to capture the dynamics of the process. For many reasons, the Jewish community in Cracow is an outstanding specimen. About 1900 there were about 25,000 Jews in Cracow (about 28% of city population). Almost whole family structure of the population has been revealed and analysed as a connected network. Over 1200 nodes of the network have been identified as the Jagiellonian University students between 1850 and 1918. We know what and when they were studying and often in which house they were born, what was their family social status etc. This data is used to model several key features of new emergent social class: what was the impact of parents' families on the choice of university education and the choice influence on a future marriage. Especially interesting are results about influence of family status on the completion of a doctoral qualification which, in turn, provided basis for discussion of best formula to describe how this influence is spreading.

Artist migration through the biographer's lens: A case study based on biographical data retrieved from the Austrian Biographical Dictionary

(Maximilian Kaiser, Matthias Schlögl, Katalin Lejtovicz, Peter Alexander Rumpolt)

76-108

A lexicon like the *Österreichisches Biographisches Lexikon 1815–1950* (Austrian Biographical Dictionary) seems to be as all in one cast because it is built on a set of formal rules for writing articles and some strict but basic criteria for the selection of new entries. The human reader can find information within that resource on a wide range of topics as well as detailed information about the life and career paths of historical individuals. An attempt to systematically analyze this information in a larger scale, however, is condemned to fail without the help of computational methods. In a first stage it is needed to convert unstructured text in structured information. Pieces of information, the so called biographical building blocks, can be identified in two ways: through natural language processing methods and by manual annotations. Both processes which are intertwined and - in the case of the APIS project - done in a custom-built virtual research environment provide the existing biographical data at hand for the analyses following in later stages. This paper aims at describing the data collection process on the example of place names which can be found in artist biographies and at demonstrating possible use cases for historical network research. In this context it is also outlined how this field of research can benefit particularly from biographical data.

Given that the amount of available digital sources is constantly growing, studies on the acquisition and transmission of knowledge in the public sphere tend to be methodologically selective in their analysis of historical discourses. These studies, as a rule, tend to offer a very partial image from an intellectual history perspective of the production, representation, reception, and evaluation of knowledge. A more productive approach would be to analyse the ways in which pre-modern knowledge production functioned based on the broadest possible sample of sources. The project will seek, with a computer-based methodology, to consistently compile and systematically analyse all 356 journals and periodicals published in Halle between 1688 and 1815, as well as to gather information on all the actors participating in the process (authors, editors, printers et cetera). The periodical press of the Enlightenment is particularly relevant for this type of approach. Newspapers and periodicals allowed for a broader circulation of the ideas of the educated and for faster reactions to news, controversies, and publications, as they became a widely accepted form of interpersonal communication. They thus filled a public and institutional gap and provided an open and adaptable medium for various public discourses without thematic limitations. We qualitatively go beyond standard discourse analysis by using a wealth of text and network data digitally available to reconstruct the breadth and depth of all the types of discourses of the Enlightenment, the combined quantitative methods for text and network analysis will allow us to trace the impact of the many involved parties right into the structure of the discourses.

During the last two decades single photographs and photograph corpora have gained in popularity as sources for historical research. In addition to their important function as carriers of the past, photographs also contain valuable information about past social relations. However, to utilise this information a researcher needs a more structured dataset, a photograph corpus containing rich metadata, which allows us to explore and analyse contextual information stored in alphanumeric form. My paper will exemplify how photography corpora could be used as a source for network analysis seeking to explore, reconstruct and visualise hidden historical social networks. The empirical case of my paper revolves around regional and interregional networks of East German dissident movement. The main empirical material explored for network analysis and visualisations consists of a large enriched photograph corpus on East German dissident movement maintained by Robert Havemann Foundation in Berlin. Based on this corpus my paper will explore the structure and dynamics of regional and interregional networks of East German opposition. The results introduce evidence that regional connectedness based on personal mobility among the East German dissidents both changes and increases over time, thus resulting in continuously evolving patterns of social interaction. Further, the analysis of Roland Jahn's geospatial networks evidences the usefulness and power of historical network analysis when it comes to tackling changes in patterns of social interaction.

McShane, Bronagh Ann

Visualising the Reception and Circulation of Early Modern Nuns' Letters

Journal of Historical Network Research 2 (2018) 1-25.

Keywords

Network analysis, network visualisation, early modern, nuns' letters, reception

Abstract

This article applies network analysis tools to letters written by and about English Benedictine nuns living in Brussels during the seventeenth century in order to demonstrate the ways in which such an approach expands our picture of early modern religious communities, makes visible the protagonists of religious controversy, and advances debates about enclosure and anonymity. The dataset for this network analysis is taken from the RECIRC project database (the project is entitled "The Reception and Circulation of Early Modern Women's Writing, 1550-1700" http://recirc.nuigalway.ie/). The RECIRC project is producing a large-scale quantitative analysis of the ways in which women's writing was received and circulated in the early modern English-speaking world; its database will be open-access from the project's close. The project has captured reception data on a range of female-authored sources, including texts produced in English convents established in Europe during the sixteenth and seventeenth centuries. The metadata

extracted from the nuns' letters that form the basis of this study have generated 1,188 reception records, each tracing a connection between a female author and a receiver. Network analysis is shown here to illuminate debates about the nature and extent of enclosure imposed upon early modern nuns, as well as the sheer breadth and diversity of their epistolary relationships. Furthermore, it exposes otherwise invisible protagonists in religious controversy, and progresses methodological debates about the presentation of data relating to anonymity.

1 Introduction*

Scholars such as Ruth Ahnert, Sebastian E. Ahnert, Evan Bourke and Ingeborg van Vugt have pioneered the application of network analysis tools to early modern sources. Ahnert and Ahnert used quantitative network analysis tools to visualise and analyse the Protestant correspondence networks that operated in England during the reign of the Catholic Queen Mary I (1553-58). Based on metadata extracted from 289 letters written either to or by Protestants living in England between 1553 and 1558, their study revealed hitherto overlooked individuals (many of them women) who were fundamental to the operation of the network and, thus, to upholding Protestant resistance in Marian England.[1] Drawing on a corpus of more than 4,000 letters, Bourke employed network analysis tools to assess the importance of female involvement in the Hartlib circle, an intellectual correspondence network formed in London in 1641. His results revealed that women such as Dorothy Moore Dury (*c*.1613-64) and Katherine Jones (1615-91), Viscountess Ranelagh, were integral to the network despite having been

---

* **Acknowledgments**: I am grateful to Marie-Louise Coolahan and Marian Lyons for helpful comments and feedback on earlier drafts of this article. Thanks are also due to David Kelly and Evan Bourke for assistance with the production of the datasets. The research for this article was funded by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013 / ERC Grant Agreement n. 615545).
Author: Bronagh Ann McShane, Moore Institute for Research in the Humanities and Social Sciences, National University of Ireland, Galway, Ireland, e-mail: bronagh.mcshane@nuigalway.ie.

1 Ruth Ahnert and Sebastian E. Ahnert, "Protestant Letter Networks in the Reign of Mary I: A Quantitative Approach," *English Literary History* 82:1 (2015): 1-33. See also: Ruth Ahnert, "Maps Versus Networks," in *News Networks in Early Modern Europe*, eds. Noah Moxham and Joad Raymond (Leiden: Brill, 2016), 130-57. Ahnert and Ahnert's current project, "Tudor Networks of Power, 1509-1603" (http://gtr.rcuk.ac.uk/projects?ref=AH/M004171/1), involves reconstructing correspondence networks based on metadata contained in the *State Papers Online* archive. See Ruth Ahnert and Sebastian E. Ahnert, "Reconstructing Correspondence Networks in the State Papers Archive", unpublished paper delivered at "Reception, Reputation and Circulation in the Early Modern World" conference held at the National University of Ireland, Galway, 22-25 March 2017. Online podcast available at: https://soundcloud.com/mooreinstitute/ruth-ahnert-and-sebastian-ahnert?in=mooreinstitute/sets/reception-reputation-and-circulation.

previously overlooked in most scholarship on the Hartlib circle.[2] Building on the extensive work already conducted on epistolary networks that made up the Republic of Letters, van Vugt argued for the importance of employing a multi-layered network approach in order to represent more accurately the hybrid and complex nature of historical epistolary networks.[3] Her study advocated a "disclose" reading methodology which combines quantitative, distant reading practices with qualitative, close reading of sources. Such a combination is reflected in this network analysis of early modern English nuns' letters.[4]

In tandem with developments in the field of network analysis, the last decade has witnessed a proliferation in scholarship on the English convents in exile. Proscription of formal religious communities in the aftermath of the Protestant Reformation in England resulted in the foundation of 22 enclosed English convents in Europe during the period 1598 to 1700; 21 of these were established as new foundations in locations across France and Spanish Flanders (in what is now modern-day Belgium) while one, the Brigittines of Syon Abbey, a medieval foundation, ultimately settled in Lisbon, Portugal.[5] The majority of these convents remained in operation on the Continent until the late eighteenth century when the turmoil of the French Revolution forced their disbandment. Thanks to the work of recent editorial and prosopographical projects, most notably the pioneering "Who Were the Nuns?" project, led by Caroline Bowden (Queen Mary University of London), an abundance of sources produced by and about members of these English convents has been made available, both online and in print.[6] This has, in turn, expanded considerably our knowledge of the nature of contemplative

---

2 Evan Bourke, "Female Involvement, Membership and Centrality: A Social Network Analysis of the Hartlib Circle," *Literature Compass* 14:4 (2017): 1-17.

3 See "Mapping the Republic of Letters" (http://republicofletters.stanford.edu/), "Circulation of Knowledge/ePistolarium" (http://ckcc.huygens.knaw.nl/), and "Cultures of Knowledge: Networking the Republic of Letters, 1550-1750" (http://www.culturesofknowledge.org/).

4 Ingeborg van Vugt, "Using Multi-Layered Networks to Disclose Books in the Republic of Letters," *Journal of Historical Network Research* 1 (2017): 25-51. This methodological approach is also advocated by Ruth Ahnert, who argues that "network analysis provides valuable navigation for close reading". See "Tudor Networks of Power, 1509-1603, project abstract" (http://gtr.rcuk.ac.uk/projects?ref=AH/M004171/1).

5 Claire Walker's pioneering study of the English convents in exile offers a foundational analysis of the establishment and growth of these communities: Claire Walker, *Gender and Politics in Early Modern Europe: English Convents in France and the Low Countries* (Basingstoke: Palgrave Macmillan, 2003).

6 "Who Were the Nuns?: A Prosopographical Study of the English Convents in Exile 1600-1800" (http://wwtn.history.qmul.ac.uk/, hereafter WWTN). This Arts and Humanities Research Council (UK) funded project took place between 2008 and 2013. The project recovered a wealth of archival documentation on the English convents in exile. The project database holds prosopographical records of over 4,000 women who joined 22 convents established across Europe during the sixteenth and seventeenth centuries. Each individual in the database has been assigned their own unique identifiers (UIDs).

life for women who left England to join convents established in Europe during the sixteenth and seventeenth centuries.[7]

The English Benedictine Convent of the Assumption of Our Blessed Lady – the subject of the present article – was the first English convent to be established on the Continent after the Reformation. It was founded in 1598 by Lady Mary Percy (*c*.1570-1642). She was one of four daughters of Thomas Percy (1528-72), seventh earl of Northumberland, an Elizabethan martyr executed due to his involvement in the 1569 Northern Rising.[8] In 1616, at the age of forty-six, Percy was elected abbess of the Brussels convent, a position she retained until her death in 1642. During her lifetime, Percy gained a not insignificant reputation as an author and, as Jaime Goodrich has highlighted, she was closely involved in a number of projects to translate devotional works into English for use by the English community in exile.[9] In addition to her activities as a translator, Percy was a prolific letter-writer as were her fellow Benedictine nuns in Brussels. Numerous letters penned by her and other members of the Brussels community survive and are currently housed in the Archive of the Archdiocese of Mechelen (hereafter AAM).[10]

2 The dataset

The letters produced by Percy and the English Benedictine nuns provide the basis for this network analysis. They were written over the course of the seventeenth century and were sent to the Flemish Archbishops of Mechelen and their secretaries. The English nuns were writing to the Flemish archbishops because, when the Brussels convent was established in 1598, the new English Benedictine congregation had not yet been restored and therefore the community was placed under the spiritual jurisdiction of the Archbishops of Mechelen. The letters vary in form and content. However, the majority relate to a series of on-going disputes that emerged within the convent during the 1620s

---

7 For a general introduction see, Caroline Bowden and James E. Kelly, eds., *The English Convents in Exile, 1600-1800: Communities, Culture and Identity* (Burlington, VT: Ashgate, 2013).

8 Caroline Bowden, "Percy, Lady Mary (*c*.1570-1642), Abbess of the Convent of the Assumption of Our Blessed Lady, Brussels," in *Oxford Dictionary of National Biography* (hereafter *ODNB*) [https://doi.org/10.1093/ref:odnb/66980, accessed 9 Nov. 2017].

9 For example, in collaboration with the English Jesuit, Anthony Hoskins (d.1615), Percy was responsible for the translation, from a French edition, of Achilles Galhardi's *Breve compendio intorno alla perfezione Cristiana* (*An Abridgement of Christian Perfection*), published in 1612. See Jaime Goodrich, "Translating Lady Mary Percy: Authorship and Authority among the Brussels Benedictines," in *English Convents in Exile*, 109-22.

10 Archief van het Aartsbisdom Mechelen (AAM), Mechelen, Belgium, Regulieren Brussel, Engelse Nonnen, Doos 12/1-12/3, unfoliated. Today this archive is held in the Diocesan Pastoral Centre in Mechelen, a city located between Brussels and Antwerp in northern Belgium.

and 1630s which were to last until the mid-seventeenth century.[11] The disputes were complex and multifaceted but the main issue of contention centred on Abbess Percy's refusal to allow Jesuit confessors minister to the community (a confessor was a priest appointed to hear the nuns' confessions and was responsible for their spiritual welfare).[12] Since its establishment in 1598, the Brussels convent had benefited from a close relationship with a succession of prominent Jesuits, among them Anthony Hoskins, vice-prefect in Flanders (1610-13) and John Norton, procurator (financial manager) of the Jesuit province (1610-23), while the role of confessor to the convent was traditionally held by a member of that order.[13] A combination of personality clashes and power struggles within the Brussels house meant that relations between the abbess and the Jesuits steadily deteriorated during the 1620s. This ultimately led to the emergence of distinct factions within the community: a "pro-Percy" faction (those who sided with the abbess) and an "anti-Percy" faction (those who wanted Jesuit priests, or priests sympathetic to the Jesuit order, to continue in their traditional role as confessors).[14]

Data from 405 letters containing reception evidence and written between 1609 and 1693 has been entered in the RECIRC project database from the AAM. Of these, 359 are original letters while 46 are translations. Translations are instances where a letter written in English by one of the nuns was translated into French or Latin, in order that it could be understood by their Flemish superiors.[15]

---

[11] The Brussels convent disputes have received significant scholarly attention: see Walker, *Gender and Politics in Early Modern Europe*, 138-42; Jaime Goodrich, "Authority, Gender, and Monastic Piety: Controversies at the English Benedictine Convent in Brussels, 1620-1623," *British Catholic History* 33 (2016): 91-114; Emilie K.M. Murphy, "Language and Power in an English Convent in Exile, *c*.1621-*c*.1631," *The Historical Journal* (Forthcoming): 1-25 [; Paul Arblaster, "The Monastery of Our Lady of the Assumption in Brussels (1599-1794)," *English Benedictine History Symposium* 17 (1999): 54-77. For an account of similar disputes that emerged in English Poor Clare convents, see Marie-Louise Coolahan, "Archipelagic Identities in Europe: Irish Nuns in English Convents," in *English Convents in Exile*, 211-28, and Caroline Bowden, "The English Convents in Exile and Questions of National Identity, *c*.1600-1688," in *British and Irish Emigrants and Exiles in Europe*, ed., David Worthington (Leiden; Boston, MA: Brill, 2010), 297-314.

[12] The statutes of the Brussels convent, approved in 1612, granted varying amounts of authority to key office holders, among them the convent confessor (called the ordinary confessor) who was appointed by the archbishop and whose duties included providing spiritual guidance, celebrating Mass, and hearing weekly confessions. The convent also had access to "extraordinary" confessors who were appointed either by the archbishop or the abbess. Their role was to provide additional spiritual counsel and hear confessions; "The Third Parte of Those Matters ... externally appertayning to the Congregation," *Statutes Compyled for the Better Observation of the Most Glorious Father and Patriarch S. Benedict* (Ghent, 1632), 4-6. I am grateful to Jaime Goodrich for sharing a copy of the statutes.

[13] Goodrich, "Authority, Gender, and Monastic Piety", 95.

[14] For an outline of the various factions that emerged within the community during the early 1620s see, Goodrich, "Authority, Gender, and Monastic Piety", 91-114.

[15] For an extended discussion regarding the translation of the Brussels nuns' letters see, Murphy, "Language and Power".

The data fields used to build the dataset for the network analysis that follows are taken from the RECIRC database. These are best illustrated by an example; a letter written on 8 April 1623 by Mary Vavasour (d.1676), a member of the Brussels convent, to the Archbishop of Mechelen, Jacobus Boonen (d.1651).[16] In her letter, Vavasour, who was one of the ringleaders of the "anti-Percy" faction, outlined the divisions that had emerged within the community due to a growing rift between Abbess Percy and the convent confessor, Father Robert Chambers (1571-1628).[17] Vavasour reported: "The cheefest cause of all our inconvenience (as I conceave) is the very great difference between my Lady [Percy] and Father Chambers". Vavasour went on to give a lengthy account of Percy's alleged misconduct and accused her of creating an atmosphere of distrust among the nuns:

> My Lady … adviseth not in matters of government … [she] is so easely disgusted, and taketh the Religious [the nuns] so short when they differ in judgment from her … my Ladyes jelowse inquiry after some, hath caused so great disuinion of minds, and mistrustfull looking in to one another['s] actions.[18]

Vavasour's letter thus constitutes a reception of Abbess Mary Percy. It was first entered as a 'Reception Source Work' (hereafter RSW) in the RECIRC database. An RSW is a work or document in which evidence of the reception of a female author and/or her work is found. For each reception of a female author that occurs within a specific RSW, a separate reception entry is created. Since Vavasour's letter contained only one instance of reception (her account of Mary Percy), one corresponding reception record was created.[19]

This reception record contains the following data fields: "Female Author" (the female author being received or, as here, the female author written about in the letter), in this case Mary Percy; "Receiver" (the person doing the receiving or, in other words, the person referring to the female author), in this case Mary Vavasour. It is important to note that "Receiver" here does not mean letter recipient. Instead "Receiver" refers to the person who is writing about (or receiving/engaging with) the female author. As well as capturing the

---

16 Originally from York, Vavasour entered the Brussels convent in 1611 and was professed in 1616 at the age of 17. In 1652 she was elected abbess of the Brussels house, and retained this office until her death in 1676. Biographical information courtesy of WWTN (UID BB186). Boonen, who hailed from Antwerp, was appointed archbishop of Mechelen in 1621 and remained in that office until his death in 1655. For a brief discussion of Boonen's career see Craig Harline and Eddy Putt, *A Bishop's Tale: Mathias Hovius Among his Flock in Seventeenth-Century Flanders* (New Haven, CT: Yale University Press, 2000), 285, 290.

17 Chambers occupied the position of confessor to the Brussels nuns from 1599: Paul Arblaster, "Chambers, Robert (1571-1628), Roman Catholic Priest," in *ODNB* [https://doi.org/10.1093/ref:odnb/5077, accessed 18 Nov. 2017].

18 "Marie Vavasour [in Brussels] to [Jacobus Boonen], 8 April 1623", AAM, Doos 12/2.

19 While in this case, the RSW contained only one instance of reception, other RSWs might have multiple instances of reception and will, thus, have multiple reception entries linked to them.

relationship between an individual female author and a receiver, each reception record features information about the type of reception that has occurred. In the RECIRC database, reception has been classified according to particular types, ranging from adaptation, dedication and extended commentary through to reference to named author, transcription and translation. In this instance, Mary Vavasour's report about Percy is captured as "reference to named author" and "extended commentary". Finally, the archival reference for the AAM was entered into the reference field, in order to link the letter back to its original repository. The metadata extracted from the 405 letters in the AAM has generated 1,188 reception records. In summary, the two pertinent categories in the dataset for the network analysis that follows are "Female Author" and "Receiver".

There are, of course, limitations to the dataset. Firstly, the data has been filtered through the lens of reception, which means that only letters that contain reception evidence have been gathered in the RECIRC database, in keeping with the aims and scope of the project. Thus, data gathered from this archive is representative of letters that contain reception evidence rather than of the convent correspondence as a whole. Secondly, the Brussels correspondence has its own particular bias. The majority of the letters were sent *by* the nuns *to* the archbishops and their secretaries; thus, information flow is largely one-way. In most cases we do not have the replies that may or may not have been sent by the archbishops and other ecclesiastical figures to the nuns.[20] Moreover, it is likely that scores more letters were sent that do not survive (although this problem of attrition applies to all early modern archives).

3 Creating a reception network

The first step in creating this reception network was to extract all the individual reception records linked to the AAM from the project database and store them in JSON (JavaScript Object Notation) format. This data was then organised into two .CSV files in Excel: a nodes file (the female authors and receivers) and an edges file (the connections between them, i.e. the receptions). The edges were given attributes according to reception type, i.e. reference to named author, extended commentary, translation. These files were then imported to Gephi – an open-source network analysis and visualisation software tool – to analyse the network. Using Gephi, various visualisations of the Brussels Benedictine reception network were generated. In the first visualisation (Figure 1), each person is represented as a node and their connections are edges (the connections

[20] The exceptions are "Jacobus Boonen in Brussels to the Benedictine Monastery of the Glorious Assumption in Brussels, 30 September 1628", AAM, Doos 12/2, and "Jacobus Boonen in Brussels to the Benedictine Monastery of the Glorious Assumption in Brussels, 6 April 1632", AAM, Doos 12/1.

are receptions of female authors).[21] There are in total one hundred nodes (people) and 536 edges (receptions) in this network. Of these one hundred people, 67 are nuns and 33 are non-nuns. The nodes have been partitioned by colour; nuns are represented in purple, non-nuns in orange. The layout is force-directed, meaning that the most connected nodes (those with the greatest number of edges) appear closer to the centre while those with the least connections appear at the periphery.[22]

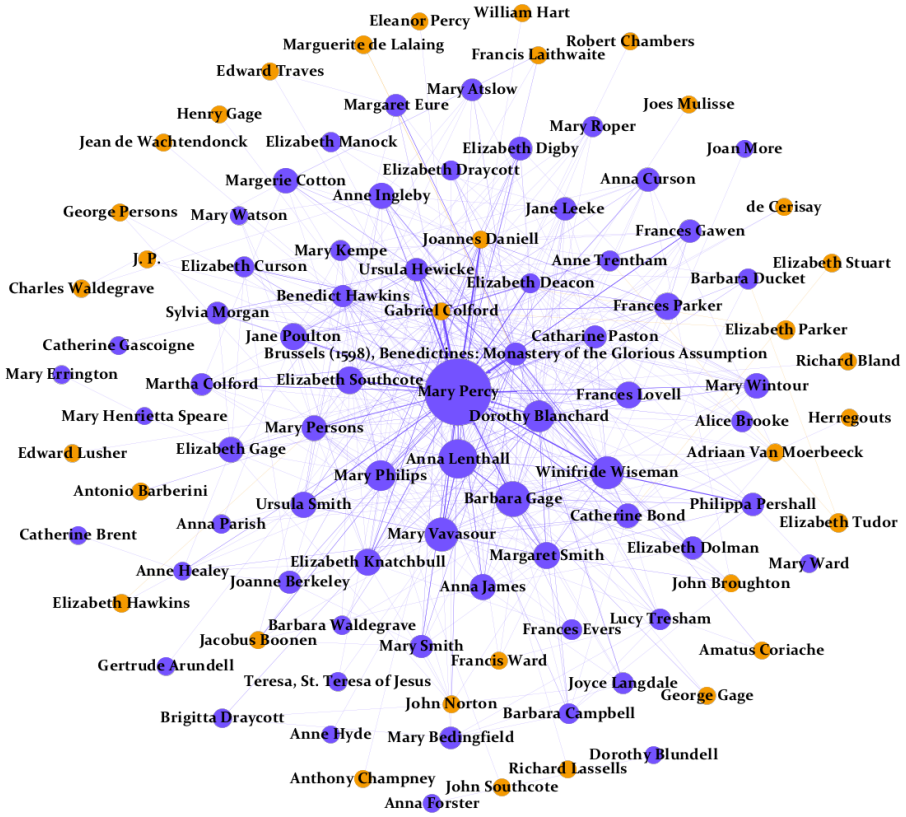


Figure 1: Brussels Benedictines Reception Network (ranked by in-degree).

---

[21] In some respects, a reception network is similar to the co-citation networks discussed in studies by Yves Gingras. See Gingras, "Mapping the Structure of the Intellectual Field Using Citation and Co-Citation Analysis of Correspondences," *History of European Ideas* 36:3 (2010): 330-39.

[22] The layout algorithm used to create the visualisations in this article is Fruchterman Reingold.

At the centre, with multiple edges radiating from her node, is Mary Percy. This is not surprising since, as discussed above, Percy was abbess of the convent between 1616 and 1642 and central to the disputes that emerged within the community during the 1620s and 1630s. Her centrality in the network thus validates the reliability of the analysis. As well as her central position in the network, Percy is also the largest node. This is because the nodes are ranked according to "in-degree", meaning that the node size corresponds to the number of individual people a person is received by, i.e. how many people wrote about that individual or engaged with their writing in some way (e.g. by translating a letter). By ranking the nodes according to this statistical measurement, key agitators in the convent controversies thus become apparent.

When the top five in-degree nodes are calculated (Table 1), the results show that four out of five of the ranked individuals occupied key positions of authority as either abbess and/or prioress (second in command to the abbess) during their religious career in Brussels.

Node	**In-Degree**
Mary Percy	64
Anna Lenthall	28
Barbara Gage	23
Mary Vavasour	21
Winifride Wiseman	20

Table 1: Top five in-degree nodes (reception network).

They are Mary Percy, Anna Lenthall, Mary Vavasour and Winifride Wiseman. Percy's case has already been explained. Her in-degree ranking of 64, more than double that of the next highest ranking node, again confirms what we might expect; she was central to the ongoing controversies in Brussels. Anna Lenthall and Winifride Wiseman, received by 28 and twenty people respectively, both occupied the office of prioress; Lenthall was prioress in 1639 while Wiseman occupied that office on three separate occasions: in 1628, 1630 and again in 1639.[23] Lenthall later went on to succeed Percy as abbess following the latter's death in 1642 while Mary Vavasour, who was received by 21 people, became abbess of the Brussels house ten years later, in 1652.[24] Thus, all four occupied top positions of authority within the convent, which undoubtedly explains why they rank highly in terms of numbers of receivers. The outlier is Barbara Gage who was received by 23 people. Unlike the four nuns discussed above, at no point did she occupy the office of abbess or prioress, although she did hold a lower ranking office, that of portress (the convent doorkeeper) for just one year

[23] WWTN (UID BB199).
[24] WWTN (UIDs BB111 [Lenthall] and BB186 [Vavasour]).

in 1650.[25] Her relatively high in-degree ranking thus prompts further investigation.

A close reading of letters written about Gage reveals her role as a key agitator in the convent disputes and this corroborates her centrality in the reception network. Of the 23 people she is received by, two are priests, three are laymen and 17 are fellow nuns.[26] Of the 17 nuns who wrote about her, more than half portray her as a divisive individual who incited discord within the community. For example, in the early 1620s Gage was identified by Dorothy Blanchard as one who "greatly disturbs the quiet and charity of the house".[27] Later in 1628, Ursula Smith identified Gage as one of two nuns who disrupted "the peace and true union" of the convent (the other nun implicated was Anne Ingleby).[28] In her lengthy epistle to Archbishop Boonen, Smith claimed that Gage and Ingleby were effectively employed by Percy as spies charged with monitoring the behaviour of other nuns (presumably those among the "anti-Percy" faction). They, in turn, used this position to further incite tensions within the community:

> in thees parsones [persons] my Lady douth most confide implying them to see that [the] Rule and descipline be well observed them selfes being noted to be the most des orderly [disorderly] and chefe brekers ther of … what soever the[y] relate all though it be very falce my Lady giveth such credet unto it that with out any excamination she maketh the Religious [the nuns] to be accuse[d] in publik chapter and pennan*c*ed for it.[29]

Not only did Gage's "falce" reports initiate disharmony between Percy and the nuns, her actions also drove a wedge between the abbess and the prioress, Agatha Wiseman. Smith related how Gage, "in [a] most audatious and contemtious mannar called the Prioresse insolent woman … 6 times before her face and … [said] she would ackno[w]le[d]g[e] no superior but my Lady Abbesse". Rather than reproaching her for her overt disregard of Wiseman's authority, Smith reported that Percy took Gage's part:

> [instead] of mentaning the due respect and subjection which the Religious … [ought] to bare to … Superior[s] my Lady taking … [Gage's] parte …

---

25 WWTN (UID BB082).

26 One of these 23 receivers is Gage herself. This is because Gage referred to an earlier letter she had sent to the archbishop in a letter dated 28 May 1638. This instance of self-reception is represented in the graph by the small arc beside Gage's node. These arcs denoting self-reception are also visible on the nodes representing Mary Percy, Ursula Hewicke, Winifride Wiseman and Ursula Smith.

27 "Dorothy Blanchard [in Brussels] to [Jacobus Boonen], before 1624", AAM, Doos 12/3.

28 "Ursula Smith [in Brussels] to [Jacobus Boonen], 3 August [1628]", AAM, Doos 12/1.

29 "Ursula Smith [in Brussels] to [Jacobus Boonen], 3 August [1628]", AAM, Doos 12/1.

> condemning the Prioresse to have doun foolishly which wardes she oftoun itterated.[30]

As well as levelling verbal insults at the prioress and provoking conflict between Percy and other nuns, Gage was accused of theft. On 24 April 1629 Elizabeth Southcote wrote to Boonen, reporting that some of her letters to the archbishop (which she kept in a small box under lock and key) were stolen on the orders of Percy by Gage and three other nuns.[31] The fact that Southcote (who was part of the anti-Percy group) was compelled to keep her letters to the archbishop under lock and key demonstrates the growing atmosphere of distrust that pervaded the convent as discord between factions escalated. While Gage played a central role in fuelling that discord, Percy herself was clearly not beyond reproach.

Gage's case reveals the complexities of the power relations in operation in the Brussels convent. As the above examples illustrate, her role as an agitator of controversy was facilitated and indeed promoted by Abbess Percy, who conferred on her a degree of unofficial status that undermined convent hierarchical and behavioural norms. This meant that her actions were frequently commented on by other nuns who portrayed her as a subversive individual; that her behaviour generated receptions (in our terms). The network analysis thus draws our attention to an otherwise unremarked protagonist, and this expands our understanding of the ways in which the disputes played out. However, since five of Gage's 23 receivers were not, in fact, members of the community, the network analysis also reveals how reports about convent controversies and the women involved in them spread beyond the convent walls. This has major implications for our understanding of the role of enclosure in early modern convents.

4 Network analysis and convent enclosure

During the early modern period, the Catholic Church imposed strict enclosure (or *clausura*) on all female religious.[32] Enclosure entailed the nuns' physical separation from society within the convent cloister and the severing of all worldly and familial ties. Convent statutes emphasised the importance of preserving enclosure, which could be undermined by contact with the outside

---

30 "Ursula Smith [in Brussels] to [Jacobus Boonen], 3 August [1628]", AAM, Doos 12/1.

31 "Elizabeth Southcote [in Brussels] to [Jacobus Boonen], 24 April 1629", AAM, Doos 12/1.

32 Claire Walker, ""Doe not Supose me a Well Mortifyed Nun Dead to the World": Letter-Writing in Early Modern English Convents," in *Early Modern Women's Letter Writing 1450-1700*, ed., James Daybell (Basingstoke: Palgrave, 2001), 159-76; Francesca Medioli, "An Unequal Law: The Enforcement of *Clausura* before and after the Council of Trent," in *Women in Renaissance and Early Modern Europe*, ed., Christine E. Meek (Dublin: Four Courts Press, 2000), 136-52; *eadem*, "The Dimensions of the Cloister: Enclosure, Constraint, and Protection in Seventeenth-Century Italy," in *Time, Space, and Women's Lives in Early Modern Europe*, eds, Anne Jacobson Schutte, Thomas Kuehn and Silvana Seidel Menchi (Kirksville, MO: Truman State University Press, 2001), 165-80.

world. Thus, the statutes of the Brussels Benedictines decreed that contact with families should be restricted because "a Religious person ought to bee very sparing in that kinde, as benig [*sic*] one dead to the world, and that desyreth onely to live to Christ alone".[33] Recent scholarship on early modern convents has pointed to the tensions that existed between the ideal of preserving enclosure and the reality of daily life for women who entered conventual life. Demonstrating the "permeability" of early modern convents, scholars such as Claire Walker and Elizabeth Lehfeldt have highlighted the interactions of cloisters with the worlds beyond their walls.[34] This network analysis contributes to that debate by visually mapping the connections maintained by the Brussels Benedictines with a diverse range of individuals, both in and beyond Spanish Flanders. Furthermore, the network visualisations graphically illustrate the scope of the nuns' networks. While the majority of the orange nodes are figures we might expect to see in the network – for example, bishops and priests charged with various aspects of the convent's governance (Jacobus Boonen, Robert Chambers, Anthony Champney [1569-1644]) – there are also a number of lay people featured, many of whom were family members of the nuns. Individuals such as Elizabeth Parker, Eleanor Percy, Elizabeth Hawkins, George Persons, Henry Gage and John Southcote, wrote letters either to or about their cloistered relations.

Quantitative research reveals the reach of the nuns' connections; qualitative research shows how their lay relatives both in and beyond Spanish Flanders intervened in convent affairs. A case in point is Elizabeth Parker *née* Tresham (1573-1647/8), the mother of Frances Parker who joined the Brussels convent *c.*1622. Elizabeth, a staunch Catholic, was the wife of William Parker (d. 1622), Baron Monteagle, and a sister of the Gunpowder plot conspirator, Francis Tresham (d. 1605).[35] From the Monteagle family residence in Essex, Elizabeth maintained correspondence with her daughter and intervened when she considered the terms of Parker's profession punitive (profession was the act of taking full religious vows). Because Parker suffered from an infirmity, her profession was subject to certain stipulations that denied her the full rights and privileges enjoyed by other choir nuns (choir nuns were those who had taken full religious vows). Among other conditions, Parker could not hold a rank in the house commensurate with her age of profession, nor could she wear the same habit worn by other choir nuns. When Parker's mother became aware of the sanctions imposed against her daughter she was outraged, writing "letters

---

33 "The First Parte of the Statutes of Those Things That Appertayne to common Disciplyne", *Statutes Compyled for the Better Observation*, 38.

34 For a concise overview of such scholarship see Elizabeth A. Lehfeldt, "The Permeable Cloister," in *The Ashgate Research Companion to Women and Gender in Early Modern Europe*, eds., Allyson M. Poska, Jane Couchman, Katherine A. McIver (Aldershot: Ashgate, 2013), 13-31.

35 Mark Nicholls, "Parker, William, Thirteenth Baron Morley and Fifth or First Baron Monteagle (1574/5-1622), Discoverer of the Gunpowder Plot," in *ODNB* [https://doi.org/10.1093/ref:odnb/21345, accessed 18 Jan. 2018].

of discontent" to Frances.[36] Such was the extent of Elizabeth's discontent that she refused to pay her daughter's dowry, ultimately leading to Frances's withdrawal from the Brussels house.[37]

Familial intervention in convent affairs also explains Eleanor Percy's appearance in the network. Eleanor, daughter of Henry Percy (*c*.1532-85), eight earl of Northumberland, was a first cousin of Abbess Percy (their fathers were brothers).[38] When tensions between Percy and nuns opposed to her escalated, Eleanor intervened by writing directly to Archbishop Boonen. In her letter, which was signed and dated London, 30 November 1629, Eleanor lamented what she characterised as the disloyal and "insolent" actions of certain nuns. Emphasising her own "fine birth", she claimed that their "rebellion" against "the spiritual authority" of her cousin undermined the "noble blood" and "antiquity" of the Percy family.[39] Hence the affront was perceived to affect the family's standing far beyond the convent walls or diocesan boundaries. Concern to uphold his family's reputation also prompted the intervention of Henry Gage (1597-1645), a captain in Spain's Flemish army and cousin of Barbara Gage.[40] When Gage became embroiled in yet another dispute, this time with Abbess Percy, Henry intervened as he considered his cousin's treatment by the abbess and other nuns to be unfair. In a letter sent from Oudenburg to Boonen on 11 May 1638, Henry emphasised his family's "very good quality" and "esteem" among the "greatest of England" and recommended his cousin to the protection of the archbishop.[41] As these letters demonstrate, despite their physical distance

---

[36] In her letter to an unknown recipient written *c*.1623, Ursula Hewicke referred to "letters of discontent" which Frances Parker had received from her mother: see "Ursula Hewicke [in Brussels] to [unknown], 1623", AAM, Doos 12/1. Unlike the English Augustinians at Louvain and the Sepulchrines at Liège, who allowed for circumscribed admission of the infirm, the English Benedictines made no formal provision for infirm entrants. See Claire Walker, "Recusants, Daughters and Sisters in Christ: English Nuns and their Communities in the Seventeenth Century," in *Women, Identities and Communities in Early Modern Europe*, eds, Stephanie Tarbin and Susan Broomhall (Aldershot: Ashgate, 2008), 61-76, and Marie-Louise Coolahan, "Nuns' Writing: Translation, Textual Mobility and Transnational Networks," in Patricia Phillippy ed., *A History of Early Modern Women's Writing* (Cambridge: Cambridge University Press, 2018), 274.

[37] For a full account of the incident involving Parker, see Walker, "Recusants, Daughters and Sisters", 71-3.

[38] See Carol Levin, "Percy, Henry, eighth earl of Northumberland (*c*.1532-1585), magnate and conspirator," in *ODNB* [https://doi.org/10.1093/ref:odnb/21938, accessed 29 Jan. 2018].

[39] "*la rebeldia de las Monjas … las dichas religiosas han legado a tal punto de insolencia, que no dexen de menoscabar tanto su noble nascimiento della que acotejen la media sangre de la casa de Noster con la sangre enterra de Northumberland no nos podemos mas refrenar de hazer saber a Vostra Seigneuria Illustrissma que no solamente yo que tambien soy Percy de nombre y nascimiento fino toda la casa de Northumberland la toma muy pesadamente, y no tengo ducla*", "Eleanor Percy in London to [Jacobus Boonen], 30 November 1629", AAM, Doos 12/2.

[40] A.J. Loomie, "Gage, Sir Henry (1597-1645), Royalist Army Officer," in *ODNB* [https://doi.org/10.1093/ref:odnb/10271, accessed 19 Jan. 2018].

[41] "*tres bonne qualité, et fort bien estimé entre les plus grands d'Angleterre*", "Henry Gage in Oudenbourg to [Jacobus Boonen], 11 May 1638", AAM, Doos 12/3.

from cloistered relations who were theoretically "dead to the world", family members – even those living in England – were attuned to events within the convent. Furthermore, they were not averse to intervening when they considered their relatives unjustly treated or their family's reputation to be at stake. The combination of quantitative and qualitative analysis, or "disclose" reading, thus draws our attention to the ways in which early modern enclosure was compromised by the realities of dissent and threats to social status and reputation.

5 Translation as a mode of reception

Whereas ranking the network according to in-degree allows us to identify the key agitators in the convent controversy, ranking by out-degree (Figure 2) immediately brings to light the importance of translation as a mode of reception. In this visualisation the largest nodes are the people engaging with the highest number of female authors, that is, the main receivers.

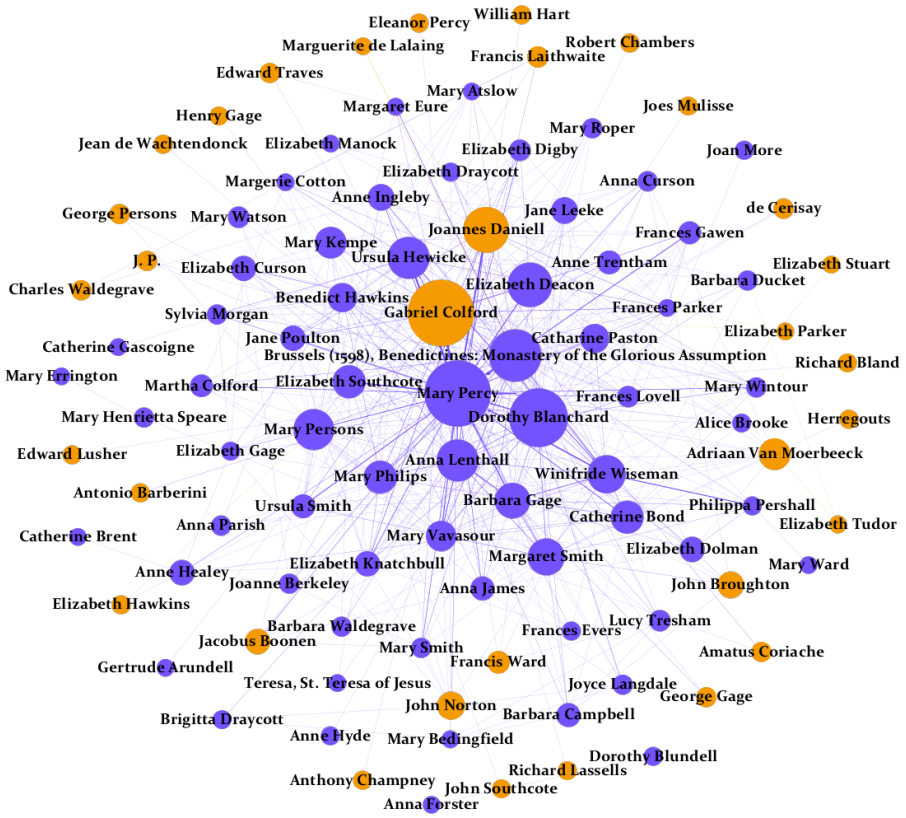


Figure 2: Brussels Benedictines Reception Network (ranked by out-degree).

Two 'non-nun' nodes that did not appear prominently in the previous visualisation now appear near the centre of the network. These are Gabriel Colford (d. 1628) and Joannes (or John) Daniel (*fl*. 1622). Colford was a layman originally from Essex but living in Belgium from around 1604.[42] His daughter, Martha, entered the Brussels convent in 1609 and professed two years later in January 1611.[43] Following his daughter's profession, Colford was employed as the convent's financial manager, a position he maintained from 1611 until his death in 1628.[44] John Daniel was an English secular priest sympathetic to the Jesuit order and confessor to the Brussels nuns during the early 1620s (secular priests did not belong to a religious order). Both men were employed as translators by various members of the Brussels community. They translated letters written in English by the nuns into French and Latin so that the nuns' letters could be understood by their Flemish male superiors. As Emilie Murphy's research has highlighted, the nuns' choice of translators "depended on their factional perspective"; they would employ particular translators depending on whether they belonged to the "pro-Percy" or "anti-Percy" faction.[45] In the case of Colford and Daniel, they were typically commissioned by nuns in the "anti-Percy" group, among them Elizabeth Southcote, Elizabeth Digby and Ursula Hewicke. However, when the network is filtered to identify receptions that are translations only, this immediately reveals how alliances between the nuns and their choice of translators were more complex than such binary divisions suggest.[46]

The filtered translation network (Figure 3) has been ranked according to out-degree, meaning that the largest node (Daniel) is the person *translating* the highest number of individual nuns' letters. The thickness of the edges corresponds to the number of translations. For example, Colford translated five of Ursula Hewicke's letters; hence, the thickness of the edge linking their nodes. Similarly, Daniel translated three of Elizabeth Digby's letters, which again is reflected in the comparatively thicker edge.

---

42 Arblaster, "The Monastery of Our Lady", 66-7.
43 WWTN (UID BB039).
44 Murphy, "Language and Power", 8.
45 Murphy, "Language and Power", 8.
46 Murphy's study has also revealed instances where nuns have acted as translators for other nuns.

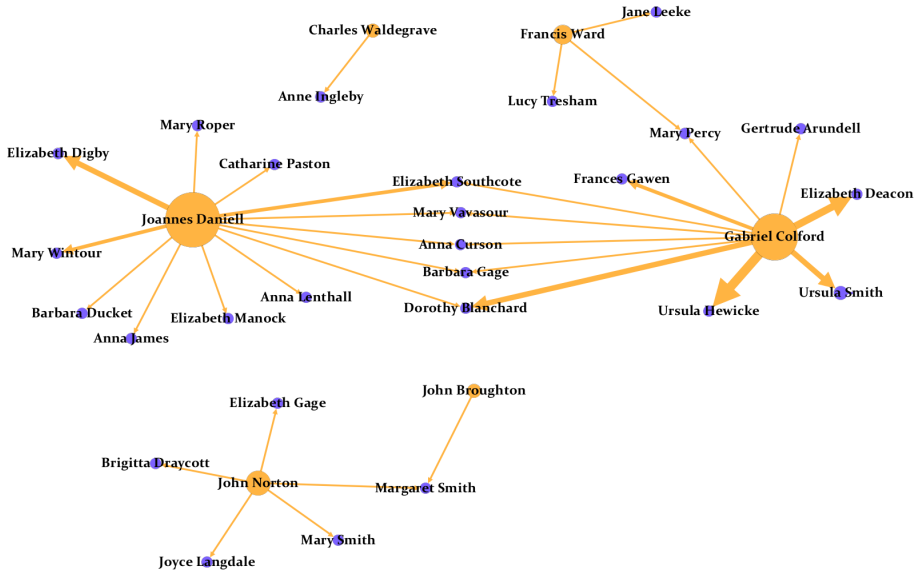


Figure 3: Filtered Translation Network with weighted edges, ranked by out-degree.

These connections reiterate the importance of Colford and Daniel as translators for nuns in the "anti-Percy" cohort since both Hewicke and Digby were identifiably part of that group. Indeed, Digby was one of the ringleaders of the "anti-Percy" group and in 1624, together with two other nuns, left the Brussels house due to ongoing disputes over the role of Jesuit confessors. With the assistance of Jesuit priests, Digby and her companions subsequently established a new Benedictine convent at Ghent, about fifty kilometres north-west of Brussels.[47] More surprising is the connection between Anna Lenthall and John Daniel since Lenthall was firmly part of the "pro-Percy" faction. In her letters to Archbishop Boonen, Lenthall defended Percy's actions and later succeeded her as abbess of the convent following the latter's death in 1642. Nevertheless, she actively solicited the services of Daniel to translate a letter she wrote to the archbishop on 8 April 1623 advertising her decision at the end of her letter: "I chose for [my] interpeter Mr Daniell".[48] This was the only occasion that Lenthall employed Daniel to translate her letters, however. All but two of the thirteen letters she sent to the archbishop after that date were sent directly in French

47 For a discussion of the foundation of the Benedictine convent in Ghent, see Laurence Lux-Sterritt, *English Benedictine Nuns in Exile in the Seventeenth Century: Living Spirituality* (Manchester: Manchester University Press, 2017).

48 "Translation of Anna Lenthall [in Brussels] to [Jacobus Boonen], 8 April 1623", AAM, Doos 12/1.

without the use of a translator, indicating her successful acquisition of that language.

When the translation filter is applied, it further nuances the network by revealing three sub-networks clustering around four external translators. These are John Norton (d.1631) and Charles Waldegrave (1592-1655), both Jesuits; John Broughton (1584-1658), a Benedictine monk; and Francis Ward (d.1665), a secular priest. These sub-networks demonstrate that the Brussels nuns were exposed to numerous potential translators from different religious orders and backgrounds. They further illustrate how the nuns' choice of translators was not always governed by factional affiliations. A case in point is the sub-network centred around John Norton (alias Knatchbull), confessor to the Brussels community during the early 1620s and a prominent Jesuit. During the early 1620s, Norton became closely associated with nuns opposed to Abbess Percy, among them his own sister, Elizabeth Knatchbull (d. 1629), who, together with the abovementioned Elizabeth Digby, left Brussels in 1624 to establish the new Ghent convent under Jesuit direction. In this venture, Knatchbull was aided by her brother who solicited patronage for the fledgling establishment and retained close connections to the rival house after its foundation, acting as the community's spiritual director.[49] Yet despite Norton's strong associations with nuns among the "anti-Percy" faction in Brussels, three of the five women he translated letters for during the early 1620s were identifiably part of the "pro-Percy" cohort: Mary Smith, Margaret Smith and Brigitta Draycott. This demonstrates that a nun's chosen translator was not always an indicator of her stance in the dispute; the need for translation services was of a separate order to factional loyalties.

6 Visualising anonymity

The visualisations discussed above have overlooked one major receiver: anonymous. Although the RECIRC project database has captured data on anonymous reception, it has not assigned unique identifiers (UIDs) to anonymous receivers. As a result, they will not appear in Gephi visualisations that trace a connection between a receiver and a female author using UIDs. A reception is designated anonymous if, for example, a letter by one of the nuns was annotated by a reader who is unidentified, either because there is no signature or the hand-writing is not recognised.[50] In the case of the Brussels correspondence, many of the letters have been underlined, suggesting

---

[49] Elizabeth Knatchbull was elected first abbess of the new Ghent foundation, a position she retained until her death in 1629. A biography of her life was written between 1642 and 1651 by Toby Matthew (1577-1655), her spiritual director, but remained unpublished until the twentieth century. For a discussion of this biography, see Nicky Hallett ed., "The Life of Lucy Knatchbull," in *English Convents in Exile, 1600-1800*, ed. Caroline Bowden, 6 vols. (London: Pickering & Chatto 2012-13), vol. 3, 159-217, and Coolahan, "Nuns' Writing", 267-70.

[50] See for example, "Apolonia Waldegrave [in Brussels] to [Jacobus Boonen], 2 September 1629", AAM, Doos 12/3.

engagement by a reader. But, since it is not clear whether this underlining was carried out by the person to whom the letter was addressed or another reader, "anonymous" has been selected to denote the type of receiver in such cases. Anonymous underlining amounts to 86 of the 1,188 receptions in this corpus. Entirely discounting instances of anonymous reception clearly distorts the analysis. The question of how to capture anonymity, therefore, poses a significant methodological challenge, and one that is not unique to this study.

In the broader context of early modern literary studies, debates about how best to account for anonymity in literary production and circulation have featured in a number of scholarly works. Marcy North has drawn attention to the scholarly neglect of anonymity despite the fact that more than 800 known authors were published anonymously between 1475 and 1640.[51] Thanks largely to North's contribution, in recent years, scholars have begun to analyse early modern anonymity as a literary convention and "textual condition" created not only by authors but also readers (or receivers) of texts.[52] Thus, as North has recently argued, the burgeoning field of anonymity studies has the potential to expand both the history of authorship and the study of early modern reception.[53] This article aims to contribute to this growing field of scholarly inquiry by experimenting with how to capture, quantify and analyse data on the ways in which English nuns' letters were received by anonymous as well as by named receivers.

Visualising anonymity presents a significant methodological challenge which, up to now, has not been adequately addressed in studies that have applied network analysis tools to early modern sources. Where anonymity has been encountered in historical sources, the trend has been to discount it entirely from datasets. For example, in their analysis of Protestant epistolary networks in Marian England, Ahnert and Ahnert excluded correspondence in which the sender or addressee was anonymous, an approach also employed by Bourke in his social network analysis of the Hartlib circle.[54] In her analysis of literary networks of Protestant disputation based on author-data extracted from polemical pamphlets published between 1548 and 1580, Aline J.E. Deicke excluded pamphlets where the author could not be identified or where author

---

51 Marcy L. North, *The Anonymous Renaissance: Cultures of Discretion in Tudor-Stuart England* (Chicago, Ill.; London: University of Chicago Press, 2003), 3. See also *eadem*, "Anonymity," in *Encyclopaedia of English Renaissance Literature*, eds. Garret Sullivan and Alan Stewart, 3 vols. (Oxford: Wiley-Blackwell, 2012), vol. 3, 8-12.

52 See for example, John Mullan, *Anonymity: A Secret History of English Literature* (London: Faber, 2007) and Janet W. Starner and Barbara H. Traister, eds., *Anonymity in Early Modern England: 'What's in a Name?'* (London: Routledge, 2016).

53 Marcy L. North, "Early Modern Anonymity," in *Oxford Handbooks Online*, ed. Colin Burrow (Oxford: Oxford University Press, 2015), 20.

54 Ahnert and Ahnert, "Protestant Letter Networks", 5; Bourke, "Female Involvement, Membership and Centrality", 3.

attributions were considered nebulous (for example, '*vir pius*' [religious man]).[55] Since the aim of these studies was to reconstruct social networks in which social identification markers were central to understanding and analysing the network, the exclusion of anonymity was justifiable. But this study's focus on the reception of female authors and/or engagement with their writing means that account must be taken of receptions involving both named and anonymous receivers.

This network analysis has trialled two approaches to capturing anonymous receptions. The first assigns an individual UID to all instances of anonymous reception, so that "anonymous person" is represented by multiple individual nodes in the network. This had to be done manually at the data processing stage; it involved identifying all anonymous receptions in the core dataset, assigning a different UID to each and adding them to the original nodes and edges .CSV files, before importing the updated files into Gephi. As Figure 4 shows, this approach significantly expands the network. The number of nodes rises by almost 200%, from 100 when anonymous receptions are excluded, to 299, an increase of 199. The number of edges (receptions) also increases, rising from 536 to 735. Incorporating the "anonymous person" nodes thus draws our attention to the sheer quantity of anonymous receptions and establishes the importance of anonymous reception to the circulation and transmission of the nuns' correspondence.

---

[55] Aline J.E. Deicke, "Networks of Conflict: Analyzing the "Culture of Controversy" in Polemical Pamphlets of Intra-Protestant Disputes (1548-1580)," *Journal of Historical Network Research* 1 (2017): 79-81.

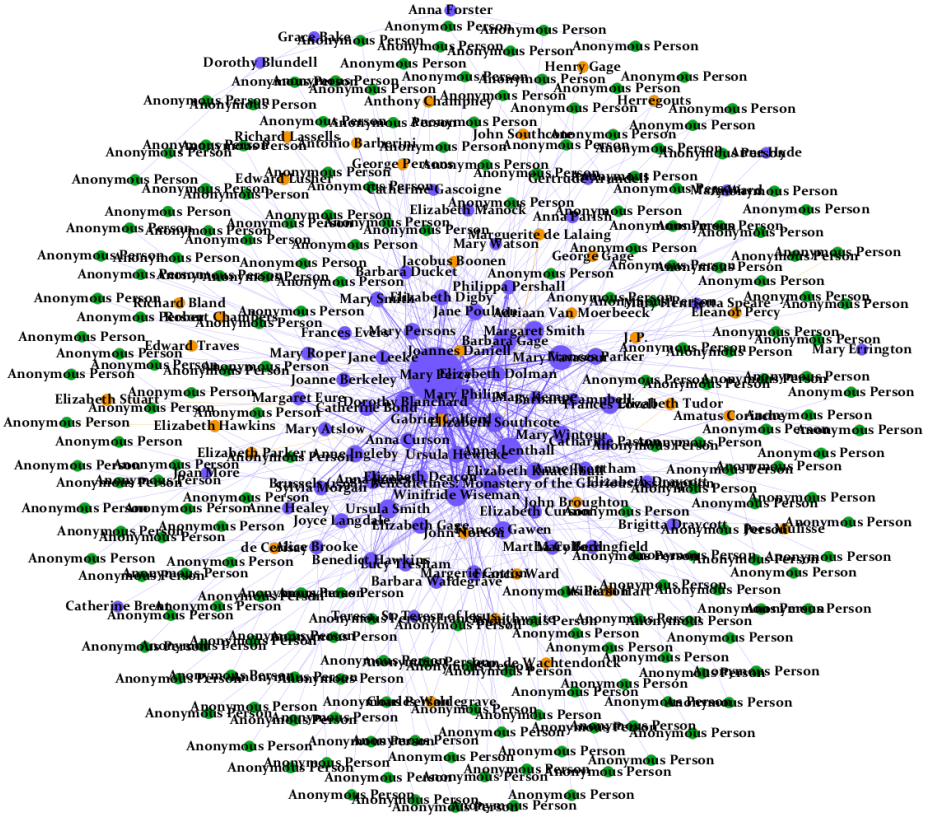


Figure 4: Brussels Benedictines Reception Network. All anonymous receptions are represented as individual "anonymous person" nodes in the network.

Despite the significant increase in the number of nodes and edges, the overall structure of the network is not impacted: Mary Percy remains the most received female author, although her in-degree (the number of people she is received by) rises significantly, from 64 to 101, an increase of 37; Gabriel Colford and Percy retain their position as the highest ranking out-degree nodes (that is, the main receivers).

There are significant drawbacks to this approach, however. The sharp rise in Percy's in-degree is misleading since not all instances of anonymous reception were necessarily carried out by different people. This method therefore skews the results by privileging instances of anonymous reception over named reception. This is because a named receiver will only add a value of one to the in-degree of a female author's node regardless of the quantity of receptions by that receiver; whereas each instance of anonymous reception increases the in-degree value of the female author's node by one. Perhaps most importantly, by incorporating multiple anonymous person nodes, the network becomes less

readable, militating against the value of visualisation as a tool to aid analysis in the first place.

Approach two counteracts this diminished legibility by assigning the same UID to every instance of anonymous reception so that all anonymous receivers are represented by a single "anonymous person" node in the network (as with the case above, this had to be done manually at the data processing stage). As Figure 5 shows, incorporating "anonymous person" as a single node changes the out-degree ranking of the network. The "anonymous person" node is now the largest node, overtaking both Mary Percy and Gabriel Colford as the main receiver; it is connected to 54 different nodes (female authors) compared with Percy and Colford, who are connected to 38 each.

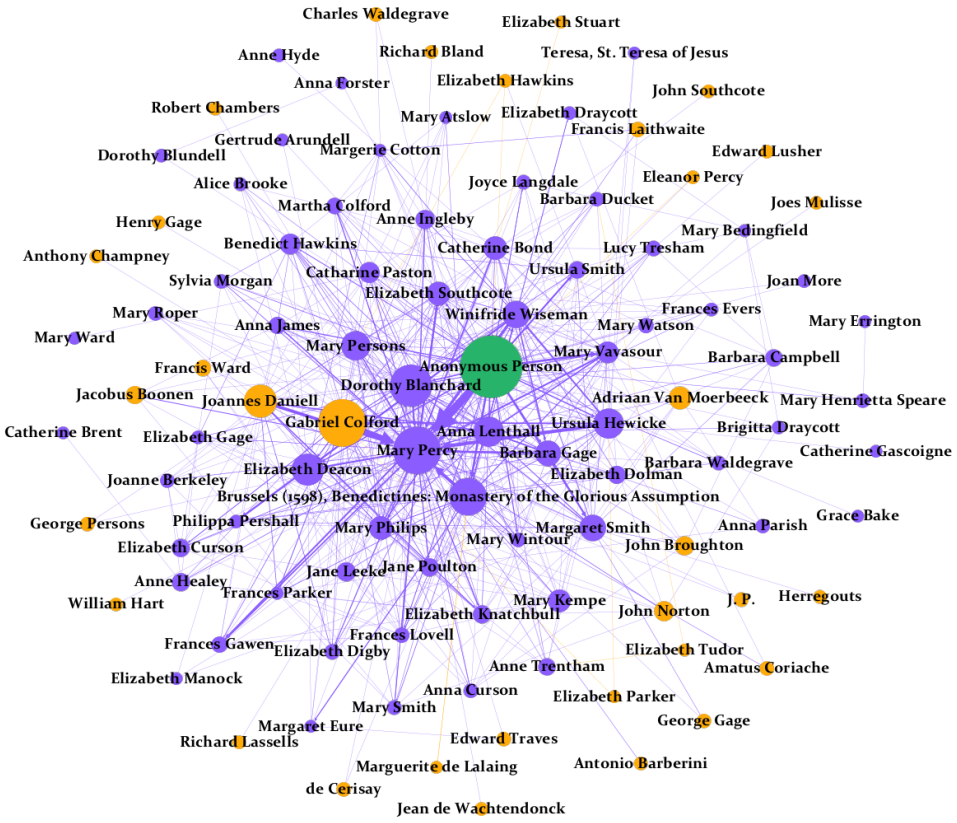


Figure 5: Brussels Benedictines Reception Network (ranked according to out-degree). All anonymous receptions are represented as a single "anonymous person" node in the network.

But the inclusion of anonymous receptions also results in the appearance of an additional female author, Grace Bake. Bake entered the Brussels convent in 1638 and remained there until her death in July 1676.[56] Just one letter written by her to Boonen survives in the archives and its anonymous reception, by a reader who underlined and annotated it, explains her appearance in this network.[57] By bringing to light hitherto hidden individuals such as Bake, the inclusion of anonymity allows for a more accurate and inclusive analysis of the Brussels' nuns epistolary activity and the reception of their letters beyond the convent walls. Again, this approach is not without limitations. Most obviously, it distorts the results by creating the impression that all instances of anonymous reception were carried out by the same person when, in fact, we cannot know whether the same anonymous person annotated just one letter or was responsible for multiple annotations on multiple letters. Thus, neither approach satisfactorily resolves the conundrum of anonymity but both highlight its importance to the circulation of the nuns' epistles.

7 Conclusion

In the last decade, the study of early modern English convents in exile has become a buoyant field of scholarly enquiry, driven in no small part by improved accessibility – both online and in print – to archival sources relating to these exiled institutions. As accessibility increases, scholars are deploying innovative tools and methodologies that can facilitate new revelations about the convents and their role in the wider post-Reformation history and culture of English and European Catholicism. Using a corpus of 405 letters produced by and about members of the English Benedictine convent in Brussels, this article has demonstrated the applicability of quantitative network analysis and visualisation methods to early modern convent sources and presented the insights that can be gleaned as a result. As this study has illustrated, however, quantitative network analysis methods are most valuable when combined with qualitative research and close reading of the primary sources. This combination opens up new ways of understanding the disputes that engulfed the English Benedictine convent during the 1620s and 1630s. What emerges confirms the centrality of key office holders but also throws up an outlier, whose centrality is not as evident using established qualitative methods. Significantly, the visualisations produced here graphically illustrate the breadth and extent of the nuns' epistolary relationships. In so doing, they progress debates about the role of enclosure in early modern convents and highlight how, despite geographical distance, family members were attuned to events within the convent cloister. The application of the translation filter nuances our understanding of the network and how it operated, bringing to light the importance of translation to the transmission of the nuns' letters. Finally, this study has sought to advance

---

56 WWTN (UID BB008).

57 "Grace Bake [in Brussels] to [Jacobus Boonen] [15 January 1639]", AAM, Doos 12/3.

discussion about the methodological challenges posed by anonymity, demonstrating how important it was to the reception and circulation of the nuns' correspondence and how challenging it is to account for using quantitative methods.

References

Ahnert, Ruth and Ahnert, Sebastian E. "Protestant Letter Networks in the Reign of Mary I: A Quantitative Approach." *English Literary History* 82 (2015): 1-33.

——— "Reconstructing Correspondence Networks in the State Papers Archive", unpublished paper delivered at "Reception, Reputation and Circulation in the Early Modern World" conference held at the National University of Ireland, Galway, 22-25 March 2017.

Ahnert, Ruth. "Maps Versus Networks." In *News Networks in Early Modern Europe*, edited by Noah Moxham and Joad Raymond, Leiden: Brill, 2016, 130-57.

Arblaster, Paul. "The Monastery of Our Lady of the Assumption in Brussels (1599-1794)." *English Benedictine History Symposium* 17 (1999): 54-77.

Bourke, Evan. "Female Involvement, Membership and Centrality: A Social Network Analysis of the Hartlib Circle." *Literature Compass* 14 (2017): 1-17.

Bowden, Caroline. "The English Convents in Exile and Questions of National Identity, *c*.1600-1688." In *British and Irish Emigrants and Exiles in Europe*, edited by David Worthington, Leiden; Boston, MA: Brill, 2010, 297-314.

Bowden, Caroline and Kelly, James E., eds. *The English Convents in Exile, 1600-1800: Communities,Culture and Identity*. Burlington, VT: Ashgate, 2013.

Coolahan, Marie-Louise. "Archipelagic Identities in Europe: Irish Nuns in English Convents." In *The English Convents in Exile, 1600-1800:*
Communities, Culture and Identity, edited by Caroline Bowden and James E. Kelly, Burlington, VT: Ashgate, 2013, 211-28.

——— "Nuns' Writing: Translation, Textual Mobility and Transnational Networks." In *A History of Early Modern Women's Writing*, edited by Patricia Phillippy, Cambridge: Cambridge University Press, 2018, 257-75.

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

Gingras, Yves. "Mapping the Structure of the Intellectual Field using Citation and Co-Citation Analysis of Correspondences." *History of European Ideas* 36:3 (2010): 330-39.

Goodrich, Jaime. "Translating Lady Mary Percy: Authorship and Authority among the Brussels Benedictines." In *The English Convents in Exile, 1600-1800:*

Communities, Culture and Identity, edited by Caroline Bowden and James E. Kelly, Burlington, VT: Ashgate, 2013, 109-22.

——— "Authority, Gender, and Monastic Piety: Controversies at the English Benedictine Convent in Brussels, 1620-1623." *British Catholic History* 33 (2016): 91-114.

Hallett, Nicky, ed. "The Life of Lucy Knatchbull." In *English Convents in Exile, 1600-1800*, edited by Caroline Bowden, vol. 3, London: Pickering & Chatto 2012-13, 159-217.

Harline, Craig and Putt, Eddy. *A Bishop's Tale: Mathias Hovius Among his Flock in Seventeenth-Century Flanders*. New Haven, CT: Yale University Press, 2000.

Lehfeldt, Elizabeth A. "The Permeable Cloister." In *The Ashgate Research Companion to Women and Gender in Early Modern Europe*, edited by Allyson M. Poska, Jane Couchman and Katherine A. McIver, Aldershot: Ashgate, 2013, 13-31.

Lux-Sterritt, Laurence. *English Benedictine Nuns in Exile in the Seventeenth Century: Living Spirituality*. Manchester: Manchester University Press, 2017.

Medioli, Francesca. "An Unequal Law: The Enforcement of Clausura before and after the Council of Trent." In *Women in Renaissance and Early Modern Europe*, edited by Christine E. Meek, Dublin: Four Courts Press, 2000, 136-52.

——— "The Dimensions of the Cloister: Enclosure, Constraint, and Protection in Seventeenth-Century Italy." In *Time, Space, and Women's Lives in Early Modern Europe*, edited by Anne Jacobson Schutte, Thomas Kuehn and Silvana Seidel Menchi, Kirksville, MO: Truman State University Press, 2001, 165-80.

Mullan, John. *Anonymity: A Secret History of English Literature*. London: Faber, 2007.

Murphy, Emilie K.M. "Language and Power in an English Convent in Exile, *c*.1621-*c*.1631." *The Historical Journal* (Forthcoming): 1-25 https://doi.org/10.1017/S0018246X17000437 (accessed 22 February 2018).

North, Marcy L. *The Anonymous Renaissance: Cultures of Discretion in Tudor-Stuart England*. Chicago, Ill.; London: University of Chicago Press, 2003.

——— "Anonymity." In *Encyclopaedia of English Renaissance Literature*, edited by Garret Sullivan and Alan Stewart, vol. 3, Oxford: Wiley-Blackwell, 2012, 8-12.

——— "Early Modern Anonymity." In *Oxford Handbooks Online*, edited by Colin Burrow, Oxford: Oxford University Press, 2015, 1-27.

Starner, Janet W. and Traister, Barbara H. eds. *Anonymity in Early Modern England: 'What's in a Name?'*. London: Routledge, 2016.

Vugt, Ingeborg van. "Using Multi-Layered Networks to Disclose Books in the Republic of Letters." *Journal of Historical Network Research* 1 (2017): 25-51.

Walker, Claire. *Gender and Politics in Early Modern Europe: English Convents in France and the Low Countries*. Basingstoke: Palgrave Macmillan, 2003.

——— "Doe not Supose me a Well Mortifyed Nun Dead to the World": Letter-Writing in Early Modern English Convents." In *Early Modern Women's Letter Writing 1450-1700*, edited by James Daybell, Basingstoke: Palgrave, 2001, 159-76.

——— "Recusants, Daughters and Sisters in Christ: English Nuns and their Communities in the Seventeenth Century." In *Women, Identities and Communities in Early Modern Europe*, edited by Stephanie Tarbin and Susan Broomhall, Aldershot: Ashgate, 2008, 61-76.

SCHAUF, ANDREW AND ESCOBAR VARELA, MIGUEL

Searching for hidden bridges in co-occurrence networks from Javanese *wayang kulit*

Journal of Historical Network Research 2 (2018) 26-52.

Keywords

Co-occurrence, Network null model, *Wayang kulit*, *Mahabharata*, Java (Indonesia)

Abstract

We propose that questions of long-standing interest in the study of *wayang kulit*, Indonesia's centuries-old shadow puppet theatre tradition, can potentially be posed in structural terms and investigated using the tools of network science. Here, we construct weighted character co-occurrence networks based on the Javanese *wayang kulit* incarnation of the Mahabharata epic, supplementing nodes with metadata specifying characters' tribal affiliations and historical origins in either Indian or Javanese traditions. In order to identify characters who play unique structural roles which other approaches may overlook, we generate null model ensembles of artificial networks that share the empirical networks' degree sequences, underlying episodic structures, and node metadata. By ranking nodes by the extent to which their betweenness centrality exceeds a null model's expectations, we reveal characters whose appearances in a story, while not necessarily

large in number, tend to serve the specific topological function of bridging groups of other characters. Decomposing betweenness centrality values into their inter-faction components then clarifies how these bridge-like characters are situated among the epic's various social factions. We observe that female characters, despite being few in number and appearing relatively infrequently, appear to dominate these rankings disproportionately. Analyses involving closeness centrality reveal low-closeness outliers whose appearances, although relatively frequent, keep them structurally isolated and distant from the rest of the Mahabharata universe; these include the epic's antagonists, the Korawa. Characters with historical origins in the Javanese tradition are found to be embedded just as closely within the network as are characters from the Indian canon when their degrees are taken into account using null models.

1 Introduction*

1.1 Social complexity in the Javanese *wayang kulit* incarnation of the *Mahabharata*

Like many stories of great historical and cultural importance, the *Mahabharata* is a tale of two warring factions. In Javanese *wayang kulit* shadow puppet theater, one of Southeast Asia's most important theater traditions, the *Mahabharata* is depicted episodically in a series of self-contained but connected stories, with a typical overnight performance covering one story (*lakon*) from the epic. The primary focus of these stories is the power struggle between two groups of adversaries: the *Pandawa* and the *Korawa*. However, the dramatic intrigue of the *Mahabharata* often stems from characters and relationships that straddle these identity lines. Some characters, due to complicated family ties or histories, juggle conflicting loyalties to both of the warring tribes. Deities intercede in the conflict, sometimes even simultaneously aiding members of both tribes against one another without ever decisively taking sides. This complexity is further enriched in the Javanese *wayang kulit* incarnation of the epic through the inclusion of characters and stories of indigeneous Javanese origin. For example, the *Punokawan* are a group of clown-servants with historical origins in pre-Hindu Javanese mythology, assimilated into *Mahabharata* stories to serve as advisors to the protagonist *Pandawa* tribe. Some plot elements, such as the so-called *carangan*

* **Acknowledgments:** We would like to thank Prof. Lai Choy Heng for helpful conversations.
Corresponding author: Andrew Schauf, NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, e-mail: andrew.schauf@u.nus.edu

("branch"[1]) stories, are also understood to represent Javanese additions to the original Indian epic[1].

1.2 Character co-occurrence networks and social complexity

As a recent explosion of research from the digital humanities demonstrates, network representations of stories can bring into sharper focus this notion that characters, rather than simply falling in line with one of two clearly-delineated sides of a conflict, are often bound together in a more complex tapestry of overlapping allegiances and rivalries[2]. By casting this web of relationships in a more tangible form, certain insights that might be missed by traditional scholarship can sometimes be brought to light. Each character and relationship within a fictional universe, while possessing intrinsic traits that are revealed and developed through events depicted in the narrative, can meanwhile also be considered in terms of the structural role it plays within this network. From this structural perspective, a work of fiction can be analyzed in terms of how it joins together — or, conversely, how it keeps separate — its various characters and social factions. Does the story tend to depict a certain faction more often via in-group interactions, or through their encounters with the outside universe? How does the story use certain characters or factions to connect, or to otherwise come between, the other characters and groups that appear within the work's fictional

---

1 Alan H. Feinstein, Lakon carangan: Ringkasan lakon, vol. 3 (Akademi Seni Karawitan Indonesia, 1986); Laurie Jo Sears, Shadows of empire: Colonial discourse and Javanese tales (Duke University Press, 1996); Sri Mulyono, Wayang, asal-usul, filsafat dan masa depannya (Gunung Agung, 1978); Jacob Kats, De wajang poerwa: een vorm van Javaans toneel, vol. 1 (Foris Publications, 1923); Victoria M. Clara van Groenendael, Wayang Theatre in Indonesia: An Annotated Bibliography (Foris Publications, 1987); James R. Brandon, Pandam Guritno, and Roger A. Long, On thrones of gold: Three Javanese shadow plays (University of Hawaii Press, 1993); Bernard Arps, Tall Tree, Nest of the Wind: The Javanese Shadow-play Dewa Ruci Performed by Ki Anom Soeroto: A Study in Performance Philology (NUS Press, 2016).

2 Aris Xanthos et al., "Visualising the dynamics of character networks," Digital Humanities (2016), 417–419; Michaël C. Waumans, Thibaut Nicodème, and Hugues Bersini, "Topology analysis of social networks extracted from literature," PloS ONE 10, no. 6 (2015): e0126470; Peer Trilcke, Frank Fischer, and Dario Kampkaspar, "Digital network analysis of dramatic texts," Proceedings of the Digital Humanities 6, no. 7 (2015): 8; Seung-Bo Park, Kyeong-Jin Oh, and Geun-Sik Jo, "Social network analysis in a movie using character-net," Multimedia Tools and Applications 59, no. 2 (2012): 601–627; Gyeong-Mi Park, Sung-Hwan Kim, and Hwan-Gue Cho, "Structural analysis on social network constructed from characters in literature texts," Journal of Computers 8, no. 9 (2013): 2442–2447; Frank Fischer et al., "Network Dynamics, Plot Analysis: Approaching the Progressive Structuration of Literary Texts," Digital Humanities 2017 (Montréal, 8–11 August 2017). Book of Abstracts (McGill University, 2017); Yeon-Mu Choi and Hyun-Joo Kim, "A directed network of Greek and Roman mythology," Physica A: Statistical Mechanics and its Applications 382, no. 2 (2007): 665–671.

universe? The quantitative tools provided by network science give scholars a promising framework with which to more systematically explore the deeper complexity beneath a story's most salient binary rivalry.

Character co-occurrence networks provide one such description of the complex scaffolding of character interactions upon which a story is played out. In these networks, the strengths of the connections between character nodes reflect how frequently those characters appear together within some suitably-defined temporal or spatial window. While the choice of co-occurrence window can affect the features of the resulting network, construction of such a network typically requires little or no prior subjective interpretation. Counting co-occurrence events by such a definition does not account for the qualitative aspects of an encounter, nor for the nature of the underlying social tie; co-occurrence may not even necessarily signify a particularly direct exchange between two characters. By no means does a co-occurrence network give a complete description of the fictional universe supposedly revealed by a narrative, which is ultimately too abstract, multifaceted, and subjective to be definitively characterized — let alone quantified — by any single approach. But by representing a story's aggregate of non-simultaneous character encounters as a structural object, a co-occurrence network can capture meaningful information about the patterns of interactions through which a particular realization of a story depicts this underlying universe.

1.3 Comparing distinct retellings of common stories

As the *wayang kulit* tradition has demonstrated to a unique extent for hundreds of years, a single storyline can be told and retold in numerous ways. Historically, *wayang kulit* scholarship has naturally involved comparisons of different realizations of the same set of canonical storylines, whether by comparing Javanese *wayang* to the Indian traditions from which it is understood to have evolved, by studying the evolution of its stories through time, or by comparing its various distinct regional traditions. Differences among different *dhalang* (puppetmasters) in delivering a common story are also of interest, as the ability to present novel re-interpretations of well-known stories is held highly within *wayang* culture. And especially since performances are often deliberately tailored to correspond to local current events, variations between specific performances are also of particular interest within this tradition.

Each of these distinct realizations may emphasize different characters, relationships, and events than another version of the same story does, and these are precisely the types of differences which a co-occurrence network perspective is potentially capable of discriminating. As we will discuss later, direct comparative analyses of multiple distinct realizations of the *Mahabharata* epic, particularly those representing different chronological stages from throughout the historical evolution of its Javanese *wayang kulit* incarnations, may involve

certain complications which place them beyond the scope of the current work. However, the potential for these comparative or dynamical historical network analyses in future work motivates our current initial step in that direction.

1.4 Historical origins and co-occurrence structure in *wayang kulit*

The traditional terminology of *wayang kulit* often carries vivid structural imagery. *Silsilah* (Arabic for "chain" or "linkage") are family trees of characters traditionally included in *wayang* encyclopedias. Even the name of the *carangan* (Javanese for "branch") stories of the Javanese tradition offers a description of their structural function within the epic, casting them as asides of secondary importance deviating from the epic's proper core of *parwa* or *pokok* ("trunk") storylines from the Indian canon. But are the structural claims seemingly implicit in this terminology actually borne out in the scaffolding of character co-occurrences established throughout the course of the stories? Even in lieu of sufficiently detailed data sets which would allow for a more direct observation of how Javanese elements were assimilated into the structure of the Indian *Mahabharata* over time, we propose that co-occurrence structures representing its current incarnations may provide some novel — albeit indirect — insights into the historical processes by which they were formed. Network science provides a clear framework with which to approach the question: How centrally have Javanese innovations to the *Mahabharata* come to be embedded within the epic? This structural perspective could thus complement more traditional approaches to address these issues of long-standing relevance in the study of *wayang kulit* history.

1.5 Scope of the current work

As a first step along these lines, we make an initial network-theoretical venture into the world of *wayang kulit* by constructing weighted, undirected co-occurrence networks based on one representative description of the most popular stories of the Javanese *wayang kulit Mahabharata* tradition[3]. Nodes are endowed with metadata describing characters' affiliations with various social factions based on information which we have compiled in an interactive Digital Wayang Encyclopedia[4]. Viewing this approach as an exploratory framework which may reveal previously unseen patterns within the epic — as a "defamiliarization strategy" with which to glean fresh insights from widely-

3 Purwadi, Kempalan Balungan Lakon Wayang Purwa [Compendium of Classical Wayang Plot Scenarios] (Cendrawasih, 2008).

4 Miguel Escobar Varela, Digital Wayang Encyclopedia (available online under the following URL: https://villaorlado.github.io/wayangnetworks/html, accessed 23 November 2018).

studied works[5] — we seek to identify characters and relationships whose structural roles may not be so immediately visible, even when viewed from other quantitative perspectives. That is, we begin by deliberately seeking out characters whose importance, as quantified by various network centrality measures, cannot simply be explained by how frequently they appear.

The most substantial advantages to be gained through the adoption of a network perspective involve not just superficial renamings of certain quantities (e.g. a character's number of episodes as its "degree"), but rather account for larger-scale structural patterns and chains of non-simultaneous co-occurrence events that extend through a network. Unlike other network-theoretical measures such as degree which only take into account the local environment immediately surrounding a node, higher-order centrality measures such as betweenness[6] and closeness centrality[7] describe a node's position amid extended chains of co-occurrence relationships. Betweenness centrality sums the fractions of shortest paths between node pairs which pass through a given node or link, and closeness is proportional to the inverse sum of the lengths of all shortest paths between a given node and all other nodes. By considering larger-scale patterns that form as the aggregate of non-simultaneous encounters between multiple characters, these measures are uniquely network-theoretical in nature, and so may reveal insights that are invisible to other methods. While the methods used in the current work can be applied to a wide variety of these centrality measures, here we focus specifically on betweenness and closeness centrality because of their intuitive interpretations in the context of co-occurrence networks. They allow us to address specific issues of interest in the study of *wayang kulit*: how various social factions mediate — or otherwise find themselves situated between — the *Pandawa* and *Korawa*, and whether characters with historical origins in the Javanese tradition tend to be closely embedded among other characters within the network, or rather are relegated to peripheral positions or "branches" detached from the network core.

Since weighted-network centrality measures inherently tend to take on higher values for more strongly-connected nodes, ranking nodes by their raw centrality values may lead us to conflate a node's overall connection strength with the specific topological features we wish to target when exploring these questions. We account for this using null models, which clarify the extent to which a character's centrality values are explained by its overall number of

5 Franco Moretti, Graphs, maps, trees: abstract models for a literary history (Verso, 2005).

6 Linton C. Freeman, "A set of measures of centrality based on betweenness," Sociometry, 1977, 35–41.

7 Linton C. Freeman, "Centrality in social networks: I. Conceptual clarification," Social networks 1, no. 3 (1978): 215–239.

appearances within the context of the epic, and so allow us to identify nodes whose centrality values can be more rightfully attributed to special, higher-order patterns of co-occurrence established throughout the course of the epic. We then use these insights to guide our explorations of how *wayang kulit* brings together the various social factions of the *Mahabharata* universe.

These explorations reveal several structural features distinguishing the epic's tribal factions: the protagonist *Pandawa* are often depicted in encounters with non-*Pandawa* characters, whereas other tribal factions are most often depicted through encounters with their in-group peers, with one character acting as the community's primary bridge to the rest of the network. While these secondary tribes both form tight-knit structural communities, their differing portrayals are also reflected in the results of null model comparisons: the *Punokawan*, who act as allies to the *Pandawa*, are positioned closer to the rest of the network than are the antagonist *Korawa* tribe. Female characters often emerge as outliers in the search for structural bridges, reflecting common features of their roles within wayang kulit. Characters who were integrated into the *Mahabharata* during the evolution of the Javanese *wayang kulit* tradition from its original Indian roots are shown to be embedded just as centrally within the network (in a sense we will make more precise later) as are canonical Indian characters.

2 Methods

2.1 Constructing co-occurrence networks

We construct weighted, undirected co-occurrence networks based on lists of characters appearing in each scene (*adegan*) from 23 commonly-performed stories (*lakon*) as extracted from a representative compendium of plot synopses from the Javanese *wayang kulit* incarnation of the *Mahabharata*[8]. Since many details of performances are not standardized, no single, definitive source text or recording exists; rather, these synopses outline the key plot points typically depicted in each scene and specify which characters are involved[9]. We thus consider these stories and scenes as the intrinsic narrative units of the epic, and rather than using sliding windows based on clock time within a particular recording or proximity within a particular transcription, we use appearances

8 Purwadi, Kempalan Balungan Lakon Wayang Purwa [Compendium of Classical Wayang Plot Scenarios] (Cendrawasih, 2008).

9 English language synopses for these stories are accessible through our Digital Wayang Encyclopedia (Miguel Escobar Varela, Digital Wayang Encyclopedia (available online under the following URL: https://villaorlado.github.io/wayangnetworks/html, accessed 23 November 2018).

within these traditional subdivisions of the epic to define character co-occurrence. Here, we consider two levels of co-occurrence granularity corresponding to the epic's subdivision into 23 *lakon* and then further into 223 *adegan*. In the following, we focus primarily on the finer-resolution *adegan* co-occurrence network unless otherwise noted, occasionally referring to the coarser *lakon* co-occurrence network to gain insights into how findings are affected by the choice of co-occurrence granularity.

Lists of characters appearing in each episode (that is, in each story or each scene, depending on the choice of co-occurrence window) are used to construct a bipartite character-episode affiliation network $B = (\boldsymbol{C}, \boldsymbol{E}, \boldsymbol{L})$, where $\boldsymbol{C}$ denotes the set of nodes representing characters, $\boldsymbol{E}$ denotes the set of nodes representing episodes, and $\boldsymbol{L}$ denotes the set of links representing character occurrences in these episodes, with the presence of a link $(c, e) \in \boldsymbol{L}$ indicating that character c appears in episode e. A character co-occurrence network G is then constructed as the weighted, unipartite projection of this bipartite network onto the set of character nodes: $G = (C, \boldsymbol{L_E}, w)$, where $\boldsymbol{L_E}$ is a set of co-occurrence links and $w : \boldsymbol{C} \times \boldsymbol{C} \rightarrow \mathbb{N}$ gives the weights for each pair (c_1, c_2) of characters $c_1, c_2 \in \boldsymbol{C}$ as

$$w\big((c_1, c_2)\big) = \|\{e \in \boldsymbol{E} \mid (c_1, e), (c_2, e) \in \boldsymbol{L}\}\|, \quad (1)$$

with $\|\cdot\|$ indicating the number of elements in the set. Here, $(c_1, c_2) \in \boldsymbol{L_E}$ only if $w\big((c_1, c_2)\big) > 0$. Link weights thus represent the number of distinct episodes in which both of the linked characters appear. In the following, we refer to a node's number of links within the character-affiliation network B (that is, the total number of episodes in which the corresponding character appears) as its *degree*, as distinct from the total sum of its link weights within the character co-occurrence network G, which we refer to as its *node strength*.

2.2 Identifying centrality outliers

When identifying the shortest paths between pairs of nodes, weighted-network generalizations of centrality measures often define the distance associated with the traversal of a link as the reciprocal of the link's weight[10]. In fictional co-occurrence networks, a densely-interconnected core of main characters is typically surrounded by numerous peripheral characters within a network of small diameter. These frequently co-occurring main characters tend to share links that have weights of a higher order of magnitude than those of peripheral characters, and so inherently tend to support the majority of shortest paths between nodes in the network. Large-degree nodes will thus tend to exhibit the

10 Mark E. J. Newman, "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality," Physical Review E 64, no. 1 (2001): 016132.

highest centrality values, even when their primary structural roles — as characterized by the topological "shape" of their connections, and not just by overall magnitudes of their link weights — may not correspond to the features which one wishes to target using a certain centrality measure. We thus wish to systematically identify characters whose appearances, though possibly few in number, are arranged to serve a specific structural role targeted by the centrality measure.

In order to identify nodes whose centrality measures are higher or lower than expected given the overall strengths of their connections, we uses ensemble null models. By averaging centrality values of each character over an ensemble of artificial networks, each of which gives characters the same levels of overall prominence that they exhibit in the actual network, we can estimate the ranges of centrality values that are expected to result from the network's combination of node degrees. A variety of approaches exist for estimating the weights of each link within a weighted network given only a specified degree sequence[11,12]. In the current context, however, link weights count discrete co-occurrence events, and so are constrained to take values in the natural numbers (Equation 1). So, we design our null model to reflect the co-occurrence network's origins as the weighted projection of an unweighted affiliation network, in addition to reproducing its degree sequences and any relevant node metadata.

We consider two methods for generating these artificial networks. First, we use a *rewired* model in which networks are generated by iteratively rewiring the empirical character-episode affiliation network $B = (\boldsymbol{C}, \boldsymbol{E}, \boldsymbol{L})$. To produce each artificial network realization B_i' of an ensemble B', we begin with a copy of the empirical link set, $\boldsymbol{L'_i} = L$. At each step of an iterated process, a randomly-selected pair of distinct links $(c_1, e_1), (c_2, e_2) \in \boldsymbol{L'_i}$ is removed and replaced with $(c_1, e_2), (c_2, e_1) \in \boldsymbol{L'_i}$ until a set number of iterations (here, $50\|\boldsymbol{L}\|$) is reached. The resulting rewired link set $\boldsymbol{L'_i}$ defines a new network realization $B_i' = (\boldsymbol{C}, \boldsymbol{E}, \boldsymbol{L'_i})$, the degree sequences of which are identical to those of the original network.

We also consider a *configuration model*[13], adapted here for the bipartite network context at hand. Given a character-episode affiliation network B with target character degree sequence $\{m_c\}$ and episode degree sequence $\{n_e\}$, we

11 Iacopo Mastromatteo, Elia Zarinelli, and Matteo Marsili, "Reconstruction of financial networks for robust estimation of systemic risk," Journal of Statistical Mechanics: Theory and Experiment 2012, no. 03 (2012): P03011

12 Giulio Cimini et al., "Estimating topological properties of weighted networks from limited information," Physical Review E 92, no. 4 (2015): 040802.

13 Fan Chung and Linyuan Lu, "The average distance in a random graph with given expected degrees," Internet Mathematics 1, no. 1 (2004): 91–113.

generate each network $B_i' \in B'$ by allowing each potential link affiliating a character node c with an episode node e to be realized with a probability

$$P\big((c,e) \in \mathrm{L}\big) = \frac{m_c n_e}{\|L\|}. \tag{2}$$

The degree sequences of each individual network may not attain their exact target values in each B_i', but their mean values in the ensemble B' approach those of the target sequences as the total number of network realizations increases. As a result, nodes with low total strength in the empirical network will remain disconnected from any network realizations wherein no links are realized for that node under the probabilistic generation of links in accord with Equation 2. These disconnected nodes will be excluded from the shortest path counting used to compute centrality values for all other nodes.

Projection of each artificial character-episode affiliation network $B_i' \in B'$ onto C then yields an ensemble of the corresponding co-occurrence networks $G_i' \in G'$ upon which the measures of interest are computed. In the following, we present results for the *adegan* co-occurrence network based on *rewired* and *configuration* model ensembles of 1000 network realizations each, as well as a 500-network rewired ensemble and an 800-network configuration model ensemble for the *lakon* co-occurrence network.

For a centrality measure b measured on a node c in G, the expected value of $b_G(c)$ according to a null model ensemble of network realizations $G' = \{G'_1, G'_2, \ldots\}$ is given by

$$\langle b_{G'}(c)\rangle_{G' \in G'} = \frac{1}{\|G'\|} \sum_{G' \in G'} b_{G'}(c)\,. \tag{3}$$

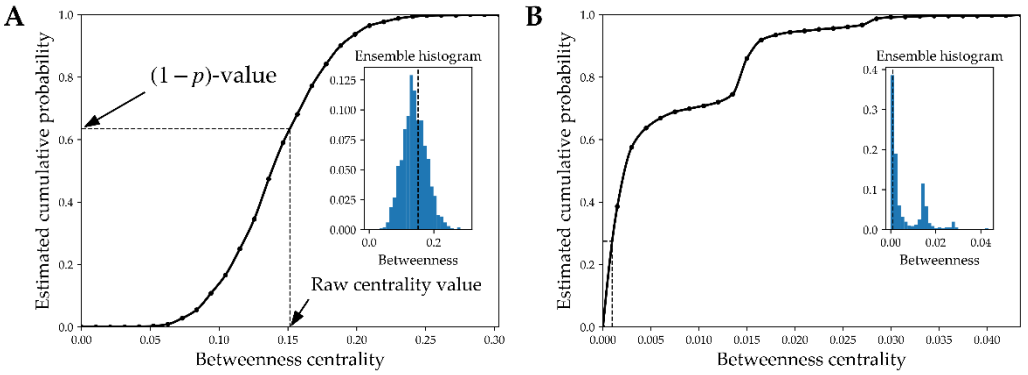
Bootstrapping from histograms based on the set of null model values $\{b_{G_1'}(c), b_{G_2'}(c), \ldots\}$, we estimate the relative values of the probability distribution $P\big(b_{G'}(c)\big)$ at histogram bin centers, and interpolate these to estimate a corresponding continuous cumulative probability distribution $P\big(x < b_{G'}(c)\big)$. The "p-value"[14] corresponding to the empirical network value $b_G(c)$ is given by

$$p\big(b_G(c)\big) = 1 - P\big(b_G(c) < b_{G'}(c)\big)\,, \tag{4}$$

14 We refer to these as "p-values" in a slight deviation from the usual statistics terminology; here, p = .05 would indicate a 5% chance that the null model network would achieve a value greater than the empirical value, while p = .95 would indicate a 5% chance that a null model network achieves a value less than than empirical value.

as illustrated in Figure 1. The p-values corresponding to node and link betweenness centrality, closeness centrality, and global network metrics are computed similarly. In the following rankings, we sort nodes and links by these p-values; elements with a shared p-value are subjected to a secondary sort based on the ratio $b_G(c)/\langle b_{G'}(c)\rangle_{G'\in G'}$ (which we will henceforth denote as $b/\langle b\rangle$)[15].

Figure 1. Computation of *p*-values: Histograms of null model ensemble centrality values (insets) and bootstrapped cumulative probability functions used to compute p-values for **(A)** Gatotkaca and **(B)** Banowati from a rewired null model based on the *adegan* co-occurrence network.



2.3 Social factions and centrality

To investigate how characters of different types are embedded among one another throughout the network, nodes are endowed with metadata describing a character's affiliations with various social factions. We consider four partitions of the set of character nodes into f disjoint sets of interest $\boldsymbol{F}_i$, with $\bigcup_{i=1}^{f} F_i = C$ and $\boldsymbol{F}_i \cap \boldsymbol{F}_j = \emptyset$ for $i \neq j$. Each character is labelled by its *Tribe* affiliation[16] (*Korawa*, *Pandawa*, *Punokawan*, or Unaffiliated), by the *Mahabharata* tradition of its historical *Origin* (India or Java), by *Species* (Animal, Demon, Human, God, or *Raksasa*), and by *Gender* (Female or Male). More detailed information on these

15 Stefan Hennemann and Ben Derudder, "An alternative approach to the calculation and analysis of connectivity in the world city network," Environment and Planning B: Planning and Design 41, no. 3 (2014): 392–412.

16 Given the complexity and ambiguity of some allegiances, here we assign a tribe label only where a character's primary affiliation is unambiguously clear.

characters and the choices of metadata labels assigned can be consulted at our Digital Wayang Enyclopedia[17].

In order to quantify how various characters and social factions are situated between one another amid the network's co-occurrence patterns, we quantify the extent to which a node or link falls between members of two factions by restricting the shortest path counts used to compute betweenness such that they consider only shortest paths that join members of each of these two specific factions. The betweenness centrality of a node c in the network $G = (\boldsymbol{C}, L_E, w)$ is given by

$$b_G(c) = \frac{2}{(\|\boldsymbol{C}\| - 1)(\|\boldsymbol{C}\| - 2)} \sum_{s,t \in \boldsymbol{C}} \frac{\sigma_G(s,t|c)}{\sigma_G(s,t)}, \tag{5}$$

where $\sigma_G(s,t)$ counts the number of distinct shortest paths between nodes s and t in the network G, and $\sigma_G(s,t|c)$ counts the number of those shortest paths which pass through the character node c[18]. These shortest paths are identified by computing the length D of a path described by node sequence $\{c_i\}$ as the sum of the inverse weights of the links traversed[19]:

$$D(\{c_i\}) = \sum_{j=1}^{\|\boldsymbol{C}\|-1} \frac{1}{w\left((c_j, c_{j+1})\right)}. \tag{6}$$

For two factions F_1 and F_2, we define the F_1-F_2 *inter-faction betweenness centrality* in which we restrict shortest path counting to only consider paths connecting pairs having one member each from the two factions of interest:

$$b_{(F_1,F_2),G}(c) = \frac{2}{(\|\boldsymbol{C}\| - 1)(\|\boldsymbol{C}\| - 2)} \sum_{s \in F_1, t \in F_2} \frac{\sigma_G(s,t|c)}{\sigma_G(s,t)}. \tag{7}$$

17 Miguel Escobar Varela, Digital Wayang Encyclopedia, https://villaorlado.github.io/wayangnetworks/html/.

18 Linton C. Freeman, "A set of measures of centrality based on betweenness," Sociometry, 1977, 35–41.

19 Mark E. J. Newman, "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality," Physical Review E 64, no. 1 (2001): 016132.

We also define the *faction-world betweenness centrality* with respect to a faction F by restricting shortest path counting to paths connecting nodes of a given faction to target nodes outside the faction:

$$b_{(F,\overline{F}),G}(c) = \frac{2}{(\|\boldsymbol{C}\| - 1)(\|\boldsymbol{C}\| - 2)} \sum_{s \in F,\, t \in \overline{F}} \frac{\sigma_G(s,t|c)}{\sigma_G(s,t)}. \quad (7)$$

where $\overline{F}$ is the set of character nodes from C not in F.

3 Results

3.1 Global network metrics and metadata modularity reveal community structure along tribal lines

In order to understand the macroscopic differences between the empirical network and the null model, we first compare several global network metrics from the empirical network to the corresponding null model ensemble mean values (Table 1). The empirical network is less topologically dense[20], larger in diameter, and exhibits higher clustering[21] than most of its null model counterparts, as might be expected based on experience with real social networks. Modularity[22,23] values corresponding to each of the metadata-based partitions of the epic's characters (*Tribe* affiliations, Indian or Javanese historical *Origin*, *Species*, and *Gender*) suggest that this clustering is indeed related to preferential co-occurrence based on shared metadata characteristics. The highest of these values is observed for partition of nodes based on tribal affiliation; this tribe-based clustering is indeed confirmed in mean allocations of node strength for each tribe (Figure 2A) and is readily visible in a force-directed network visualization (Figure 3A), particularly for the *Korawa* and *Punokawan* factions.

---

20 Density is equal to $2\|\mathbf{L_E}\|/[\|\mathbf{C}\|(\|\mathbf{C}\| - 1)]$, the fraction of potential links that have non-zero weight in the character co-occurrence network.

21 Duncan J Watts and Steven H Strogatz, "Collective dynamics of 'small-world' networks," Nature 393, no. 6684 (1998): 440; Jukka-Pekka Onnela et al., "Intensity and coherence of motifs in weighted complex networks," Physical Review E 71, no. 6 (2005): 065103.

22 Mark E. J. Newman and Michelle Girvan, "Finding and evaluating community structure in networks," Physical Review E 69, no. 2 (2004): 026113; M. E. J. Newman, Networks: An Introduction (Oxford University Press, 2010).

23 These modularity values are computed with reference to an assumed unipartite network configuration model as the null model, rather than with reference to the bipartite network null models generated here.

3.2 Null models identify centrality outliers at all levels of overall prominence

As illustrated by the visible lack of simple monotonic trend between raw centrality values and the corresponding *p*-values for both betweenness (Figure 4B) and closeness (Figure 4D), null model *p*-values indeed provide information that is qualitatively distinct from that provided by the corresponding raw centrality values. While in some sense the null model is intended to rescale the raw values in a way that accounts for node degrees, the null model's manner of accounting for degrees is also distinct from discarding link weights outright, as is confirmed in similar plots of the unweighted-network counterparts of these centrality measures which treat all co-occurrence weights as identically equal to 1 (Figures 4A and 4C). Over the entire range of node degrees, then, the approach distinguishes both high- and low-centrality outliers from characters whose centrality values more closely conform with null model expectations. Viewing the empirical network's betweenness centrality values within the context of the null model's predicted dependence of betweenness on degree clearly reveals number of high-betweenness outliers of both low and high degree (Figure 5A). High-betweenness outliers (ranked in Table 2) include many less-prominent characters whose appearances, though few, nonetheless serve to bridge other characters via chains of co-occurrence interactions.

Table 1. Global metrics for the adegan co-occurrence network: Empirical network values compared to results from *rewired* and *configuration* null model ensembles.

Network metric	Empirical network	*Rewired* mean	*Rewired* *p*-value	*Configuration* mean	*Configuration* *p*-value
Number of links $\|\|\boldsymbol{L}\|\|$	1032	1032	.5369	1030.05	.4947
Largest component	146	146	.5369	128.12	.0000
Diameter	5	3.78	.0000	3.44	.0000
Mean shortest path	1.34	1.28	.1122	1.11	.0000
Density (topological)	.10	.14	.9999	.15	.9999
Mean clustering coeff.	.72	.38	.1227	.05	.0000
Species modularity	.08	-.006	.0000	-.007	.0000
Origin modularity	.06	-.006	.0000	-.007	.0000
Tribe modularity	.17	-.017	.0000	-.018	.0000
Gender modularity	.01	-.001	.0020	-.002	.0007

Figure 2. (A) Allocations of node strength among tribes, averaged over each tribe for the *adegan* co-occurrence network, and similarly for **(B)** an alternative partition of the nodes by historical origin, but with *Korawa* and *Punokawan* communities considered separately.

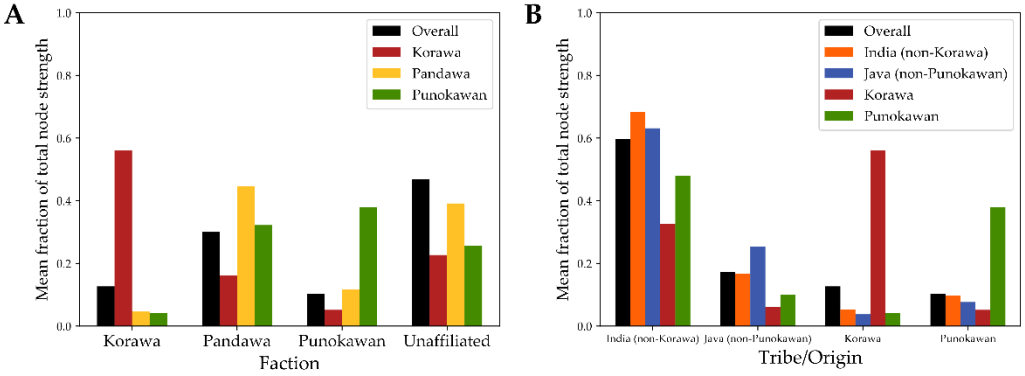
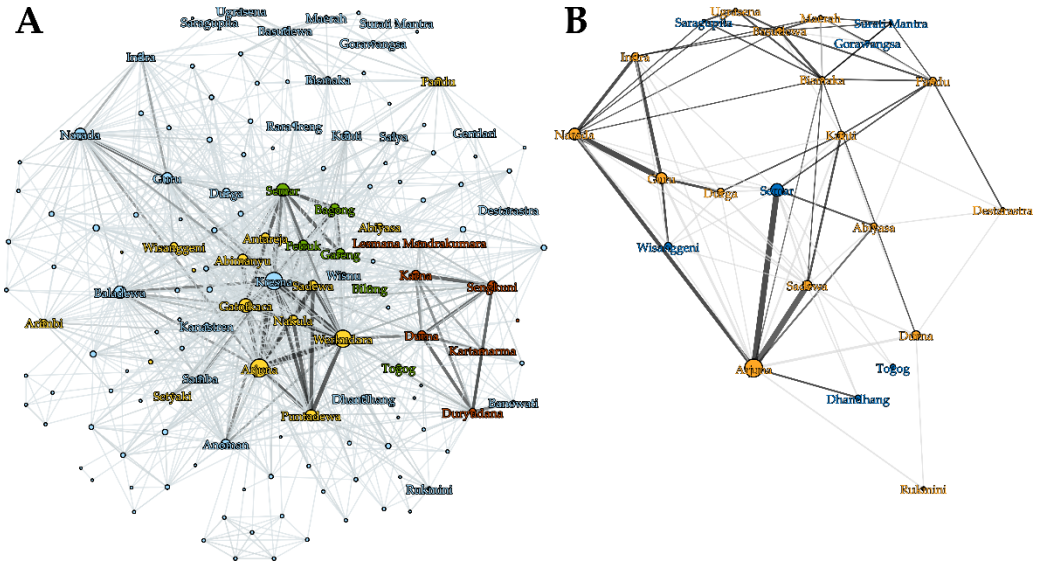


Figure 3. *Adegan* co-occurrence network: (A) Force-directed network visualization showing tribal affiliations (Korawa (red), Pandawa (yellow), Punokawan (green), and Unaffiliated (light blue)). **(B)** Detail of high-betweenness outliers showing characters' historical origins (India (orange) and Java (blue)). Darker link color indicates a lower link betweenness *p*-value.



However, even for higher-profile characters whose importance is already readily apparent through their high degrees and raw centrality values, these *p*-values offer a more complete account of their structural roles than do the raw values alone. Characters like Arjuna, Werkudara, and Durna not only have high betweenness centrality values, but these values are also significantly higher than expected; not only do they appear often, but the contexts of their appearances

situate them on paths between other characters to an even greater extent than can be explained by such high frequencies of appearances alone. This distinguishes them from other high-profile characters such as Kresna, Gatotkaca, or Sengkuni, who are highly central but are not situated between other characters significantly more than their frequent appearances would, in themselves, lead one to expect. Viewing closeness centrality values in the context of the null model's predicted dependence of closeness on degree (Figure 5B) reveals a number of low-closeness outliers as ranked in Table 3. Higher-degree characters tend to conform more closely with null model expectations, with the most salient exception of three extreme low-closeness outliers, who are all members of the antagonist *Korawa* tribe.

Figure 4. Character $(1 - p)$-values versus raw centrality values for the *adegan* co-occurrence network from a *rewired* null model, distinguishing Indian characters (orange), *Punokawan* (green), and non-*Punokawan* Javanese characters (blue): **(A)** Unweighted betweenness centrality, **(B)** Weighted betweenness centrality, **(C)** Weighted closeness centrality, and **(D)** Unweighted closeness centrality.

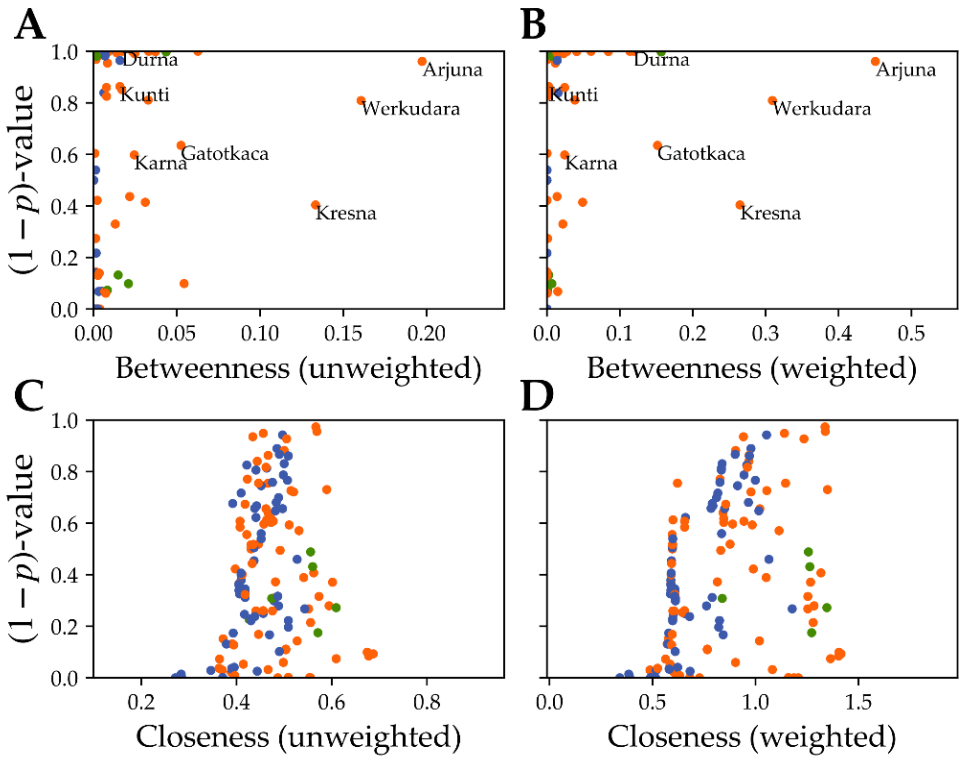


Table 2. High betweenness centrality outliers as ranked by *p*-value (ascending) and then by the ratio $b/\langle b \rangle$ (descending) for the *adegan* co-occurrence network.

Adj. rank	Raw rank	Character	Species	Origin	Tribe	Gender	Degree	Node strength	Betweenness centrality	*p*-value (B.C.)	$b/\langle b \rangle$ (B.C)
1	6	Pandu	Human	India	Pandawa	M	8	36	.060	.0000	40.62
2	7	Narada	God	India	Unaffiliated	M	25	105	.114	.0000	4.6
3	6	Durna	Human	India	Korawa	M	26	110	.119	.0000	4.4
4	13	Rukmini	Human	India	Unaffiliated	F	4	16	.027	.0005	89
5	11	Durga	God	India	Unaffiliated	F	10	42	.041	.0010	14.99
6	8	Guru	God	India	Unaffiliated	M	22	107	.084	.0010	4.87
7	16	Bismaka	Human	India	Unaffiliated	M	7	26	.024	.0015	20.11
8	4	Semar	God	Java	Punokawan	M	34	192	.157	.0025	2.54
9	23	Saragupita	Human	Java	Unaffiliated	M	4	9	.014	.0034	50.22
10	26	Surati Mantra	Raksasa	Java	Unaffiliated	M	2	8	.001	.0059	10.78
11	24	Ugrasena	Human	India	Unaffiliated	M	5	12	.014	.0061	27.1
12	25	Dhandhang	Demon	Java	Unaffiliated	M	4	22	.014	.0069	42.61
13	17	Basudewa	Human	India	Unaffiliated	M	9	36	.023	.0094	11.08
14	26	Gorawangsa	Raksasa	Java	Unaffiliated	M	3	10	.013	.0182	41.25
15	30	Destarastra	Human	India	Unaffiliated	M	4	13	.005	.0189	17.79
16	29	Togog	God	Java	Punokawan	M	4	22	.006	.0190	18.07
17	38	Maerah	Human	India	Unaffiliated	F	4	15	.001	.0320	2.45
18	21	Wisanggeni	Human	Java	Pandawa	M	9	39	.014	.0359	5.98
19	1	Arjuna	Human	India	Pandawa	M	62	310	.451	.0379	1.27
20	18	Abiyasa	Human	India	Pandawa	M	6	33	.012	.0458	12.65

3.3 Null model *p*-values eludicate bridging features within the network's core-periphery structure

Considering these *p*-values now within the context of a network visualization by highlighting high-betweenness outliers as in Figure 3B, we observe some intricate finer features embedded within the network's basic coreperiphery structure. These include an aggregate of bridge-like features anchored to the core of the network on one side by high-betweenness outliers such as Durga, Guru, and Narada, and bypassing the network core to approach Pandu – father of the protagonist *Pandawa* tribe and the #1-ranked high-betweenness outlier – and the *Korawa* on the other end. Having identified these characters as high-betweenness outliers, we can look at their allocations of betweenness centrality among specific inter-faction components, and so clarify their positions as

bridges between certain factions. Durga, for example, acts mostly as a bridge between gods and other characters (Figure 6A).

Figure 5. (A) Betweenness and (B) Closeness centrality versus node degree for the *adegan* co-occurrence network plotted atop density plots based on histogram data from the *rewired* null model.

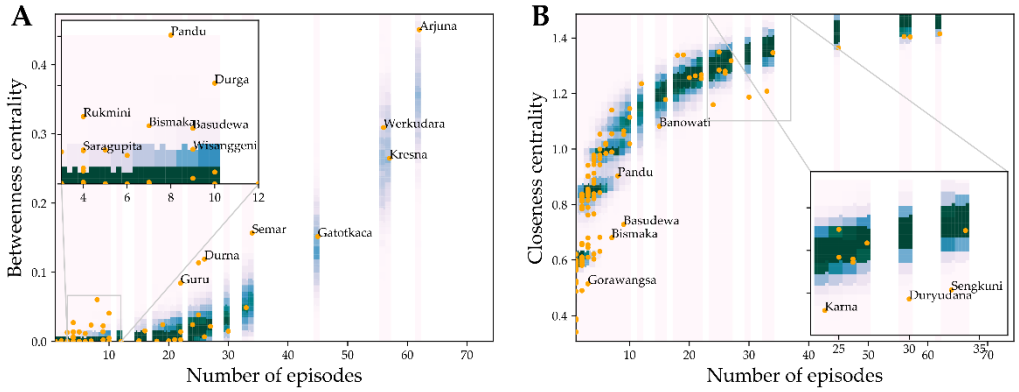


Table 3. Low closeness centrality outliers as ranked by p-value (descending) and then by the ratio $b/\langle b\rangle$ (ascending) for the *adegan* co-occurrence network.

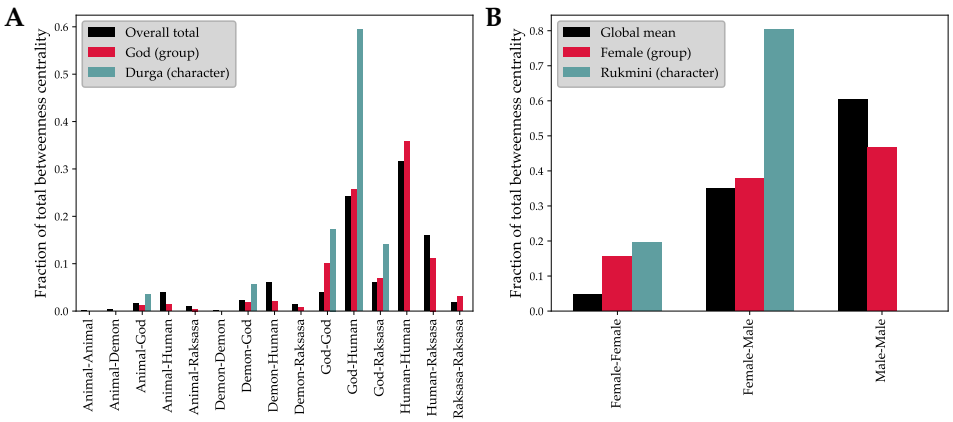
Adj. rank	Raw rank	Character	Species	Origin	Tribe	Gender	Degree	Node strength	Closeness centrality	p-value (C.C.)	$b/\langle b\rangle$ (C.C.)
1	1	Jaya Bajra	Raksasa	Java	Unaffiliated	M	1	2	.341	.9999	.57
2	6	Gorawangsa	Raksasa	Java	Unaffiliated	M	3	10	.515	.9999	.65
3	67	Basudewa	Human	India	Unaffiliated	M	9	36	.728	.9999	.68
4	127	Duryudana	Human	India	Korawa	M	30	93	1.187	.9999	.88
5	128	Sengkuni	Human	India	Korawa	M	33	121	1.208	.9999	.88
6	125	Karna	Human	India	Korawa	M	24	88	1.160	.9990	.89
7	5	Surati Mantra	Raksasa	Java	Unaffiliated	M	2	8	.490	.9990	.71
8	65	Bismaka	Human	India	Unaffiliated	M	7	26	.681	.9984	.68
9	3	Ngembat Landeyan	Human	Java	Unaffiliated	M	1	1	.388	.9953	.65
10	57	Ugrasena	Human	India	Unaffiliated	M	5	12	.632	.9922	.69

3.4 Structural distinctions between the protagonist *Pandawa* tribe, their *Punokawan* allies, and their *Korawa* rivals

As indicated previously by modularity values (Table 1) and mean node strength allocations for the partition of the network by Tribe (Figure 2), most members of the *Korawa* and *Punokawan* are depicted primarily through interactions with members of their own tribes. Applying the faction-specific betweenness

centralities discussed above, we examine the relative contributions made by each tribe to the network-wide sum of intra-faction betweenness for each tribe (that is, *Korawa-Korawa*, *Pandawa-Pandawa*, and *Punokawan-Punokawan* inter-faction betweenness centrality as defined in Equation 7) in order to observe which (if any) factions are embedded internally between members of each group (Figure 7A). The *Korawa* and *Punokawan* communities are each internally bridged only by members of their own faction; meanwhile, some *Pandawa* are internally connected via non-*Pandawa* characters: the god Kresna and the *Punokawan* leader Semar.

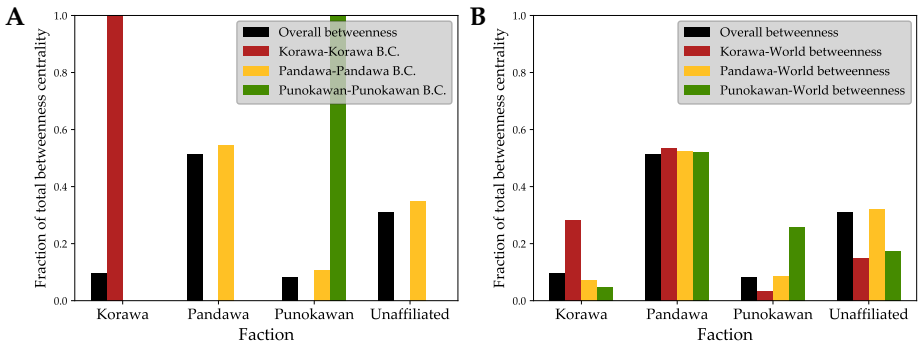
Figure 6. Female high-betweenness outliers: Allocations of overall betweenness centrality among its interfaction components for **(A)** Durga with respect to *Species*, and for **(B)** Rukmini with respect to the *Gender* partition.



Examining how these groups are bridged to the outside universe in terms of the overall allocations of faction-world betweenness among tribes for each group (Figure 7B), we find clues that the *Korawa* and *Punokawan* communities are each bridged to the wider universe via a primary in-group representative. Indeed, rankings of the top high-betweenness outliers (Table 2) include a primary character representing each of these tribes: Durna of the *Korawa* and Semar of the *Punokawan*. Decomposing these characters' betweenness centrality values into their inter-faction components with respect to the *Tribe* partition, we confirm their roles as representatives who specifically bridge their own factions with the outside universe (Figure 8). Durna's co-occurrences with his mentor Werkudara of the *Pandawa* give their link the third highest raw link betweenness centrality value in the network, and the highest of any *Korawa-Pandawa* link. While the role of Semar as the representative that bridges the other *Punokawan* to the world is obvious, adjusted rankings also identify the unique role of a much lower-profile *Punokawan* character, Togog. Togog appears in just 4 scenes,

but is ranked at #15 in adjusted betweenness rankings, and at #1 in adjusted rankings of high *Punokawan*-world and *Korawa-Punokawan* betweenness outliers.

Figure 7. Internal and external betweenness by tribe: Fractions of total global **(A)** Intra-faction betweenness and **(B)** Faction-world betweenness contributed by each tribe.



Togog is a member of the clown-servant *Punokawan* who differs from other *Punokawan* in that he associates with the *Korawa* rather than the *Pandawa*. Null model rankings uncover the otherwise-obscured structural manifestations of this character's unique role in the narrative: Togog is isolated from the other *Punokawan*, with the shortest paths that bridge Togog himself to other nodes often passing directly through non-*Punokawan* characters such as Sengkuni (*Korawa*) rather than through the *Punokawan* leader Semar (Figure 8B).

Both the *Korawa* and *Punokawan* communities are of similar size and interact preferentially with members of their own tribes to a similar extent. Their overall degrees are similar, on average, and they even show similar raw average closeness centrality values (Table 4). However, the *Korawa* have smaller mean node strength, and comparisons with the null model demonstrate that the co-occurrence patterns of the highest-profile *Korawa* (excluding the bridge-like Durna) leave them significantly more distant from the rest of the network than is expected despite their frequent appearances, as seen in Table 3 and Figure 5B for a rewired null model. Under a configuration null model, which de-emphasizes peripheral characters when counting shortest paths, the low-closeness outlier rankings of Duryudana, Sengkuni, and Karna rise even further to #1, #2, and #4, respectively. Meanwhile, the *Punokawan* are embedded more closely to the core of the network in accord with their frequent appearances, as reflected in their smaller mean closeness *p*-value (Table 4).

Figure 8. Inter-tribal bridges: Allocations of overall betweenness centrality among its inter-faction components for **(A)** Durna of the *Korawa* and **(B)** Togog of the *Punokawan* with respect to the *Tribe* partition.

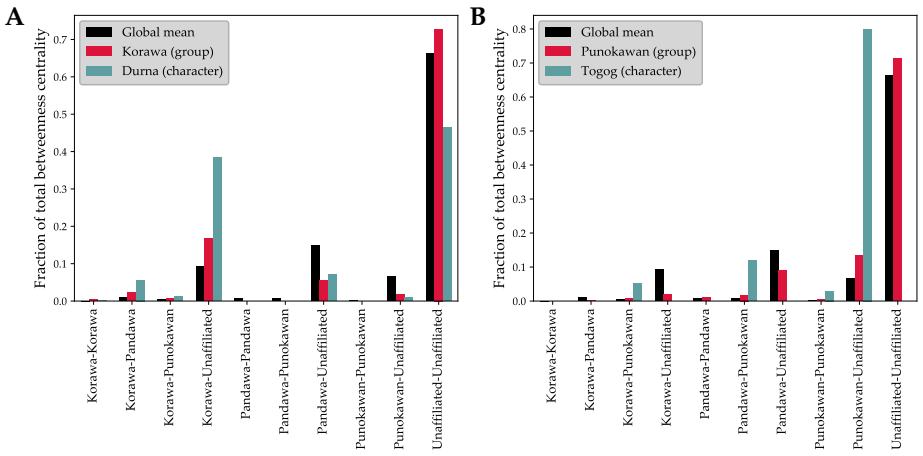


Table 4. Mean centrality values and p-values by faction for the *adegan* co-occurrence network with a rewired null model ensemble.

Faction	Mean degree	Mean node strength	Mean betweenness	Mean *p*-value (B.C.)	Mean closeness	Mean *p*-value (C.C.)
Female	3.77	15.35	.002	.7281	.779	.5406
Male	7.96	38.5	.018	.6627	.833	.6061
Korawa	18.71	69.29	.030	.6314	1.081	.7211
Pandawa	19.38	105.68	.069	.4030	1.127	.5005
Punokawan	17.83	101.83	.029	.6194	1.096	.6826
Unaffiliated	4.14	18.09	.006	.7197	.750	.5924
India	9.96	46.64	.025	.6050	.893	.5918
Java	4.01	19.8	.003	.7523	.746	.5926

3.5 Female characters often top adjusted betweenness rankings

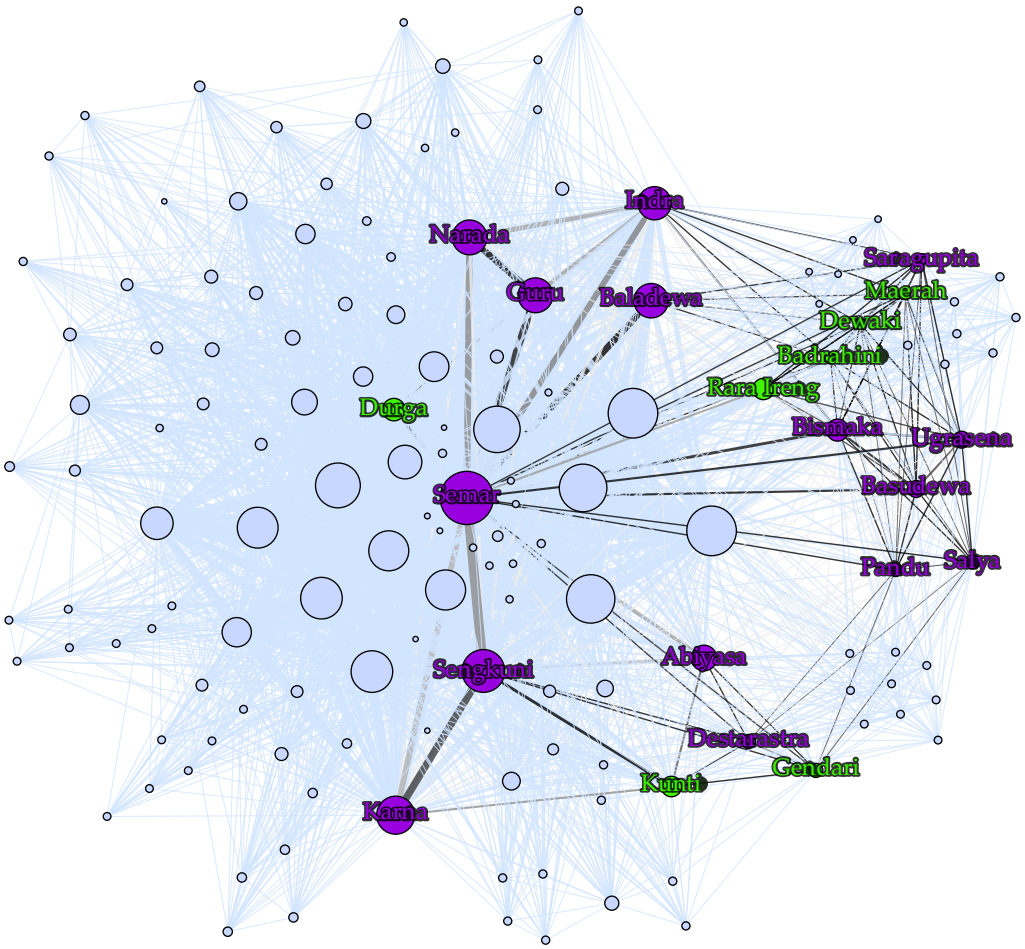
A cursory inspection of the outliers with high adjusted rankings of betweenness (Table 2) or its faction-specific variants also appears to reveal an unforeseen structural pattern related to another type of social faction: female characters often appear at or near the top of rankings. Rukmini (who ranks at #4 among overall high-betweenness outliers, #1 in Human-world betweenness, #1 in Male-Female betweenness, #1 in India-India betweenness, and #2 in *Pandawa*-world betweenness) and Durga (#5 among overall betweenness outliers, #1 in God-world betweenness, #1 in God-Human betweenness) are identified among the most bridge-like characters. A configuration null model, which tends to exclude low-strength characters in its shortest path counting, further elevates these rankings, placing Rukmini at #2 and Durga at #4 overall. These characters' bridge-like positions are readily apparent in the network visualization of Figure 3.

Although this trend does not generalize to all female characters, who on average show less extreme *p*-values than do male characters (Table 4), inspection of the shortest paths that explain these characters' betweenness centrality values reveals this to be a structural manifestation of the epic's depiction of these female characters' particular social roles. Rukmini is seen to act as a bridge between male and female characters (Figure 6B); the women she bridges to the predominately-male outside universe are those with whom she shares a domestic role within the kedhaton (palace). As a wife of the god Kresna, one of the most central characters in the *Mahabharata*, Rukmini's co-occurrences with her husband place her closely adjacent to his central position; meanwhile, the fact that she appears more frequently with Kresna's other wives than does Kresna himself places her on all of shortest paths joining those wives and all other characters via Kresna. Durga's high ranking similarly results from her adjacency to her high-profile husband, Guru, combined with her interactions with other characters who do not appear so frequently with Guru.

Under the coarser co-occurrence window of the lakon co-occurrence network, female characters Rara Ireng (#1 overall in a rewired null model), Kunti (#5 overall, and #1 in *Korawa*-world betweenness), and Gendari (#9 overall, #2 in *Korawa*-world betweenness) play bridge-like roles that are readily visible in force-directed network visualizations (Figure 9). Female characters like Arimbi (#1 in *Korawa-Pandawa* betweenness) and Banowati (#1 in *Korawa-Punokawan* betweenness) also top many of the faction-specific adjusted betweenness rankings. However, as the broader co-occurrence window results in more links for high-profile characters like Semar who tend to appear in many lakon, regardless of whether or not they actually encounter other characters directly within the story's scenes, shortest paths between lower-profile

characters who may interact more directly with one another are effectively short-circuited through these higher-profile characters.

Figure 9. *Lakon* co-occurrence network: Force-directed network visualization highlighting the highest-ranking betweenness centrality outliers according to a *configuration* null model, distinguishing Female (green) and Male (purple) outliers.



The high-betweenness outliers identified here thus cannot be explained in terms of their close adjacency to higher-profile characters, but instead tend to form bridges which do not traverse the core of the network. For example, Rara Ireng is the wife of the *Pandawa* hero Arjuna, but the shortest paths which explain her unexpectedly high betweenness do not involve Arjuna, but rather bridge characters who appear in a story depicting her own birth with other peripheral characters. Similarly, Kunti appears more often than any of her male consorts. Her highest-ranked link – which in fact has the second-highest raw betweenness

centrality of all links in the lakon co-occurrence network – does not correspond to a domestic relationship with a protagonist, but rather to an adversarial one with her enemy (and would-be suitor) Sengkuni of the *Korawa*. These examples emphasize the qualitative differences in perspective offered by different levels of co-occurrence resolution.

3.6 How deeply have Javanese characters been assimilated into the epic's co-occurrence structure?

Turning now to the relationship between characters' structural positions and historical origins, Figures 4C and 4D for the *adegan* network illustrate that characters with their historical origins in the Javanese wayang kulit incarnation of the *Mahabharata* tradition (plotted in blue, with *Punokawan* plotted in green) span a full range of closeness centrality p-values just as do characters of Indian origin (plotted in orange). This is echoed in faction mean node metrics (Table 4), which demonstrate that while characters from the Indian tradition have a much higher mean closeness value than do Javanese characters ($.89 > .75$), this appears to be almost entirely explained by their higher mean degrees and not by any higher-patterns of assortative co-occurrence by tribe, as both groups share almost identical mean closeness centrality p-values of $p \approx .59$. The same holds under the lakon co-occurrence window, where Indian-origin characters' higher mean closeness ($.96 > .74$) obscures the fact that both groups share almost-identical mean closeness p-values ($p \approx .51$). Although the *Punokawan* have the highest mean degrees and raw centrality values among Javanese characters by far, many non-*Punokawan* Javanese characters exceed null model expectations for closeness more than do the *Punokawan*, such that the overall mean p-value for Javanese characters is lower ($p = .59$) than that of the *Punokawan* alone ($p = .68$). That is, non-*Punokawan* Javanese tend to exceed the corresponding null model expectations to a greater extent than do the *Punokawan*, and overall, Javanese-origin characters conform to the null model expectations to a similar extent as do Indian-origin characters. Javanese characters, while not as ubiquitous throughout the stories as are characters from the original Indian canon, are not generally relegated to peripheral positions that would give them lower-than-expected closeness values any more than are Indian-origin characters, nor do they form detached co-occurrence "branches".

High-betweenness outliers, as ranked in Table 2 and highlighted in the network visualization of Figure 3B, include a number of Javanese-origin characters interspersed among bridge-like network features at the adegan level. Attempting to locate any hidden clusters of Indian- and Javanese-origin clusters by examining rankings of India-Java betweenness centrality outliers, we find that this faction-specific betweenness almost exactly reproduces the rankings of Table 2. However, in the adjusted Java-Java intra-faction betweenness rankings, which excludes India-Java paths, the *Korawa* bridge Durna drops in rank from

#3 to #24. The dependence of his betweenness on paths between characters of different historical origins reflects his role as a bridge between the *Korawa*, a community of homogenously Indian origin, to the rest of the universe. Observations such as these suggest that the observed higher-than-expected modularity along the lines of historical origin (Table 1) seems to be largely explained by the homogeneiety of historical origins within the epic's two most tight-knit tribal communities. While non-*Punokawan* Javanese characters are thoroughly interspersed throughout the network, the isolation of the *Korawa* in still preserved. Mean node allocation profiles for an alternative partition of the nodes, which distinguishes characters by origin but with *Korawa* and *Punokawan* communities considered separately, demonstrates that there is relatively little assortative co-occurrence based on historical origin, except insofar as it occurs along tribal community lines (Figure 2B). Javanese- and Indian-origin characters have apparently come to be thoroughly inter-mixed with one another throughout the network, but this has occurred in such a way that the *Korawa* remain structurally isolated.

4 Conclusion

By examining one representative version of the Javanese *wayang kulit Mahabharata* canon, we have presented an initial proof-of-concept demonstrating how several issues of interest in the cultural and historical study of this and other related art forms can be approached in network-theoretical terms. The many intangible aspects of *wayang*'s appeal certainly cannot be quantified. Still, without purporting to replace existing approaches towards understanding this tradition, a network-theoretical perspective can provide concrete, meaningful information about the mechanics of a particular work of fiction to complement traditional perspectives.

The results presented here are perhaps most interesting if considered as a first step towards comparative studies involving multiple tellings of the epic, or other works of fiction. For example, since network modularity can be used to quantify the extent to which the storylines are advanced through depictions of in-group interactions rather than inter-faction encounters, it could also allow us to quantitatively compare how distinct retellings – representing different regional traditions, or distinct performances by multiple *dhalang* – differ in their depictions. If null model analyses of betweenness centrality can indeed reveal some structural hallmarks of the gender roles depicted in *wayang kulit*, then the relative exaggeration or de-emphasis of these features in networks representing distinct retellings of the epic could help to inform discussions of the broader cultural differences which may have shaped their different structures. If network-theoretical tools like closeness centrality can quantify the extent to which Javanese characters have been assimilated into central roles within the epic, they could also potentially be used to quantitatively track the incorporation of those elements into the epic over time.

These types of comparative or dynamical network analyses would require the preparation of data sets of sufficient resolution and scope to represent mutiple incarnations of the epic. In most cases, this would not be a straightforward matter. For example, attempts to compare Indian and Javanese *Mahabharata* traditions using co-occurrence networks would be complicated by the fact that many incarnations of the epic are not organized into episodic subdivisions of comparable resolution. Our ability to compare co-occurrence networks representing multiple stages of the historical evolution of the *wayang kulit* tradition, throughout which various Javanese innovations were assimilated into the originally Indian canon over time, is highly contingent on the availability of suitable historical sources. But if these obstacles can be creatively overcome, co-occurrence networks may serve as a useful new tool with which to more fully understand the rich diversity and complexity which makes the *wayang kulit* tradition so fascinating, and the processes which have shaped it throughout its centuries-long history.

References

Arps, Bernard. *Tall Tree, Nest of the Wind: The Javanese Shadow-play Dewa Ruci Performed by Ki Anom Soeroto: A Study in Performance Philology*. NUS Press, 2016.

Brandon, James R, Pandam Guritno, and Roger A Long. *On thrones of gold: Three Javanese shadow plays*. University of Hawaii Press, 1993.

Choi, Yeon-Mu, and Hyun-Joo Kim. "A directed network of Greek and Roman mythology." *Physica A: Statistical Mechanics and its Applications* 382, no. 2 (2007): 665–671.

Chung, Fan, and Linyuan Lu. "The average distance in a random graph with given expected degrees." *Internet Mathematics* 1, no. 1 (2004): 91–113.

Clara van Groenendael, Victoria M. *Wayang Theatre in Indonesia: An Annotated Bibliography*. Foris Publications, 1987.

Cimini, Giulio, Tiziano Squartini, Andrea Gabrielli, and Diego Garlaschelli. "Estimating topological properties of weighted networks from limited information." *Physical Review E* 92, no. 4 (2015): 040802.

Escobar Varela, Miguel. *Digital Wayang Encyclopedia*. https://villaorlado.github.io/wayangnet works/html/.

Feinstein, Alan H. *Lakon carangan: Ringkasan lakon*. Vol. 3. Akademi Seni Karawitan Indonesia, 1986.

Fischer, Frank, Mathias Göbel, Dario Kampkaspar, Christopher Kittel, and Peer Trilcke. "Network Dynamics, Plot Analysis: Approaching the Progressive Structuration of Literary Texts." *Digital Humanities 2017 (Montréal, 8–11 August 2017). Book of Abstracts*. McGill University, 2017.

Freeman, Linton C. "A set of measures of centrality based on betweenness." *Sociometry*, 1977, 35–41.

——— "Centrality in social networks: I. Conceptual clarification." *Social networks* 1, no. 3 (1978): 215–239.

Hennemann, Stefan, and Ben Derudder. "An alternative approach to the calculation and analysis of connectivity in the world city network." *Environment and Planning B: Planning and Design* 41, no. 3 (2014): 392–412.

Kats, Jacob. *De wajang poerwa: een vorm van Javaans toneel. Vol. 1*. Foris Publications, 1923.

Mastromatteo, Iacopo, Elia Zarinelli, and Matteo Marsili. "Reconstruction of financial networks for robust estimation of systemic risk." *Journal of Statistical Mechanics: Theory and Experiment* 2012, no. 03 (2012): P03011.

Moretti, Franco. *Graphs, maps, trees: abstract models for a literary history*. Verso, 2005.

Mulyono, Sri. *Wayang, asal-usul, filsafat dan masa depannya*. Gunung Agung, 1978.

Newman, Mark E. J. *Networks: An Introduction*. Oxford University Press, 2010.

——— "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality." *Physical Review E* 64, no. 1 (2001): 016132.

Newman, Mark E.J., and Michelle Girvan. "Finding and evaluating community structure in networks." *Physical Review E* 69, no. 2 (2004): 026113.

Onnela, Jukka-Pekka, Jari Saramäki, János Kertész, and Kimmo Kaski. "Intensity and coherence of motifs in weighted complex networks." *Physical Review E* 71, no. 6 (2005): 065103.

Park, Gyeong-Mi, Sung-Hwan Kim, and Hwan-Gue Cho. "Structural analysis on social network constructed from characters in literature texts." *Journal of Computers* 8, no. 9 (2013): 2442–2447.

Park, Seung-Bo, Kyeong-Jin Oh, and Geun-Sik Jo. "Social network analysis in a movie using character-net." *Multimedia Tools and Applications* 59, no. 2 (2012): 601–627.

Purwadi. *Kempalan Balungan Lakon Wayang Purwa [Compendium of Classical Wayang Plot Scenarios]*. Cendrawasih, 2008.

Sears, Laurie Jo. *Shadows of empire: Colonial discourse and Javanese tales*. Duke University Press, 1996.

Trilcke, Peer, Frank Fischer, and Dario Kampkaspar. "Digital network analysis of dramatic texts." *Proceedings of the Digital Humanities* 6, no. 7 (2015): 8.

Watts, Duncan J, and Steven H Strogatz. "Collective dynamics of 'small-world' networks." *Nature* 393, no. 6684 (1998): 440.

Waumans, Michaël C, Thibaut Nicodème, and Hugues Bersini. "Topology analysis of social networks extracted from literature." *PloS ONE* 10, no. 6 (2015): e0126470.

Xanthos, Aris, Isaac Pante, Yannick Rochat, and Martin Grandjean. "Visualising the dynamics of character networks." *Digital Humanities*, 417–419. 2016.

MINAKOWSKI, MAREK JERZY

Family Network of an Emerging Jewish Intelligentsia (Cracow, 1850-1918)

Journal of Historical Network Research 2 (2018) 53-75.

Keywords

Jewish intelligentsia, mass genealogy, vocational studies, Jewish studies

Abstract

Mass-genealogical research of the Jewish community in Cracow in 19th and early 20th century provides extraordinary opportunity to investigate the process of the emergence of a new social class: society of traditional merchants and peddlers produced modern attorneys and doctors. Now we are able to capture the dynamics of the process. For many reasons, the Jewish community in Cracow is an outstanding specimen. About 1900 there were about 25,000 Jews in Cracow (about 28% of city population). Almost whole family structure of the population has been revealed and analysed as a connected network. Over 1200 nodes of the network have been identified as the Jagiellonian University students between 1850 and 1918. We know what and when they were studying and often in which house they were born, what was their family social status etc. This data is used to model several key features of new emergent social class: what was the impact of parents' families on the choice of university education and the choice influence on a future marriage. Especially interesting are results about influence of family status on the completion of a doctoral qualification which, in turn, provided basis for discussion of best formula to describe how this influence is spreading.

1 Introduction*

Somewhere between 1860 and 1920 there was a significant shift in the societies of Central and Eastern Europe.[1] The dominant position of landed nobility was replaced by the rule of intelligentsia.[2] Let us look into formation of specific group: Jewish intelligentsia in the former Free City of Cracow, a quasi-independent country created by the Vienna Congress in 1814 and in 1846 incorporated to Austria as Grand Duchy of Cracow. Cracow and especially its Jewish community is a unique place to analyse. It was the old capital of Polish-Lithuanian Commonwealth – a country which was a home for most of European Jews until the end of 18th century. The Jewish community thrived here continuously since Middle Ages until 1939, in the same district of the city and known for their conservatism. In 1860 they solicited Shimon Sofer (Schreiber), to become the chief rabbi of Cracow. Sofer, who held office until his death in 1883 (and passed his office to son-in-law), was a son of Moses Schreiber (Chasam Sofer, the one who coined the motto of the ultra-orthodox Jews (Haredim): "'new' is forbidden by the Torah".[3]

But the "new" happened. Cracow was a home to one of the oldest universities in Europe, the Jagiellonian University (established 1364). The University was attractive to the Cracow Jews because provided good education while other possible destinations (universities in Vienna, Budapest, Breslau, Warsaw and Lvov) were all about 300 kilometers far from Cracow. Among 28,873 students who enrolled at the Jagiellonian University between 1850 and 1918, there were 3,955 Jews (13.7%). Among the University's 7,280 graduates (who re-

---

* **Acknowledgements:** The author thanks Prof. Daniel S. Hirschberg from University of California, Irvine, for creating and making available online his database of „Early Family Trees" of Cracow Jewish families at https://www.ics.uci.edu/~dan/genealogy/Krakow/ - it was of tremendous help in reconstruction of the family network and, most importantly, it showed that it was possible at all.

Corresponding author: Dr Marek Jerzy Minakowski, Dr Minakowski Publikacje Elektroniczne, ul. Komandosów 5/10, 30-334 Kraków, Poland, e-mail: mj@minakowski.pl ORCID: https://orcid.org/0000-0001-5677-9277

1 Rafał Smoczyński and Tomasz Zarycki, Totem inteligencki. Arystokracja, szlachta i ziemiaństwo w polskiej przestrzeni społecznej (Warszawa: Wydawnictwo Naukowe Scholar, 2017).

2 Jerzy Jedlicki, Maciej Janowski, and Magdalena Micińska, Dzieje Inteligencji Polskiej Do Roku 1918 (Warszawa: Instytut Historii PAN, Wydawnictwo Neriton, 2008).

3 David Ellenson and Daniel Gordis, Pledges of Jewish Allegiance: Conversion, Law, and Policymaking in Nineteenth- and Twentieth-Century Orthodox Responsa (Stanford University Press, 2012), 70.

ceived title of doctor[4]) there were 1,760 categorized as "of Mosaic faith", i.e. Jews (24.2%)[5] Among those students, 1,080 were matched to birth certificates in the records of Cracow Jewish community. If we only count those who were born between 1881 and 1900, we have 671 such Cracow-born Jewish students. Only 57 of those 671 had a father who was also a university student – for over 90% of them being a student was the infamous "new". The 671 represent 3.7% of all 18,025 Jewish births registered in Cracow during 1881-1900.[6]

In our analysis we will employ the framework of mass genealogy. Mass genealogy was first conceptualized by M.J. Minakowski[7] and presented (in English) during EUSN2017 in Mainz.[8] In short, it is a discipline on the crossroads between (classic) genealogy, historical demography and social network analysis. It analyses whole societies (as demography) as graphs (as SNA) of people connected by genealogical ties (as genealogy). The main difference between mass genealogy and "ordinary" (or "historical") genealogy (auxiliary science of history) is that while "ordinary" genealogy consider family ties as properties of individuals (e.g., in order to discuss people's biographies and their family circles), mass genealogy looks from the network perspective: it considers people as nodes in a massive social network (several hundred thousand people at once) and focuses on society (considered as multi-generational network of people connected by family links) and not on individuals who

4 In the discussed period the title granted after successfully completed studies was "doctor", shortened as "Dr." which upon graduation was first part of the name, in the form: "Dr. Samuel Garfunkel" (in civil records, censuses etc.). The title was awarded by faculty and Jewish students became Doctor of Law, medicine or philosophy. Our analysis sometimes sounds clumsy because it is important not to mix it with ordinary English meaning of doctor (that is, doctor of medicine) or PhD (which means literally "Doctor of Philosophy" while philosophers were in minority here).

5 Own calculations based on: Corpus studiosorum Universitatis Iagellonicae 1850-1918 (Archiwum Uniwersytetu Jagiellońskiego, 1999). Jews were not considered a nationality; official censuses in Austrian Empire did not even allow for providing Yiddish as language (Jews were entering Polish or German instead).

6 Own calculations based on: "Akta Stanu Cywilnego Izraelickiego Okręgu Metrykalnego w Krakowie" (n.d.), 29/1472/0, Archiwum Narodowe w Krakowie, https://szukajwarchiwach.pl/29/1472/0/1/str/1/1000#tabJednostki.

7 Marek Jerzy Minakowski, "Modelowanie rozkładu kapitału ekonomiczno-społeczno-kulturowego poprzez genealogię masową," Przeszłość Demograficzna Polski 38 (2016): 63–88, https://doi.org/10.18276/pdp.2016.4.38-03.

8 Marek Jerzy Minakowski, "Mass Genealogy: Top 1% of 19-Th Century Polish Society as a Single Family Network (PageRank-like Analysis)" (Third European Conference on Social Networks EUSN 2017, Mainz, Germany, 2017), https://www.eusn2017.uni-mainz.de/files/2016/08/EUSN2017_Booklet_25_09.pdf.

happen to be networked with other individuals (in order to create a family tree with a few dozen people so that it could be presented in one, readable chart).[9]

The analysed database consists of 83,000 people, out of whom 64,000 were born in 19th century. It is a single, connected graph (the giant component of a larger graph of 102,000 people)[10]. To build it, we analysed all 10,433 marriage records from the Jewish Registry of Births, Deaths and Marriage in Cracow (Izraelicki Okręg Metrykalny Kraków, literally: Israelite Metrical District Cracow) from 1811 to 1914 (they are available in the National Archives in Cracow) and a considerable part of birth records (all records 1856-1876 and 1891-95 and selected records from other years). Unfortunately, most of Jews in the analysed period did not register their marriages, but fortunately – most of those who enrolled at University, were born in registered marriages (in their birth records they are marked as legitimate offspring).[11]

Usually, in a marriage record, the details of six people were included: both spouses and their parents. The birth record format was varying: initially it contained three people (child and its parents) but since 1868 mother's father begins to be mentioned and since 1873 also her mother, which makes five people in each birth record. In the years 1817-1900 alone, there were 55,300 births recorded in the district. We were not able to closely analyse all those birth records but already most (28,800) of the children born within period 1817-1900 have been placed in the aforementioned giant component of the graph (for the children born between 1856 and 1876 it is even 75.7%).

---

9 Further methodological discussion of mass genealogy is published in: Marek Jerzy Minakowski, "Genealogia Masowa – Metodologia Tworzenia i Publikacji Bazy Danych," in Edytorstwo Wobec Masowości Źródeł Najnowszych, vol. 3, Edytorstwo Źródeł XIX i XX Wieku, Teoria i Praktyka (Warszawa: Wydawnictwa Uniwersytetu Warszawskiego, 2018).

10 Network science terms are used according to their definitions in: Albert-László Barabási and Márton Pósfai, Network Science, 1st edition (Cambridge, United Kingdom: Cambridge University Press, 2016).

11 In the period 1891-1895, 4,555 children were born in Israelite Metrical District in Cracow. For 4,505 we know whether they were born legitimately or not. For 229 we know that they studied in Jagiellonian University before 1918. The results are: 75% of future students were born legitimately while only 36% of non-students were born legitimately. A detailed analysis of legitimate and illegitimate births in 19th century Cracow Jewish Community was presented in Marek Jerzy Minakowski, "Jews of 19th-Century Cracow in Civil Registers" (The XIth Congress of the European Association for Jewish Studies, Kraków: Fundacja Alef dla Rozwoju Stiudiów Żydowskich, 2018). The extended version is to be published in the next volume of Przeszłość Demograficzna Polski as Marek Jerzy Minakowski and Anna Lebet-Minakowska, "Legitimate, Legitimized and Illegitimate Families of 19th-Century Cracow Jews," Przeszłość Demograficzna Polski 40 (2018).

In our database people can also be connected to places. We have access to data from censuses which were held every ten years (1857, 1870, 1880, 1890, 1900, 1910 and 1921). We did not manage to collect all data from all censuses, but we focused on the last available census, from 1921.[12] Out of the 1,080 Jagiellonian University students whose birth records we found among Cracow Jews, 465 were found in the 1921 census. From the 1921 census we collected all the data about families of former students, their parents, the families of their siblings and many others. The census provides the detailed address (house number which we converted to coordinates) and such information like a profession (means of living), religious denomination, nationality and language.

From the birth records 1855-59 and 1891-95 we were able to extract data about all Jewish birth places that occurred within these periods (house numbers converted to coordinates with accuracy of about 20 meters).

The last dimension to be included in our database was the families' social standing. To measure this, we used the 1883 electoral roll.[13] Jews in Cracow elected their local authorities according to the tax list; we found and analysed the 1883 list of all 565 Jewish families who were eligible to vote, with the sum of the local tax they had to pay. We managed to identify most of them. Knowing the genealogy of the whole community, we were able to calculate the relative status of everybody (assuming that people who had many rich uncles were from "better" family than people who had little number of affluent cousins).

The above database has been published at Wielcy.pl;[14] it can be browsed and individual records can be analysed. Now we are going to show some interesting results found in the graph as a whole. We believe that many other interesting results can be found and please consider it an invitation to further research.

2 Size and shape of the students' network

The preceding discussion concerned the whole Jewish society of Cracow. Let us focus our attention on the issue of intelligentsia itself. The term "intelligentsia" as a social class can be operationalized in many ways[15] but for the sake of

12 "Spis Ludności Miasta Krakowa z r. 1921" (n.d.), 29/91/0, Archiwum Narodowe w Krakowie, https://szukajwarchiwach.pl/29/91/0/#tabZespol.

13 "Spis Członków Zboru Izrelickiego w Krakowie" (1883), 29/33/0 Akta miasta Krakowa Series: 3.2.3 Akta Magistratu w układzie rzeczowym. File/unit: Kr 7055. Classification scheme code: fasc. 29 konw. 003112/IV/1920, Archiwum Narodowe w Krakowie.

14 Marek Jerzy Minakowski, "Wielka Genealogia Minakowskiego," 2018, http://wielcy.pl/.

15 Jedlicki, Janowski, and Micińska, Dzieje Inteligencji Polskiej Do Roku 1918.

current discussion let us operationalize it as a group of people who were enrolled to the Jagiellonian University in Cracow and their close families. Considering the family structure as a graph, let us consider "close family" as people who are connected by a path no longer than five links (A, B) such that A is a child, parent, sibling or a spouse of B. A path of five will be, e.g., a father-in-law of one's first cousin (i.e., the father of a wife of a son of a brother of one's father) or son-in-law of one's wife's niece (husband of a daughter of a daughter of a sister of one's wife). The value "five" seems to be a limit of what people can consider their families (people who can be met during family events organized by somebody located on the path between them). The choice of "five" was arbitrary and should be considered rather an upper limit, aiming to eliminate ties that are not *family* ties.

Therefore, for further discussion, let us select from our network all paths (A … B) such that: (a) the path is composed of no more than five links, (b) each link is a relation of being a child, parent, sibling or a spouse, (c) both A and B were enrolled as students of the Jagiellonian University between 1850 and 1918; strictly speaking: A and B are listed in the *Corpus Studiosorum Universitatis Iagellonicae*[16], and (d) both A and B: either have a birth record in the Israelite Metrical Department of Cracow or their parents have a marriage record there or any of their parents have birth certificate there.

The sum of the paths (all nodes and links between them) forms a network. In the giant component of the network there are 5,905 people, each connected to everybody else. Within the network, 1,236 were enrolled as students at the Jagiellonian University between 1850 and 1918. The other 4,669 people are at a distance of no longer than three links (edges) to some student.

For 925 (75%) of the 1236 Jewish students there is a birth record in Israelite Metrical Department of Cracow. The total number of Jewish students of this period who have birth record in Israelite Metrical Department of Cracow was 1,080, so these in our network make 86% of them. For the next 136 there is a birth record of any of their parents or their parents' marriage record. If we count all students who either have birth certificate here or any of their parents have birth certificate here or their parents have marriage certificate here the numbers are: 1061 are in the giant component in our graph, which is 86% of all 1,236 Jewish students in our graph (some are from outside) and also 86% of all 1,230 students who satisfy the above conditions for birth or marriage certificate (some are outside of the giant component).

16 Corpus studiosorum Universitatis Iagellonicae 1850-1918.

Faculty	Jewish students in the network		Percentage for non-Jewish students
	number	percentage	
Law	611	49.4%	40.7%
Medicine	340	27.5%	16.1%
Philosophy	263	21.3%	36.9%
Theology	0	0.0%	5.0%
No data / other	22	0.6%	0.5%
Total	1236	100.0%	100.0%

Table 1. The faculty of the Jagiellonian University where students enrolled to; in the cases where a student changed faculty, only first enrolment is counted. All students in the analysed network of Jewish students from Cracow connected by family ties. Non-Jewish students according to volumes 7-8 of the *Corpus Studiosorum Universitatis Iagellonicae* (total number in these volumes: 7,620, non-Jewish 6,660). Own calculations.

The shape of the network can tell us whether the Jewish intelligentsia in Cracow was composed of some distinct communities. We could imagine, for instance, that there were parts originating from different religious groups or from distinct districts of the city. It could have happened also that there was some counter-elite, as among the readers circle of 19th century *Kurier Warszawski*[17] where two groups of Catholics (one from old aristocracy and other descendants of Frankist sect) formed two opposed elites of Warsaw intelligentsia in the middle of 19th century.

First approach was to visualize the network in Gephi[18] using the ForceAtlas2 layout[19] and then apply the Louvain modularity algorithm for community detection[20] several times to see whether there are some outstanding communities. This method did not provide any meaningful results: the calculated communities did not stand out, they were mixing and overlapping.

17 Marek Jerzy Minakowski, "Sieć społeczna wokół Kuriera Warszawskiego na podstawie jego nekrologów z lat 1821–1861," Przeszłość Demograficzna Polski 39 (2017): 209–51, https://doi.org/10.18276/pdp.2017.39-09.

18 "Gephi - The Open Graph Viz Platform," accessed August 13, 2018, https://gephi.org/.

19 Mathieu Jacomy et al., "ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software," PLOS ONE 9, no. 6 (June 10, 2014): e98679, https://doi.org/10.1371/journal.pone.0098679.

20 Vincent D. Blondel et al., "Fast Unfolding of Communities in Large Networks," Journal of Statistical Mechanics: Theory and Experiment 2008, no. 10 (October 9, 2008): P10008, https://doi.org/10.1088/1742-5468/2008/10/P10008.

The other try was to apply the Louvain method ten times with the resolution = 15. This resulted in splitting all 1,236 students into 5–8 separate communities (five times: five communities, three times six communities, once seven and once eight communities). The largest community size was at least 316 or 26% (median: 450 or 37%). The question was: whether these communities are consistent and reflect the real divisions within the society or are just an accidental artefact of used method (which had to return something even if the result was to be artificial). The number of nodes that always (ten times) belonged to the largest community was 0. The sizes of largest sets of nodes that always kept together were, in decreasing order: 39, 37, 22, 21 and 20. If the conditions were to be relaxed and instead of ten subsequent application of the method, five applications were to be performed in two separate experiments, the number of nodes that five times belonged to largest community was either 36 or 13 and the sizes of largest sets of nodes were, in decreasing order, either 44, 40, 36, 35, 32 or 70, 50, 46, 41, 32. Therefore we can conclude that there were no deep, real divisions within the analysed society that could be detected.

For measuring modularity of a partition (a scalar value between -1 and 1 that measures the density of links inside communities as compared to links between communities), the Neuman's algorithm was employed[21]; for the above calculations of Louvain method with resolution = 15, it produced scores for modularity from 0.61 to 0.77 (average: 0.70, median: 0.69). When resolution was set to 1.0, modularity score was varying from 0.913 to 0.916 but the number of detected communities was much larger: it was from 63 to 67. Each of the smaller communities have between 6 and 41 students; 85% of the communities had between 10 and 30 students.

It seems, then, that the whole social graph does not exhibit visible structural divisions. Instead, we may assume that the whole society of the Jewish intelligentsia in Cracow between 1850 and 1918 was composed of about 65 families of about 15-20 students or alumni of the Jagiellonian University, closely connected within themselves and more distantly connected with other such families.[22]

---

21 M. E. J. Newman, "Modularity and Community Structure in Networks," Proceedings of the National Academy of Sciences 103, no. 23 (June 6, 2006): 8577–82, https://doi.org/10.1073/pnas.0601602103.

22 Some connections between families were stronger than another. Such families can be merged; e.g., we can also assume resolution = 3 and find about 30 communities/families with modularity of about 0.905.

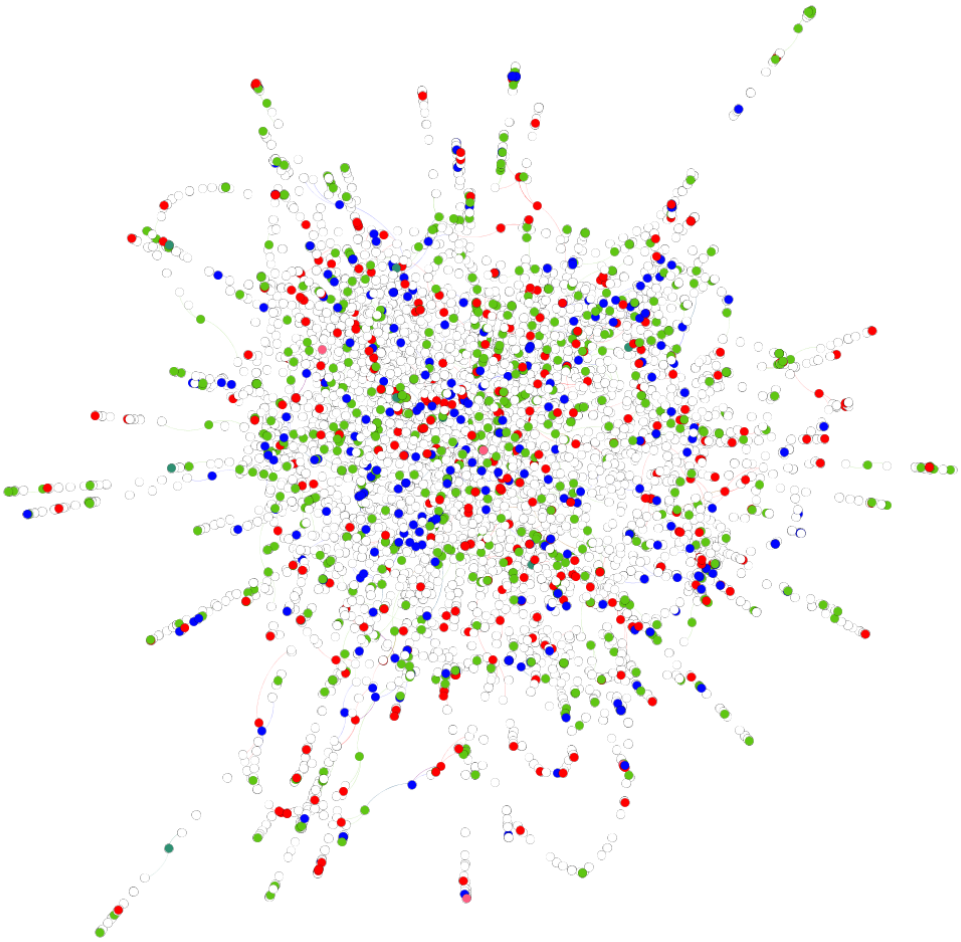


Figure 1. The network of the Jewish intelligentsia (students and links between them). ForceAtlas2 in Gephi. Law: green; medicine: red; philosophy: blue; non-students: grey.

Not only the structure, but also quality of the communities can be important. We could hypothetically assume that there were families specializing in medicine, families of lawyers and families of teachers (studying philosophy, i.e., all sciences and humanities except law or medicine). But the close look into the network shows that there are no visible distinctions between faculties they chose.

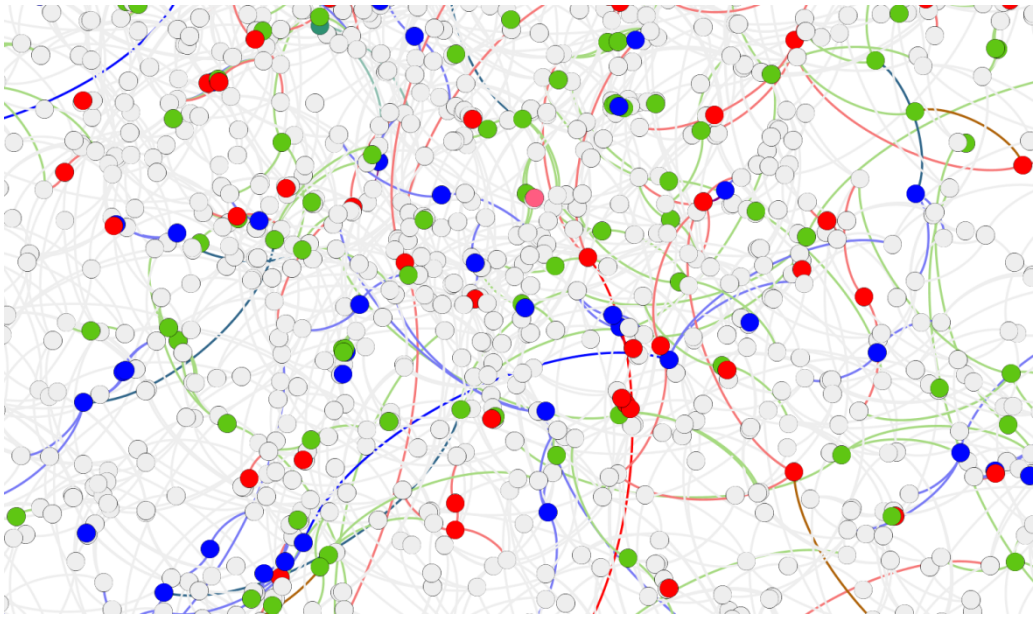


Figure 2. The network of Jewish intelligentsia (students and links between them). ForceAtlas2 in Gephi. Zoom showing internal structure. Law: green; medicine: red; philosophy: blue; non-students: grey.

At least, there are no visible sign. But this may be misleading. Actually, the author was strongly confident that there must be some regularities, that people were sharing professions (legal, medical, teaching) in families, from father to son or between cousins. Also, the idea of marital homophily (homogamy) was very compelling, claiming that "birds of a feather flock together" and people tended to marry people of similar interests and similar profession.[23] Yet another argument could be found in the institution of arranged marriage, still widespread in Jewish societies in this part of Europe those times: it is enough to remind that the novel S. Aleichem's *Tevye the Dairyman*[24] was first published in 1894, known better as *Fiddler on the Roof* (Broadway musical and Hollywood film).

The data was prepared in the following way: we took all 1226 students of Jagiellonian University between 1850 and 1918 who either have birth record in Israelite Metrical District of Cracow or any of their parents have it or their parents married there. Then we found the shortest paths between them, no longer than five links, where each links is one of following, directed relationships:

23 Henryk Domański and Dariusz Przybysz, Homogamia małżeńska a hierarchie społeczne (Warszawa: Wydawnictwo IFiS PAN, 2007).

24 Sholem Aleichem, Tevye the Dairyman and Motl the Cantor's Son, trans. Aliza Shevrin, New edition (New York, N.Y: Penguin Classics, 2009).

being child, father, mother, husband, wife or a sibling. We found 17,456 such pairs, between 1,124 students (102 had no pair: the shortest path to another student was longer than five or was infinite).

For each path its length was recorded and the side from which it begins. "Side" is the attribute of first link in the path (child, father, mother, husband, wife or a sibling). The paths are the shortest ones, so if a path is labelled as "father side", it means that the person on the end of the path is closer to one's father than to one's mother, one's sibling, one's children or one's spouse(s). Thus, for instance, the path to one's father-in-law was labelled "spouse side" while path to one's grandchildren was be labelled "child side".

Out of the 1,124 students, for 1,110 we know the faculty they enrolled as one of the three: law, medicine and philosophy (for four of them the faculty was not recorded while ten of them enrolled to lately opened faculties of pharmacy and agriculture). For the 1,110, the distribution was following:

Faculty	**Number**	**Percentage**	**Confidence interval**
Both men and women			
Law	571	51%	49% – 54%
Medicine	298	27%	24% – 30%
Philosophy	241	22%	19% – 25%
Total above	1110	100%	
Men only			
Law	571	66%	63% – 69%
Medicine	221	26%	23% – 28%
Philosophy	72	8%	6% – 10%
Total men	864		
Women only			
Law	0	0%	0% – 0%
Medicine	77	31%	26% – 38%
Philosophy	169	69%	73% – 79%
Total women	246		

Table 2. Faculties where Jewish students from Cracow enrolled first. Absolute number, probability and confidence intervals. Own calculations.

The confidence intervals (higher and lower limits of confidence) in the Table 2 and in following discussion are calculated according to the formula:

$$\mathrm{CI} = a \pm 1.96 \times \sqrt{\frac{a \times (1-a)}{n}}$$

Formula 1: Confidence interval. *a* is the rate of all paths ending with a student of the given faculty and *n* is the number of all pairs (shortest paths) satisfying the tested conditions. The value of *n* in most cases is above 100, so the normal distribution was assumed (which explains why for 0.95 significance level the 1.96 factor was used).

The differences between male and female students are so great that they had to be analysed separately. One thing is that women were not allowed to study law; but this was an obvious trap that could be avoided. More important difference was in the proportion of female students of medicine and philosophy: while for men the proportion was 1.23:1, for women it was 0.46:1. That is why many differences resulting from sex could be falsely attributed to the faculty.

The tests were performed as follows: all students have been divided according to the faculty (law, medicine or philosophy). Then, for each group all other students from their families (satisfying the tested conditions) were found and the probability that they enrolled into the same faculty has been calculated. For instance, for all lawyers we took all people from their families with path no longer than three; the probability that they were also lawyers was 54%: there were 1,312 such paths out of 2,439 paths from a lawyer to somebody else with distance no longer than three. In this case, the 95% confidence interval was between 52% and 56%. In general, the confidence interval for being a lawyer was between 49% and 54% (see Table 2), so we could not refute zero-hypothesis: the average member of (such defined) family of a lawyer had a little greater possibility of studying a law (instead of medicine or philosophy), but the difference could also be due to an accident.

The tests were taken in cases where *n* was enough to provide any meaningful results for all these three faculties. For the network distance of no longer than five (as defined above), the results are presented in the Table 3.

Faculty	Confidence interval	N	Reference values	Notes
all sides, distance ≤5:				
law	66%–68%	6946	63%–69%	
medicine	24%–27%	2633	23%–28%	
philosophy	5%–9%	849	6%–10%	
father's side, distance ≤5:				
law	63%–67%	2233	63%–69%	
medicine	23%–29%	808	23%–28%	
philosophy	5%–12%	225	6%–10%	
mother's side, distance ≤5:				
law	67%–71%	2360	63%–69%	
medicine	23%–29%	774	23%–28%	
philosophy	2%–7%	233	6%–10%	
wife's side, distance ≤5:				
law	62%–70%	595	63%–69%	
medicine	17%–26%	328	23%–28%	
philosophy	-1%–5%	98	6%–10%	Below
sibling's side, distance ≤5:				
law	65%–70%	1658	63%–69%	
medicine	22%–28%	652	23%–28%	
philosophy	6%–13%	277	6%–10%	
all sides, distance ≤ 4:				
law	65%–69%	3388	63%–69%	
medicine	22%–27%	1254	23%–28%	
philosophy	8%–14%	428	6%–10%	
all sides, distance ≤ 3:				
law	66%–70%	1928	63%–69%	
medicine	21%–27%	682	23%–28%	
philosophy	9%–18%	230	6%–10%	
all sides, distance ≤ 2:				
law	64%–70%	886	63%–69%	
medicine	17%–26%	329	23%–28%	
philosophy	8%–21%	113	6%–10%	
sibling's side, distance ≤ 4				
law	65%–71%	1042	63%–69%	
medicine	16%–24%	377	23%–28%	Low
philosophy	7%–17%	172	6%–10%	
sibling's side, distance ≤ 3				

law	66%–73%	753	63%–69%	
medicine	13%–22%	276	23%–28%	Below
philosophy	8%–22%	107	6%–10%	
sibling's side, distance ≤ 2				
law	64%–72%	574	63%–69%	
medicine	11%–20%	212	23%–28%	Below
philosophy	7%–24%	72	6%–10%	

Table 3. Confidence intervals for proportions of Jewish students from Cracow (1850-1918) that were related to other students of the same sex who enrolled in the same faculty. Males related to males. Various sides and distances tested. N: number of all other students (from any faculty) related to a student of this faculty. Own calculations.

What we can see in the Table 3, in most cases there are no significant results: the sons of Jewish families had rather equal tendency to study in various faculties. The irregularities can be found in (a) philosophers and their wives' families and in (b) students of medicine and their brothers and families of their brothers and sisters. We will discuss it later, after looking at the Table 4, where results for female students are shown.

Faculty	**Confidence interval**	**N**	**Reference values**	**Notes**
all sides, distance ≤5:				
medicine	26%–38%	234	26%–37%	
philosophy	73%–79%	656	63%–74%	Low
father's side, distance ≤5:				
medicine	32%–57%	61	26%–37%	
philosophy	74%–85%	206	63%–74%	Low
mother's side, distance ≤5:				
medicine	18%–34%	112	26%–37%	
philosophy	62%–74%	223	63%–74%	
sibling's side, distance ≤5:				
medicine	25%–52%	49	26%–37%	
philosophy	75%–88%	146	63%–74%	Over
sister, distance ≤5:				
medicine	35%–73%	26	26%–37%	High
philosophy	70%–90%	60	63%–74%	
all sides, distance ≤4:				
medicine	33%–52%	99	26%–37%	High

philosophy	76%–85%	297	63%–74%	Over
all sides, distance ≤3:				
medicine	38%–64%	59	26%–37%	Over
philosophy	80%–90%	189	63%–74%	Over
all sides, distance ≤2:				
medicine	35%–71%	30	26%–37%	High
philosophy	80%–93%	106	63%–74%	Over

Table 4. Confidence intervals for proportions of Jewish students from Cracow (1850-1918) that were related to other students of the same sex who enrolled in the same faculty. Females related to females. Various sides and distances tested. N: number of all other students (from any faculty) related to a student of this faculty. Own calculations.

Table 4 contains similar data to discussed Table 3 but related to women (female students). Women could not study law at Jagiellonian University in the discussed period; they could study philosophy since 1897 and medicine since 1899. We can see two interesting things in this table.

First, there is higher dependence in the closest distances. Unlike male students, women were prone to choose the faculty according to the choice of other women from their family circle. Their number is not big enough to make more detailed analysis but it seems that while for boys studying law or medicine was important for their future job (so that even poor boys could try to clench their teeth and do their best to get a degree and started new, prosperous life), for girls studies were rather a case of social status and – as such – it was more closely connected to the status of their families.

Another important result is that for women there is a strong dependence between them and their siblings' side, that is their sisters, their sisters-in-law or mothers-in-law of their brothers and sisters. We can connect it to the result we noticed in male students, where especially students of medicine were often connected to their brothers and their families. In both observed cases (female philosophers and male students of medicine) the high result (over the threshold of confidence) appears in the moment when there is no such effect for the families of their parents.

We can suggest the following interpretation of the presented results, which is compatible with the presented data. The Jewish community of Cracow did not contain inherent differentiation that could result in studying in different disciplines by different social groups. Predilection towards specific disciplines was not something that could be inherited or otherwise acquired from atmosphere in grandparents' house. It was not shared between cousins. But brothers and sisters who attended the same schools and had common friends did influence each other, which also applied to the families of one's

girlfriends and boyfriends; we can also expect that the crucial cause was to have friends who are already studying.

3 Wealth

The next attribute in the network that can be taken into account is wealth. The relationships between economic position of a family and educational chances can be different. There is an obvious assumption that children from affluent families can have better start and receive better support of their families but we should remember that in 19th and early 20th century the social divisions were more visible and among obstacles in receiving university education were such hard ones like hunger, child labour, illiteracy or homelessness.

We don't know the value of property and income of everybody but knowing the structure of the family network we can guess it. Let us assume that the economic position is smoothly spread across families: that cousins or grandchildren of rich (or poor) people are also rich (or poor). We have the list of 571 most wealthy heads of families of Cracow Jews, who paid local tax at least 5 Austrian gulden in 1883[25]. If rich people are related to rich and poor are related to poor, we may suppose that people who have many rich people in their close families are also rich and people who have no such relationships (or a few, or very distant) are poor.

Rate	Number of payers
4	1
5	297
6	1
7	1
8	82
10	72
12	31
15	43
20	23
25	8
30	12

Table 5. Number of taxpayers for every tax rate in the 1883 election list for Cracow Jewish Religious Community

The hypothesis can be formulated as following: (a) the educational success, as measured in the probability of finishing studies as doctor (of philosophy, medicine or law, i.e. PhD, MD or J.D.) depended on (b) how close was a student to

25 "Spis Członków Zboru Izrelickiego w Krakowie."

as many "rich people" as possible. The closeness can be defined in different ways, for instance the following:

$$wealth\ index\ of\ A\ = \sum_{d(A,B)=0}^{5} \frac{t(B)}{2^{d(A,B)}}$$

Formula 2. Wealth index. A is one of our students, B is a person listed in the 1883 voters list, *t(B)* = tax rate paid by B in 1883 and *d(A,B)* = distance from A to B.

The results of our analysis are presented in Table 6.

Faculty	Average wealth index	Total	Dr rate	Av. Wealth Index if Dr	Av. Wealth Index if Dropout
Law	6.50	620	69%	6.95	5.51
Medicine	6.46	328	67%	6.83	5.72
Law or Med.	6.49	948	68%	6.91	5.59
Philosophy	6.56	263	16%	6.58	6.56

Table 6. Average wealth index of Jewish students from Cracow according to faculty they enrolled

There is a strong difference between students of philosophy and students of law or medicine: while only 16% of Jewish students enrolled in the Faculty of Philosophy received PhD title (42 out of 263), in the faculties of law and medicine over two thirds (average: 68) were awarded the title of doctor.

The statistically significance of the results have been checked using Kruskal-Wallis test (kruskal.test in R). The result for all students and for each of the faculties were not statistically significant (p-value > 0.05). However, if law and medicine were to be considered together, the p-value was 0.0478 which is should be enough for claiming statistical significance. The results for law and medicine are very similar: average doctor had wealth index either 6.95 (law) or 6.83 (medicine) and average dropout (person who did not receive title of Dr.) had either 5.51 (law) or 5.72 (medicine); the difference between them being small is also not statistically significant, so we will consider them as one kind of studies; we can call them "vocational studies" because being doctor of medicine or doctor juris provided immediate professional status (in medicine or law); on the other hand, doctor of philosophy (i.e., all other sciences and humanities except for law, medicine and theology) could be a teacher or a clerk but it wasn't so directly related.

Now we can reverse the process: having concluded that there was a visible and statistically significant connection between "wealth index" and "success in vocational studies" we can try to optimize the wealth index to see whether the effect can be strengthened with some changes of algorithm.

First: is it important to consider the actual value of tax rate? Maybe it is enough to count whether some relative was a member of established community (having voting rights) with no differentiation in the actual income? The key values to compare should be: (a) the ratio of "average wealth index if dr" and "average wealth if dropout" and p-value of the Kruskal-Wallis test for the difference. For the above, it was 1.236 and 0.0478. If we assume that everybody paid the same tax rate (t = 1), we receive 1.175 and 0.0429: the difference is smaller but more significant. If only those who paid over 5 gulden (the richer ones) receive t=1, the rate drs/dropouts was 1.20 but the result is not significant (p-value 0.129). If those who paid no more than 5 gulden received t=1 and the richer received t=2, the rate was 1.225 and p-value was 0.0425. If the richer (paying more than 5) received t=5, the rate was 1.223 and p-value 0.0684. Therefore the tax rate should not be ignored.

One can argue that we should not mix cause and result and count the former students as 1883 taxpayers (people whose distance was = 0) because for them the position was the result of being successful student and not a cause for it. If we ignore them, the original ratio is 1.161 bat the p-value is over the threshold: 0.0616. If, instead of Kruskal-Wallis test, we could use one-way ANOVA test (aov in R) the result were significant (p-value is 0.0349) but the author is not sure whether use of ANOVA is justified here. Therefore, we would be happy to ignore the cases of d=0 but it is not easy.

The next aspect that can be tested is the effect of network itself. In the denominator of Formula 2 we have 2^d. What if, instead, we had d or 3^d or 4^d? For 2^d we had ratio of 1.236 and p-value 0.0478. For 3^d we have ratio 1.368 and p-value 0.0290: both much better. For 4^d even better: ratio is 1.494 and p-value is 0.0264. This time the p-value is so small that we can try what happens if we remove cases when d=0: in such case we obtain ratio 1.337 and p-value 0.0398. For doctors of law he have ratio of 1.341 and for medicine: 1.292 but still, for more granular data the distinction is not statistically significant (for law p-value is: 0.0868 and for medicine: 0.239). If, instead, we assume just d in denominator (and $d>0$), the ratio is 1.169 but p-value is 0.0945.

We can summarize this part of discussion with the conclusion that there was a visible and statistically significant relationship between vocational studies (law or medicine) and social standing of family: Jewish students from "better" families were more likely to obtain title of Doctor of Medicine or Doctor of Law. Another important conclusion is that this relationship is best visible when we assume that the influence of other member of family is reversely proportional to powers of 4, that is a member of family removed by n links had

influence four times stronger than a member removed by *n*+1 links (his parent, child, spouse or a sibling).

4 Jews from Cracow, other Jews and Christians

We can compare Jews from Cracow with non-Jews (i.e., Christians) and Jews from outside Cracow. We have data for 31% of Christian students from this time (5,642 of 17,938 - those whose surname starts with S-Z)[26]. Out of the 5,642, 5,520 enrolled in faculties of philosophy, law, medicine or theology and either got their doctoral diploma in the same faculty or did not get it at all (see Tables 7 and 8).

Faculty	Jews (Cracow, networked)			Other Jews			Christians (S-Z)		
	All	Dr	%	Drop-out	Dr	%	All	Dr	%
Law	606	437	72%	1223	634	52%	2231	557	25%
Medicine	337	219	65%	838	313	37%	916	475	52%
Philosophy	248	28	11%	536	35	7%	2045	167	8%
Theology	0	0	0%	0	0	0%	328	10	3%

Table 7. Percentage of those who successfully finished studies (Drs) according to faculty, religion and origin

The data from Table 7 can be presented in the form of confidentiality intervals (calculated with discussed Formula 1). The results are in the Table 8.

Faculty	Jews (Cracow, networked)	Other Jews	Christians (S-Z)
Law	68.5% – 75.7%	49.0% – 54.6%	23.2% – 26.8%
Medicine	59.9% – 70.1%	34.1% – 40.6%	48.6% – 55.1%
Philosophy	7.4% – 15.2%	4.4% – 8.6%	7.0% – 9.4%
Theology	0.0% – 0.0%	0.0% – 0.0%	1.2% – 4.9%

Table 8. Percentage of those who successfully finished studies (Drs) according to faculty, religion and origin, confidentiality intervals

In the fields of medicine and law there are big, important differences: "The big Jewish family of Cracow" is doing much better than other Jews and Christians. Other Jews are still better than Christians in law but worse than Christians in medicine (this was probably connected to the fact that studying medicine and

26 In the volumes of the "Corpus Studiosorum" covering letters A-R (21,253 entries) we have analysed only Jewish students (2996) and in the volumes S-Z (7,620 entries) all students (Jewish and Christian) have been analysed.

not becoming "Dr." was useless – which was not the case for law or philosophy; that was why even Christians did they best to complete their studies and receive title of MD; in this case they even outperformed Jews from outside Cracow).

5 Geographic networks

It was also interesting to analyse networking aspects related to geographic neighbourhoods. What made people study?

From all birth records since 1891 to 1895 we know that 4,555 children were born in Jewish Cracow in this period, out of whom 228 started to study in Jagiellonian University before 1918. The exact address of birth (street and house number) was recorded for almost everybody and it was possible to collect the detailed location. We are able to propagate the location through the whole network and connect everybody with the district closest in their family network if it was unique. If there were two concurrent possibilities, we left it blank.

The outcome showed that there was no visible difference, families were mixed. We can also compare the rate: doctors vs. dropouts for every district, but the result is very similar, about 55-56%. The only difference was for the people who had two or more simultaneous possibilities – they had rate of 61% which suggests that it was not place but change of place (mobility) that mattered, but the difference was not statistically significant (χ^2=0.12). It is likely that further analysis of mobility of Jewish families (especially moving from the old Jewish District in the direction of the city centre) will provide results that will be statistically significant but it is too early now.

6 Final remarks

In recent times we can see growing interest among scholars who are studying processes of status attainment, educational opportunities and new class formation worldwide. This can be seen in sociologists studying disadvantaged groups like African-Americans or destitute immigrants in the United States or class formation in countries of the former Soviet Bloc in Europe. Such research can be based on contemporary data, acquired from living subjects through direct questionnaires and from national and international statistical bureaus. Such research is fascinating because it helps in understanding contemporary societies and establish and test theories that help in understanding them.

One cannot be an expert on everything, however. The author of the present article grew in quite another intellectual framework: the paradigm of history of East European intelligentsia where enrolment to university was considered not a step in one's professional career but rather a door to some kind of hereditary nobility. One of the most famous students of the Jagiellonian University of the

analysed time, Tadeusz "Boy" Żeleński (enrolled 1894 in the faculty of medicine, doctor of medicine 1900), famous poet, essayist and translator of French literature, promotor of emancipation of women and man who had a great impact on self-consciousness of Polish intelligentsia, published an essay in 1932 under the title "Bourgeois nobility" comparing high school finishing examination ("matura", finishing high school and opening door to university) as contemporary form of ennoblement.[27] The history of Polish intelligentsia has been summarized in detail by the already quoted *Dzieje inteligencji Polskiej*.[28] The pre-1918 Eastern European societies still retained much of medieval feudal structure and had features different from the societies commonly analysed by contemporary sociologists.

The Jews in Cracow cannot be subject to contemporary analysis because virtually all of them have been killed by the German invaders between 1939 and 1942. But when they were still thriving and nothing suggested their extinction, they were subjects to process of acculturation: virtually all educated Jews in Cracow (using the title "Dr" attached to their surname) in the 1921 census declared Polish nationality and Polish as their main spoken language. [29] They were not "Jews", they were rather "Poles of Mosaic faith", unlike their non-educated cousins, who spoke Yiddish at home.

The data about educated Jews in Cracow can be therefore used as a bridge allowing cross-disciplinary research between the classic notion of university education as "bourgeois nobility" and present sociological analysis of status attainment in minority groups. The author is not able to do this on his own but it seems that shared work, using the already collected or easily available data can be of great interest to scholars in all these areas: general sociological analysis (of education, status attainment or social networks), history of Eastern European intelligentsia and history of the Ashkenazi Jews and their culture.

7 Conclusions

The analysis of family network of Jewish Community of Cracow in 19th and early 20th century is far from being complete. It already reveals some interesting and statistically significant results, however.

Not all students of Jagiellonian University were born equal: those born in Jewish families in Cracow, closely related to rich citizens of the Jewish City, had better opportunity to be granted Doctor of Law or Doctor of Medicine

27 Tadeusz Żeleński, "Burżuazyjne Szlachectwo," Wiadomości Literackie 9, no. 33 (450) (August 7, 1932).

28 Jedlicki, Janowski, and Micińska, Dzieje Inteligencji Polskiej Do Roku 1918.

29 "Spis Ludności Miasta Krakowa z r. 1921."

than those who were Christians or were Jewish but from outside Cracow or were Jewish and from Cracow but had less strong connection to local Jewish elite.

The best matching formula for calculating the influence of relatives (formula that provides strongest and most significant results) we found in this community was following: in the network where every link between A and B means "A is a child, parent, spouse or sibling of B", extending a path by one link makes influence of the distant end about four times weaker.

References

"Akta Stanu Cywilnego Izraelickiego Okręgu Metrykalnego w Krakowie," n.d. 29/1472/0. Archiwum Narodowe w Krakowie. https://szukajwarchiwach.pl/29/1472/0/1/str/1/1000#tabJednostki (accessed 23 November 2018).

Aleichem, Sholem. Tevye the Dairyman and Motl the Cantor's Son. Translated by Aliza Shevrin. New edition. New York, N.Y: Penguin Classics, 2009.

Barabási, Albert-László, and Márton Pósfai. Network Science. 1 edition. Cambridge, United Kingdom: Cambridge University Press, 2016.

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 2008, no. 10 (October 9, 2008): P10008. DOI: 10.1088/1742-5468/2008/10/P10008.

Corpus studiosorum Universitatis Iagellonicae 1850-1918. Archiwum Uniwersytetu Jagiellońskiego, 1999.

Domański, Henryk, and Dariusz Przybysz. Homogamia małżeńska a hierarchie społeczne. Warszawa: Wydawnictwo IFiS PAN, 2007.

Ellenson, David, and Daniel Gordis. Pledges of Jewish Allegiance: Conversion, Law, and Policymaking in Nineteenth- and Twentieth-Century Orthodox Responsa. Stanford University Press, 2012.

"Gephi - The Open Graph Viz Platform." Accessed August 13, 2018. https://gephi.org/.

Jacomy, Mathieu, Tommaso Venturini, Sebastien Heymann, and Mathieu Bastian. "ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software." PLOS ONE 9, no. 6 (June 10, 2014): e98679. DOI: 0.1371/journal.pone.0098679.

Jedlicki, Jerzy, Maciej Janowski, and Magdalena Micińska. Dzieje Inteligencji Polskiej Do Roku 1918. Warszawa: Instytut Historii PAN, Wydawnictwo Neriton, 2008.

Minakowski, Marek Jerzy. "Genealogia Masowa – Metodologia Tworzenia i Publikacji Bazy Danych." In Edytorstwo Wobec Masowości Źródeł Najnowszych, Vol. 3. Edytorstwo Źródeł XIX i XX Wieku, Teoria i Praktyka. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego, 2018.

——— "Jews of 19th-Century Cracow in Civil Registers." Kraków: Fundacja Alef dla Rozwoju Stiudiów Żydowskich, 2018.
——— "Mass Genealogy: Top 1% of 19-Th Century Polish Society as a Single Family Network (PageRank-like Analysis)." Mainz, Germany, 2017. https://www.eusn2017.uni-mainz.de/files/2016/08/EUSN2017_Booklet_25_09.pdf.
——— "Modelowanie rozkładu kapitału ekonomiczno-społeczno-kulturowego poprzez genealogię masową." Przeszłość Demograficzna Polski 38 (2016): 63–88. https://doi.org/10.18276/pdp.2016.4.38-03.
——— "Sieć społeczna wokół Kuriera Warszawskiego na podstawie jego nekrologów z lat 1821–1861." Przeszłość Demograficzna Polski 39 (2017): 209–51. https://doi.org/10.18276/pdp.2017.39-09.
——— "Wielka Genealogia Minakowskiego," 2018. http://wielcy.pl/.
Minakowski, Marek Jerzy, and Anna Lebet-Minakowska. "Legitimate, Legitimized and Illegitimate Families of 19th-Century Cracow Jews." Przeszłość Demograficzna Polski 40 (2018).
Newman, M. E. J. "Modularity and Community Structure in Networks." Proceedings of the National Academy of Sciences 103, no. 23 (June 6, 2006): 8577–82. DOI: 10.1073/pnas.0601602103.
Smoczyński, Rafał, and Tomasz Zarycki. Totem inteligencki. Arystokracja, szlachta i ziemiaństwo w polskiej przestrzeni społecznej. Warszawa: Wydawnictwo Naukowe Scholar, 2017.
"Spis Członków Zboru Izrelickiego w Krakowie," 1883. 29/33/0 Akta miasta Krakowa » Series: 3.2.3 Akta Magistratu w układzie rzeczowym » File/unit: Kr 7055. Classification scheme code: fasc. 29 konw. 003112/IV/1920. Archiwum Narodowe w Krakowie.
"Spis Ludności Miasta Krakowa z r. 1921," n.d. 29/91/0. Archiwum Narodowe w Krakowie. https://szukajwarchiwach.pl/29/91/0/#tabZespol (accessed 23 November 2018).
Żeleński, Tadeusz. "Burżuazyjne Szlachectwo." Wiadomości Literackie 9, no. 33 (450) (August 7, 1932).

KAISER, MAXIMILIAN; SCHLÖGL, MATTHIAS; LEJTOVICZ, KATALIN AND RUMPOLT, PETER ALEXANDER

Artist Migration Through the Biographer's Lens: A Case Study Based on Biographical Data Retrieved from the Austrian Biographical Dictionary.

Journal of Historical Network Research 2 (2018) 76-108.

Keywords

Digital art history, prosopography, biographical dictionary, spatial history of art, digital humanities

Abstract

A lexicon like the *Österreichisches Biographisches Lexikon 1815–1950* (Austrian Biographical Dictionary) seems to be as all in one cast because it is built on a set of formal rules for writing articles and some strict but basic criteria for the selection of new entries. The human reader can find information within that resource on a wide range of topics as well as detailed information about the life and career paths of historical individuals. An attempt to systematically analyze this information in a larger scale, however, is condemned to fail without the help of computational methods. In a first stage it is needed to convert unstructured text in structured

information. Pieces of information, the so called biographical building blocks, can be identified in two ways: through natural language processing methods and by manual annotations. Both processes which are intertwined and - in the case of the APIS project - done in a custom-built virtual research environment provide the existing biographical data at hand for the analyses following in later stages. This paper aims at describing the data collection process on the example of place names which can be found in artist biographies and at demonstrating possible use cases for historical network research. In this context it is also outlined how this field of research can benefit particularly from biographical data.

1 Introduction to the Austrian Prosopographical Information System (APIS)*

The *Österreichisches Biographisches Lexikon 1815–1950* (Austrian Biographical Dictionary) was launched as a project by the Austrian Academy of Sciences in the 1950s.[1] First concepts and plans for the publication of only three to four volumes soon proved as insufficient. Up to the present day, the lexicon consists of 15 printed volumes with 69 issues. The documentation, on which basis the printed version of the ÖBL was made, started as card-index boxes. Each card contained the fundamental facts, a short description of the most important professional relations and a list of reference literature. A transformation process to a digital research environment has been introduced step by step from a first concept of a relational database system in 1997 called *ÖBL-DOC* to the *Gideon* database in 2009.[2] Now the workflow from the distribution of new biographies

* **Acknowledgements:** This paper was made possible by the framework of the research project "Mapping historical networks: Building the new Austrian Prosopographical/Biographical Information System (APIS)" which is funded by the Austrian National Funds for Research, Technology and Development (Österreichische Nationalstiftung für Forschung, Technologie und Entwicklung).

Corresponding authors: Mag. Dr. Maximilian Kaiser, Institute for Modern and Contemporary History, Austrian Academy of Sciences, Maximilian.Kaiser@oeaw.ac.at, Katalin Lejtovicz MSc, Austrian Centre for Digital Humanities, Austrian Academy of Sciences, Katalin.Lejtovicz@oeaw.ac.at, Mag. Matthias Schlögl, Austrian Centre for Digital Humanities, Austrian Academy of Sciences, Matthias.Schloegl@oeaw.ac.at & Mag. Peter Alexander Rumpolt, Institute for Urban and Regional Research, Austrian Academy of Sciences, Peter.Rumpolt@oeaw.ac.at.

We thank the anonymous peer reviewers for their comments and suggestions which had a considerable influence on the final version of this article.

1 It will be abbreviated in the following as ÖBL.

2 See Christine Gruber and Roland Feigl, "Von der Karteikarte zum biografischen Informationsmanagementsystem. Neue Wege am Institut Österreichisches Biographisches Lexikon und biographische Dokumentation", in *Biografische Lexika im Internet. Internationale Tagung der „Sächsischen Biografie" in Dresden (30. und 31. Mai 2008)* (= Bausteine aus dem Institut

to the authors to the publishing of the final articles can be done digitally. In respect of the prosopographical research topic, the former member of the editorial board Friederike Hillbrand-Grill has noted already in the 1990s that a digital approach would help researchers by conducting quantitative studies.[3] Soon after the ÖBL went online in 2011, a transnational initiative was launched to cross reference entries between the different national biographical dictionaries such as the *New German Biography* (NDB), the *Historical Lexicon of Switzerland* (HLS), the *Slovenian Biography* (SBI), the *Austrian Music Lexicon* (OeML) and smaller regional biographies from Germany.[4]

Since then, in addition to the printed version of the ÖBL, six "online only" issues have been published. The goals of this new edition are the extension of the time period covered by the lexicon up to the year 2010 and the update of rather scarce biographies from the first two volumes. For Central Europe, the ÖBL is the only comprehensive scientific reference work which covers the lives and careers of individuals who had impact through their life and work in the Habsburg Monarchy or in the successor state of Austria. At the beginning of the research project "Mapping historical networks: Building the New Austrian Prosopographical/Biographical Information System" (APIS) the data at hand was a corpus of approximately 18,000 biographies in the form of partly structured XML-files. The main tasks, were the enrichment of the unstructured text, the development of methods for analysing this kind of data and the answering of historical research questions based on the visualization of networks and maps.[5]

The application of digital humanities methods in biographical research is a relatively new field of research. Back then there didn't exist any form of comprehensive text corpora under open-access licensing which could have provided structured biographical data beyond the fundamental facts. In 2015, when the APIS project started, the workshop *Biographical Data in a Digital World* (BD2015) was held in Amsterdam. This was the first time that researchers from

für Sächsische Geschichte und Volkskunde 14), ed. Martina Schattkowsky and Frank Metasch (Dresden: Thelem Universitätsverlag, 2009), 55–75; Ágoston Zénó Bernád, "Das Österreichische Biographische Lexikon 1815–1950. Ein Nachschlagewerk im Wandel der Zeit", in *Hortus amicorum. Köszöntőkötet Egyed Emese tiszteletére [Festschrift für E. Egyed]*, ed. Katalin Ágnes Bartha, Annamária Biró, Zsuzsa Demeter, and Gabriella-Nóra Tar (Kolozsvár: Erdélyi Múzeum Egyesület, 2017), 464.

3 Friederike Hillbrand-Grill, *Ein neues Forum der Personengeschichte* (= Österreichisches Biographisches Lexikon, Schriftenreihe 2) (Wien, 1992), 13–14.

4 If you search for the artist Heinrich von Angeli (1840–1925), the database gives you not only a list of names but also shows you that an additional biography can be found in the database of the NDB. See *Biographie-Portal*, www.biographie-portal.eu (accessed 19 February 2018).

5 The APIS project is a collaboration of the Institute for Modern and Contemporary History (INZ), the Austrian Centre for Digital Humanities (ACDH) and the Institute for Urban and Regional Research (ISR) at the Austrian Academy of Sciences.

the field of the digital humanities and biography research came together to discuss the diverse topics, tasks and challenges. In autumn of the same year, the conference *Europa baut auf Biographien* took place as a kick-off-event of APIS in Vienna. The proceedings of these events so far form the most complete overview of these topics.[6]

2 Artist migration as a topic in the realm of digital art history

Study trips, mobility and migration shaped the narratives of artist biographies since the beginning. This can be traced through the frequently mentioned places in biographical texts, starting by Giorgio Vasari's famous *Lives of the most eminent painters, sculptors, and architects* up to modernity. Cities like Florence or Rome were the centres for art and culture during Renaissance times, lost their hegemony over time and were replaced by new art metropoles. At the end of the 17th century Paris superseded Rome as Europe's most important art metropolis in connection with the newly established Salon exhibitions and art discourse.[7] "Best seen through the bird's view", one might think. In 2014 a group of researchers, among them the art historian Maximilian Schich and the social network expert Albert-Laszlo Barabasi, started a project called *A Network Framework of Cultural History*.[8] The findings of this study were object of discussion and controversy not limited to the field of digital art history – a research field related to the digital humanities. In the *Science* article they mapped the cultural history of mankind on the basis of freely accessible biographical data.[9] Critics pointed out that the study was based on weak data. *Freebase.com* and other data sources provided them only fundamental facts such as the places of birth and death of an artist.[10] The critical reception of the study

6 See Serge ter Braake, Antske Fokkens, Ronald Sluijter, Thierry Declerck, and Eveline Wandl-Vogt, eds., *BD2015. Biographical Data in a Digital World 2015. Proceedings of the First Conference on Biographical Data in a Digital World 2015. Amsterdam, The Netherlands, April 9, 2015* (2015), http://ceur-ws.org/Vol-1399/ (accessed 23 November 2018); Ágoston Zénó Bernád, Christine Gruber, and Maximilian Kaiser, eds., *Europa baut auf Biographien. Aspekte, Bausteine, Normen und Standards für eine europäische Biographik* (Wien: new academic press, 2017).

7 See Uwe Fleckner, Maike Steinkamp, and Hendrik Ziegler, "In die Welt geschickt. Künstlerische Mobilität vom Mittelalter bis in die Gegenwart", in *Der Künstler in der Fremde. Migration – Reise – Exil,* ed. Uwe Fleckner, Maike Steinkamp, and Hendrik Ziegler (Boston: de Gruyter, 2015), 4–5.

8 Maximilian Schich et al., "A network framework of cultural history", *Science* 345, no. 6196 (2014): 558–562, http://barabasi.com/f/491.pdf (accessed 23 November 2018).

9 The tool which was developed during the project "Cosmobilities – Grenzüberschreitende Lebensläufe in den europäischen Nationalbiographien des 19. Jahrhunderts" should be mentioned. *Cosmotool*, https://cosmotool.de.dariah.eu/cosmotool/personsearch/ (accessed 23 November 2018).

10 Malte Rehbein, "Digitalisierung braucht Historiker/innen, die sie beherrschen, nicht beherrscht", *H/Soz/Kult*, 27. November 2015, www.hsozkult.de/debate/id/diskussionen-2905 (accessed 18 April 2018).

made clear that a more detailed set of data is needed to make valid assumptions about the migration of artists.

Schich was the first to conduct such a case study with artist biographies. Historian Sarah Panter and her colleague Michael Piotrowski, a computer linguist, attempted a similar study, but from a historical perspective and with a different focus. Panter coined the new term "cosmobilities" based on two concepts framed by sociologists: "cosmopolitan" and "mobilities". This new term stands for a methodological approach which should be acknowledged here because of its "stimulating impetus for analysing transnational lives".[11]

For the sake of completeness another project should be mentioned. In the republic of letters project, the amateur architects, who went in the 18th century during their Grand Tours to Italy, were subject of a quantitative case study. The authors of this study emphasised that they looked for "aggregates and patterns, but not in merely quantitative terms".[12] They took a sample of sixty-nine persons from the digitized prosopographical *Dictionary of British and Irish Travellers in Italy, 1701–1800* as a basis for their analyses of correlations between the travel activities and the emergence of social networks.

In comparison to that, the research project APIS deals with a later epoch and isn't limited to a group of elites, but also includes people from the so-called second rank. It aims to provide researchers with access to prosopographical data hidden in the ÖBL biographies. To achieve this goal the text corpus is processed with natural language processing methods. Furthermore, manual annotation with a web-based highlighter tool and data curation is done by historians with specific domain knowledge. With each annotation in the web application, a connection to a linked open data source like the *Integrated Authority File* (GND) or *Geonames.org* (GeoNames) is established. From these resources, information is retrieved through an automatic parsing process. With this method, place entries in the APIS database receive not only an URI and a label in the form of a string but also geographic coordinates that are stored as metadata in these external data sources.[13] The following section introduces the sample data set collected within the APIS web application and exemplifies with

---

11 Sarah Panter, Johannes Paulmann, and Margit Szöllösi-Janze, "Mobility and Biography: Methodological Challenges and Perspectives", in *Mobility and Biography* (= Jahrbuch für Europäische Geschichte / European History Yearbook 16), ed. Sarah Panter (Berlin and Boston: de Gruyter, 2016), 5.

12 Giovanna Ceserani, Giorgio Caviglia, Nicole Coleman, Thea de Armond, Sara Murray, and Molly Taylor-Polesky, "British Travelers in Eighteenth-Century Italy: The Grand Tour and the Profession of Architecture", *American Historical Forum* 122, no. 2 (April 2017): 449, DOI: 10.1093/ahr/122.2.425.

13 "Integrated Authority File (GND)", available under the following URL: http://www.dnb.de/EN/Standardisierung/GND/gnd_node.html (accessed 4 December 2017); "GeoNames",http://www.geonames.org/ (accessed 4 December 2017).

the help of auto-biographical sources three different migration stories of artists which can be found in the ÖBL.

With this, it is intended to emphasize the human condition behind all the following networks, maps and statistics. In connection to *A Network Framework of Cultural History* the final section demonstrates what assertions about migrations can be made with a more detailed set of biographical data.

2.1 The "Künstlerhaus" data collection

When working with biographical data, the question how to build a sample is inevitable. The following quote originates from the English historian Lawrence Stone. At the beginning of the 1970s, when scholars started to discuss how computers can be applied in humanities research for the first time, he already pointed out that it will become more and more important for prosopographical research to reason data sampling through methodology.

> The availability of the computer will increasingly tempt some historians to concentrate their energies on problems that can be solved by quantification, problems which are sometimes – but by no means always – the most important or interesting ones. It will also tempt them to abandon sampling techniques (…).[14]

Sampling of biographical data can be done in numerous ways: cohorts can be compiled according to their belonging to a generation, the origin of a person or upon thematic criteria. In this particular study the decision was in a first instance to select a group based on their profession. In total there are about 2,500 artists in the ÖBL.[15] After that, every artist who was member of the artist association "Genossenschaft bildender Künstler Wiens" was selected. This was done by running through the yearly published lists of members which are still preserved in the archives until today. This decision was made because of the prominence and the art historical relevance of this institution. In the end the collection for this study consisted of 506 ÖBL biographies.

The association itself, also known by its alias name "Künstlerhaus", was founded in the year 1861.[16] At this point in history, except for the Albrecht-Dürer-Verein, it has been the only artist association in Vienna and the most

14 Lawrence Stone, "Prosopography", *Daedalus. Historical Studies Today* 100, no. 1 (1971): 72.

15 The "Künstlerhaus" data collection was built at the beginning of the APIS-project in 2015. All biographies published until the 14th volume (65th issue) of the ÖBL have been initially imported into the database. Approximately 2,255 (12,5 %) of the ~18,000 biographies in the lexicon were artists' biographies at this point. The data collection grew because of an additional data import of the ÖBL print (66th issue) and (4th) online edition at a later stage of the project in 2016. At the time this article was written in 2018, the database therefore consisted of ~2,500 biographies of artists (+ 1,4 %).

16 Instead of the association's full name it's alias name "Künstlerhaus", which was used by contemporaries of this time, will be used synonymous in the following.

important professional network for artists in the Habsburg Monarchy. With the founding of the Vienna Secession 1897, and some years later that of the Hagenbund, for the first time the monopoly of the Künstlerhaus was challenged. For this reason, admission to the association and the participation in the exhibitions in the building at the Karlsplatz have long been decisive for the acknowledgement of the social status of an artist.

This leads to the following initial questions: Which places are frequently mentioned in artists' biographies? What categories can be used for describing the various relation types? How can these categories be distinguished statistically? Before starting the manually performed annotation project in the APIS web application, it was assumed that relations between the biographies and the various entities wouldn't be equally distributed. This assumption was made on the basis that artists' biographies are written differently in comparison to that of lawyers, politicians, clergymen or physicians. The numbers resulting from the annotation project are shown in figure 1. In ÖBL biographies it is common practice to mention place names instead of institutions (e.g. "studied in Vienna" instead of "studied at the University of Vienna"). It also has to be considered that the circumstances for historical research has been completely different in the 1950s compared to nowadays. The digitalisation of resources from civil registers, letters, historical newspapers or literature has advanced significantly in the past decades.

Finding categories for the quantification of mobility is not an easy task. Especially when taking the timespan and the regions covered by the lexicon into account. To start at any point, it was necessary to define some basic categories. The following text passage describes a small section from the life and career path of the Austrian painter Alexander Demetrius Goltz as it is depicted in the ÖBL. "He undertook study trips to France, England, the Orient and America and in between visited Munich (1884 to 1888), Dachau, Paris and the Bretagne."[17] Every place name mentioned in this sentence is underlined to illustrate those passages which were annotated in APIS (figure 2). In this case, each of the first four places would be the destination of a study trip. Therefore, the category "was place of a study trip" was chosen. The following place names must be annotated differently because of the diverging textual description. It is obvious that the biographer wanted to point out that these stays lasted longer.

17 N. N., "Goltz, Alexander Demetrius, Maler", in *Österreichisches Biographisches Lexikon 1815–1950*, Bd. 2 (Lfg. 6, 1957), 29.

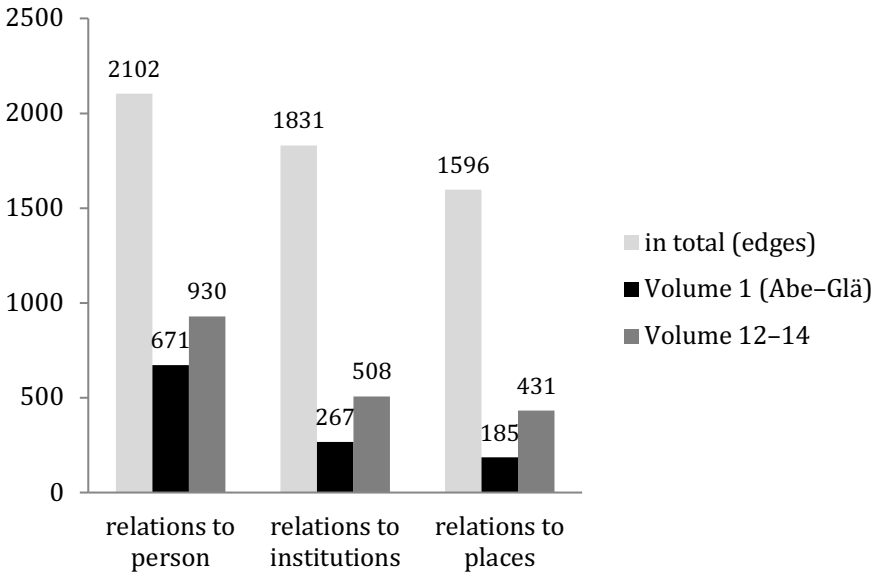


Figure 1. This diagram shows artists' biographies and their relations to other persons and other APIS entities (places, institutions). In the first volume of the ÖBL 385 artists can be found and for 82 of them the membership in the Künstlerhaus can be proven. A comparison of this group makes it necessary to take all published artists' biographies of the last three volumes together into account. Even though these volumes contain nearly the same amount of biographies, these articles exceed the earlier entries of the ÖBL in quality and detail of information. This fact can be concluded out of the number of relations to places or institutions. The most significant growth can be seen by these two kinds of entities. (graphic of the authors)

Once all 506 biographies are annotated through the APIS web application, the analyses of the used categories can be started. Figure 3 shows the results of this step and in descending order the most frequent types of person-place relations as they appear in the artists' biographies of the ÖBL.

Within the digital art history discourse, it is often mentioned that building such categories is one of the most important epistemological tasks during a project. Miriam Posner made that clear when she stated the following during an interview:

> But in my own work, I've found that I gain things, too, not only from the finished product of, for example, a network diagram, but from the iterative process that goes into making it. (…) Sometimes this tension between what goes unsaid in

text-based scholarship and what needs to be made explicit in a data-based project becomes the real question at the heart of your work.[18]

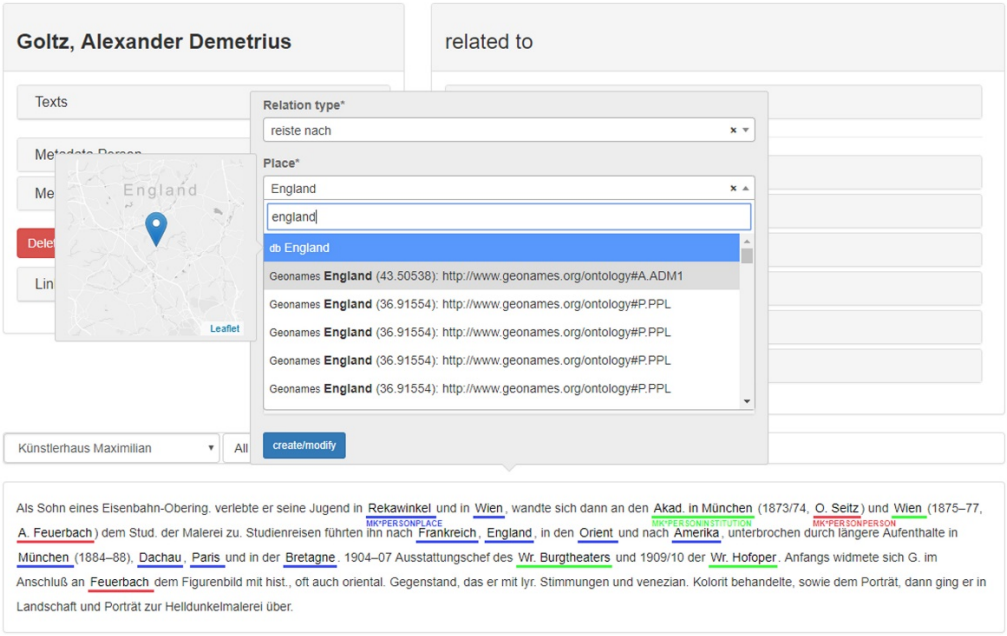


Figure 2. The screenshot shows how text passages are annotated in the web application. First, the type of relation "was place of a study trip" (in German "reiste nach") must be selected. Second, the right entity must be chosen. After typing the name of the place, an auto-complete-function brings up existing entries of the database (indicated as db) and entries from external resources (indicated as GND or GeoNames). If the researcher scrolls through the list map tiles appear and show where the place is situated on the overall map. Depending on what type of entity is annotated, the text is finally highlighted in blue (place), green (institution) or red (person). This provides the researcher with an overview of the progressing annotation process.

The difference of 0,2 percentage points between the number of places for birth and death can be explained through the fact that in some cases this information is simply not available. In other words, the author of the biography couldn't find any clues of where and how the artist had died. The least often mentioned relation types in this sample are "migrated to" (3), "went into exile" (2) or "was deported to" (2).[19]

18 Miriam Kienle, "Digital Art History 'Beyond the Digitized Slide Library': An Interview with Johanna Drucker and Miriam Posner", *Artl@s Bulletin 6*, no. 3 (2017): 125, http://docs.lib.purdue.edu/artlas/vol6/iss3/9 (accessed 23 November 2018).

19 The numbers between the round brackets stand for the frequency.

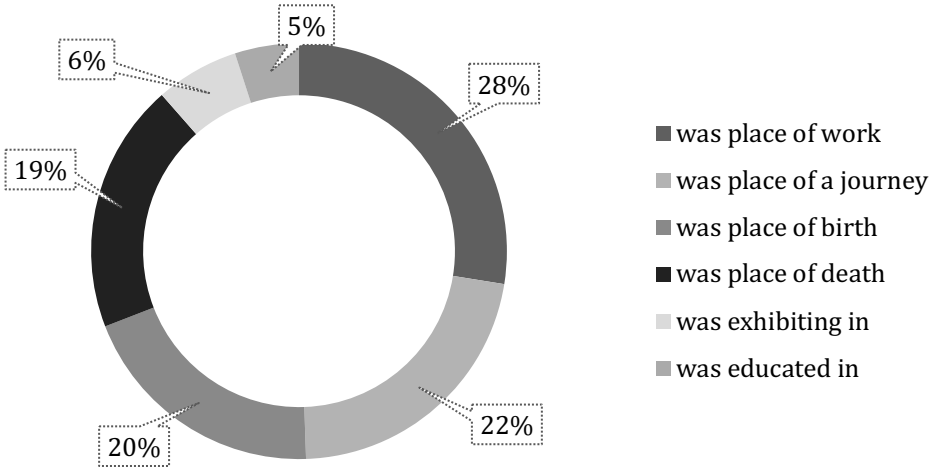


Figure 3. The diagram shows the percentage of the different person-place relation types which were used during annotating the 506 artists' biographies. These categories paraphrase on a general level the individual context in which an artist is related to a place mentioned in his biography. Therefore, places of work constitute the highest share (28 %) within all places mentioned in the biographies of artists, followed by places visited during a study trip (22 %). Places attended by reasons of education only represent 5 % of all places. (graphic of the authors)

2.2 Different angles on three migration stories

The stories, which can be told in connection with the biographies of the following three artists, are intended to illustrate the diversity of migration stories. For this purpose, Leopold Carl Müller (1834–1892), Josef Engelhart (1864–1941) and Jehudo Epstein (1870–1945) were chosen.

The first artist, Leopold Carl Müller, can be characterized as an Orient painter. He was a friend of some of the most influential painters of the so called Ringstrasse era like Hans Makart or August von Pettenkofen. A photography, which is frequently reproduced and is today in possession of the image archive of the Austrian National Library shows him together with his colleagues and friends during a study trip to Cairo. The ÖBL summarises his travelling activity as follows: "M. was on journeys a lot: 1861/62 to Hungary, 1870–72 to Venice, 1872/73 to Sicily, then several times to Egypt."[20] This line gives only a brief overview. By consulting the monograph *Carl Leopold Müller. Ein Künstlerleben in Briefen, Bildern und Dokumenten* written by Adalbert Franz Seligmann it is possible to get a better knowledge of what exactly the painter did during his

[20] Every quote from the ÖBL is a corresponding translation by the authors of the original German text in the lexicon. – Hans Schöny, "Müller, Leopold Karl (1834–1892), Maler und Zeichner", in *Österreichisches Biographisches Lexikon 1815–1950*, Bd. 6 (Lfg. 30, 1975), 423–424.

stays at the various locations. With this book, Seligmann published his transcriptions of the personal correspondence at request of the artist's relatives. After finishing his studies in Vienna, the first study trip of the artist took him to the Hungarian countryside. His destination was the small town of Szolnok, which later became famous through the founding of its own artist colony. 1861 Müller wrote to his father after arriving at the town of Szolnok:

> I have arrived in Szolnok in a good mood and safe and sound, have been talking much about politics on my way and was able to settle myself into the local condition. I have established my atelier in the house of a miller, which is not so far away from the miller's own domicile. (…) The Austrian government tries everything to provoke the Hungarian people to any violent uprise – but every provocation and try had failed because of the smart leadership of this people.[21]

In another letter, addressed to his friend Ferdinand Julius Laufberger[22], he wrote about his stay in Venice in the year 1871. There he illustrated the advantages of the Italian town in comparison to Paris during winter:

> As I had seen it coming, Pettenkofen [note: August Pettenkofer (Pettenkofen), 1822–1889] is completely unhappy that he didn't followed my example. Yesterday I received a writing from him in Paris. Within that, he informed me that he hadn't seen the sky since his arrival and that, while writing these lines by candlelight on two o'clock afternoon, in which from dirty blackened snow and thick clouds of fog can be read. (…) How can it be for a German, if he must not do, to go to Paris at this time of the year. It can't be cosy there right now under all these fools.[23]

In 1875, as Müller plans to accompany the Prince of Wales on his journey to India disappeared in sound and smoke, he decided to leave Venice in direction to Egypt.[24] He hoped that besides his other friends also Pettenkofen would join him on his study trip. Müller left in November by ship. He wrote his first letter to Laufberger at the 7 th of December that year:

---

21 These and the following passages are corresponding translations by the authors from German into English. – See Adalbert Franz Seligmann, ed., *Leopold Carl Müller. Ein Künstlerleben in Briefen, Bildern und Dokumenten* (Wien: Rikola Verlag, 1922), 14–15.

22 See Hans Schöny, "Ferdinand Julius Laufberger (1829–1889), Maler und Graphiker", in *Österreichisches Biographisches Lexikon 1815–1950*, Bd. 5 (Lfg. 21, 1970), 49.

23 Seligmann, *Müller*, 56–57.

24 To be part of this undertaking, Müller travelled to England for meeting the prince in person. The whole journey was published with plenty of illustrations in book form later. A wide range of items can be found in the collection of the Royal Collection Trust. One interesting item is an illustrated map which shows the path of the journey. See Royal Collection Trust, https://www.royalcollection.org.uk/collection/themes/trails/the-prince-of-waless-tour-of-india-in-1875-6 (accessed 2 February 2018).

> Beloved friend! I'm, as you know once again in my loved Cairo – and I am delighted. You might know that Makart, Huber [note: Carl Rudolf Huber, 1839–1896] and Lenbach [note: Franz von Lenbach, 1836–1904] are also here. I lost plenty of my time by wandering around with them and so on. Pettenkofen got aquaphobic in Triest and returned home, although he had intended to come here.[25]

On this occasion the previous mentioned photography was shot. A similar photography, showing count Karl Khevenhüller, architect Adolf Gnauth, Lenbach, Makart, Huber and Müller, can be found in Seligmann's book. Seligmann as a journalist and art critic is an interesting character, because in the 1920s he was writing some biographies for the *Neue Österreichische Biographie* (e.g. Heinrich von Angeli, Franz von Defregger, Pauline von Metternich, Johann Nepomuk Graf von Wilczek), the prequel of the ÖBL, and therefore is strongly connected with the early history of the lexicon.[26]

As a second example, the story of Josef Engelhart should be told. His fame is grounded on his role as one of the co-founder of the Vienna Secession. Legend has it, that because his painting *Kirschpflückerin* was rejected by the jury of the annual exhibition of the Künstlerhaus, the Secession was founded. It was the trigger for the upcoming events and is known in art history for exactly that reason. In the ÖBL it is written:

> On the wish of his parents he studied at the Technical University of Vienna, but then went on to the academies of Vienna and Munich and, after a long time stay at Paris (1890), made study trips to Spain and Italy.[27]

In his autobiographical novel *Ein Wiener Maler erzählt* he describes his stay in Paris in detail. During this time, he came into contact with many artists and actors.

> My cottage was behind a tenement in the Rue des Martyrs, rising up the Montmartre. A door at the end of the courtyard led up a few steps to it. Through the adjoining gardens and mansion-like houses, the whole thing had an almost rural character, and if in summer the foliage of the chestnut trees covered the front of the house, one could actually feel like being transferred into a small country town, even though one was in the heart of Paris. I moved here as subtenant of the painter Mesplès [note: Paul-Eugéne Mesplès, 1849–1924]. To the left of my studio was the property of the successful comedian Gandillot [note:

---

25 Seligmann, *Müller*, 109.

26 See Maximilian Kaiser, *Der Wiener Diskurs zur Avantgarde. Rekonstruktion und Analyse des Diskursnetzwerkes* (Dissertation, Universität Wien, 2017), 147.

27 N. N., "Engelhart, Josef, Maler und Bildhauer", in *Österreichisches Biographisches Lexikon 1815–1950*, Bd. 1 (Lfg. 3, 1956), 251.

> Léon Gandillot, 1862–1912], to the right was the garden and home of the member of the Comédie Francaise Talbot [note: Denis-Stanislas Montalant (Talbot), 1824–1904], who, despite his high years, still received a lot of young ladies whom he taught in acting.[28]

These contacts had been very useful when he travelled through Europe on behalf of the newly formed Vienna Secession. In this mission he sought allies among the most famous of European artists like Edgar Degas or James McNeal Whistler, to strengthen the international character of the artists' association Vienna Secession and its local exhibition program. Engelhart's novel is full of anecdotes. Some of these episodes, for example his meeting with Whistler, must be seen especially from his point of view and his understanding as an agent of the Secession.

The Jewish painter Jehudo Epstein, our third example for an artist's life and migration story, has also written an autobiography. In contrast to Josef Engelhart who got supported by his parents and was lucky to marry into a wealthy family, Epstein was born in a jewish ghetto in Sluzk (today's Republic of Belarus) in poverty.[29] He moved to Vienna in hope of success and planned to make his living as a painter. This is the leading topic of the book *Mein Weg von Ost nach West*.[30] Twenty-six years before this book was published, the well-known German art critic Franz Servaes wrote an article about Epstein. Servaes resided at this time in Vienna and was well connected within the art scene of the capital.[31] This article was part of a larger anthology with the name *Jüdische Künstler* which was edited by the famous Zionist Martin Buber. For the biographical side notes Servaes used quotes from a letter Epstein provided him for the purpose of writing his article.

> About the career and life of Jehudo Epstein so far, we let the artist speak for himself. He is writing the author of these lines (…) As I had learned enough of drawing, I decided to study at an academy and because it was particularly difficult for jews to study in St. Petersburg, I decided to go to Vienna. With fifteen

---

28 Josef Engelhart, *Ein Wiener Maler erzählt. Mein Leben und meine Modelle* (Wien: Andermann, 1943), 48–49.

29 Josef Engelhart was the child of the butcher Josef Anton E. (1838–1900) and Maria Apfelthaler (1842–1933). He married 1895 Dorothea (Doris) Mauthner (1871–1967), the daughter of industrialist and brewery owner Karl Ferdinand Mauthner v. Markhof. – See Wien Geschichte Wiki, https://www.wien.gv.at/wiki/index.php?title=Josef_Anton_Engelhart (accessed 30 April 2018).

30 See Jehudo Epstein, *Mein Weg von Ost nach West. Erinnerungen* (Stuttgart: J. Engelhorn & Nachfahren, 1929).

31 As an example, he was part of the organizing team in 1903 of the so called "Kunstwanderungen". This was a series of guided tours through art collections of the aristocracy and their palaces. See Kaiser, *Wiener Diskurs,* 52–53.

> guilders in my pockets, without speaking any German, without any sort of recommendation, I came to Vienna and lived through hard times.[32]

Besides his first steps in Vienna, he managed to establish himself as a painter. His breakthrough came by winning a prize sponsored by the Michael Beer foundation. With this support he could travel to Rome. About the results of this journey he wrote:

> The works of the last eighteen months, I spent in Rome, I presented to the general public through a solo-exhibition at the Künstlerhaus, from November till December, 1901.[33]

The first version of his biography in the ÖBL was published in the mid-1950s. Because of that, the initial information about the artist's life which can be found in the biography was rather scarce:

> He has been member of the Künstlerhaus since 1902, made several journeys and won international and national awards. The ashes of the emigrant were buried in Vienna 1949.[34]

The newly updated version from the year 2015 is much longer, goes much more into detail and includes the circumstances of his emigration:

> In December of 1934 E. travelled out of financial reasons to Johannesburg (his furniture from his studio and his apartment in Vienna was stored in the factory of Bernhard Altmann, one of his close friends, in 1936) and he was able to establish himself with the painting of portraits. (…) After the Anschluss in 1938, his temporary stay changed into emigration.[35]

After the first appearance of the avant-garde in Vienna, Epstein became one of its most rigid critics. In several articles he criticised art historians, mostly Hans Tietze, for their support of the younger and more international oriented generation of artists.[36] Besides all controversies, the art historian Max Eisler spoke at the dinner which was given on the occasion of Epstein's journey to Palestine in 1927.

---

32 See Franz Servaes, "Jehudo Epstein", in *Jüdische Künstler,* ed. Martin Buber (Berlin: Jüdischer Verlag, 1903), 163–165.

33 Ibid., 165.

34 N. N., "Epstein, Jehudo, Maler", in *Österreichisches Biographisches Lexikon 1815–1950*, Bd. 1 (Lfg. 3, 1956), 258–259.

35 Rene Schober, "Epstein, Jehudo (1870–1945), Maler und Illustrator", in *Österreichisches Biographisches Lexikon. Online Edition*, 30.11.2015, http://www.biographien.ac.at/ (accessed 2 February 2018).

36 See Kaiser, *Wiener Diskurs*, 136–137.

> The man, of whom I want to speak, is certainly not an admirer of my profession. (…) With this I come to the third and last encounter with Epstein, the encounter with his jewish soul. Out of his autobiography, she can be understood at its clearest (…) How he managed to become successful as a painter out of the poorest beginnings through hard work, can be cause for admiration. On this difficult path he relentlessly worked also on his humanity and didn't lose his deeply rooted jewish idealism, this is connecting him only closer with our group of people and therefore imposes the duty on us to honor a farewell solemnly, no, cordially.[37]

Eisler's speech can be read as a gesture of reconciliation with Epstein. This is certainly based on commonalities such as Jewish life in Vienna. It is also possible that he was moved by the fact that the departure of an artist which is important for his religious community was imminent.

2.3 Biographical Data and Historical Network Research

As these few examples already show, the reasons for migration are very diverse. The authors of the ÖBL biographies used a wide range of historical sources because they had to cover an artist's life completely from birth until death. These materials could be archival resources, newspapers, journals, ephemera or scholarly work. Biographical data collected from the ÖBL biographies always reflects this multitude. This implies that this data is also incomplete and historically shaped.[38] In contrast to that historical network research is built upon the systematically exploitation of one singular type of resource such as exhibition catalogues. One can therefore ask the question whether biographical data can be of any value to serious academic research. While this is a justifiable critique it has to be noted that APIS data allows for a macro perspective that wouldn't be possible based on heterogeneous data such as primary resources. To allow the proper distinction between historical network research and analysis of biographical data with methods from network analysis in the future, the term "biographical network research" should be introduced here.

Do any further applications for biographical data exist? This question can only be answered in respect of the emerging research field of digital art history.

---

37 Max Eisler, "Jehudo Epstein. Eine Tischrede", *Menorah. Jüdisches Familienblatt für Wissenschaft, Kunst und Literatur* 5, no. 3 (März 1927): 158–167.

38 Matthias Schlögl et al., "Biographik in den Digital Humanities – Kritische Bestandsaufnahme und quantitative Analysemöglichkeiten am Beispiel des Österreichischen Biographischen Lexikons 1815–1950", in DHd 2018. Kritik der digitalen Vernunft. Konferenzabstracts. Universität zu Köln, 26. Februar bis 2. März 2018, ed. Georg Vogeler (2018), 360–363. http://dhd2018.uni-koeln.de/wp-content/uploads/boa-DHd2018-web-ISBN.pdf (accessed 23 November 2018).

The application of digital methods is of particular interest when it comes to answering spatial research questions:

> To be coherent and comprehensive, art historical research needs to cover many scales (from the global to the local), to combine monographic and serial data, and take into account the plurality of cultural and artistic transfers that occur through both the creative process as well as its reception, as exemplified in work on cultural transfers, and the comparative approaches of cultural and social history.[39]

This quote originates from two scholars working on the Artl@s-project. Within this framework, serial data such as exhibition catalogues of the Paris Salon among others has been collected and entered into a database. The retrieved data was the basis for the visualisation of networks and maps. The systematic exploitation of these primary sources is a commonality with other projects in the realm of digital art history.[40] Since networks are based on vectors collected in the form of tabular data, the possibilities of filter criteria and attributes are limited due to technical standards. In APIS most of the information is captured in relations between entities rather than attributes of nodes. This implies that filtering of nodes needs to be done partly on relations to other nodes, something that is currently only insufficiently possible in the General User Interface (GUI). But on the other hand, the theoretical assumptions based on network structures are numerous. It therefore would make sense to use biographical data to bridge the gap between ideal conception and the possibilities available to us. So, it would not only be possible to read out of network structures the importance of an actor, but also ask questions about which places a person visited and institutions he or she attended during the education and career path. These are questions of particular interest with regard to cultural transfers and social history of art.

---

39 Béatrice Joyeux-Prunel, Catherine Dossin, and Sorin Adam Matei, "Spatial (Digital) History: A Total Art History? – The Artl@s Project", *Visual Resources. An International Journal of Documentation* 29, no. 1-2 (12 March 2013), 48, DOI: 10.1080/01973762.2013.761119.

40 The following list gives a brief overview about which research and exhibition projects has been using this methodology since 2010: Informatics. Data driven analyses of art related data | Impuls-Bauhaus | Netzwerk Will Grohmann | Hagenbund. An international artist network | Avant-Garde Network in the Twenties. The International Connections of Lajos Kassák and his journal MA | Artnet: Modern and Contemporary Artist Networks, Art Groups and Art Association | Artl@s | Map Tap: Mapping the Antwerp Brussels Oudenarde tapestry complex via social network analysis 1600–1700 | Exhibitium project | Exhibitions of Modern European Paintings 1905–1915 | Cornelia – a database for analysing creative communities.

3 Beyond *A Network Framework to Cultural History*: analysing migration patterns based on APIS data

Three types of analyses can be done with this biographical data at hand: a visual exploration of migration and mobility through a) networks, b) maps and c) alluvial diagrams. All three methods should be discussed in the following two sections.

As mentioned above, Schich and his colleagues concentrated in their study on fundamental facts as provided by the databases. This can be done with the APIS data as well. But before the network is analysed some basic operations are necessary. First of all, the data export gives the user a two-mode network which consists of two types of nodes: namely person and places. In our case, the relations between the entities are "was place of birth" and "was place of death". Vienna is clearly the central hub in this network. A quick first glance suggests that migration primarily took place from the periphery to the centre of the Habsburg Monarchy. This assessment coincides with findings presented in the literature on historical migration research. For instance, Heinz Fassmann states that in the 19th century in Europe in general and in the Habsburg Monarchy in particular spatial mobility in the majority of cases meant rural-to-urban migration within the borders of a state.[41]

In the next step it is transformed to a one-mode network which consists of places only. After completing this step, each edge represents a person connected to more than one place through the given relation types. Isolated nodes represent cases were a person was born and has died at the same place. At the same time as the amount of data has decreased, new interpretations become possible.[42] The comparison of these two graphs is shown in figure 4.

41 Heinz Fassmann, "Europäische Migration im 19. und 20. Jahrhundert", in *Migrationen. Globale Entwicklungen seit 1850* (= Globalgeschichte und Entwicklungspolitik 6), eds. Albert Kraler et al. (Wien: Mandelbaum Verlag, 2007), 38.

42 "The multiplication can also be applied on two-mode matrices. From this, two-mode networks can be transformed into interesting one-mode networks." See Jürgen Lerner, "Beziehungsmatrix", in *Handbuch Netzwerkforschung*, eds. Christian Stegbauer and Roger Häußling (Wiesbaden: VS Verlag, 2010), 361.

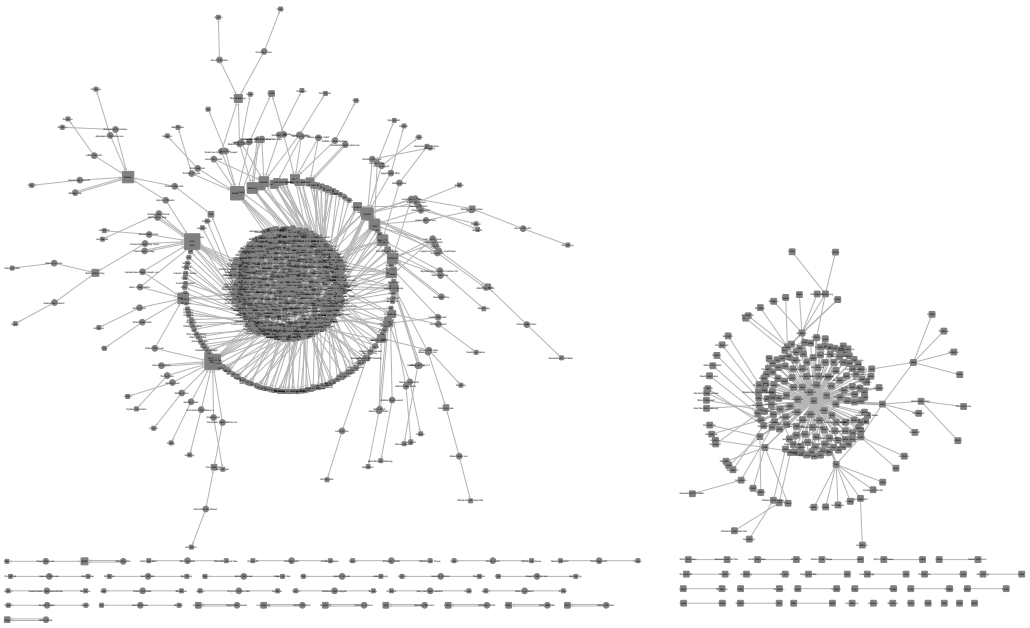


Figure 4. These two graphs show the network after the data export (left) and after the transformation to a one-mode network (right). Originally it consists of 769 nodes (265 of them are places) and 1007 edges. In its new form there are only the places with a fraction of the edges. On the other side the edge weight has increased. (graphic of the authors)

Edges can be read as migration from place A to place B. So, the newly drawn graph on the right shows, for example, how often persons have moved between those places. This is depicted through the width of an edge in this network. One advantage of the APIS data is that every place has its geographic coordinates through the linking to the norm data of the GeoNames database. This makes it possible to display the same network on a map (figure 5).

What are the most common migration paths in this network? One way of answering that question would be to count the multiplicity or numbers of the edges in this network. All values are computed through the visualisation software *visone*. In descending order, the most frequent paths outgoing from Vienna are therefore: Graz (12, meaning that 12 out of 506 persons were born in Vienna and have died in Graz), Prague, Klagenfurt and Budapest (10), Linz, Mödling and Lemberg (8). And what are the most distant migration paths in this network? The actual distance becomes better understandable in this view of the network data. Calculated from Vienna, the most distant places are New York City (6,795 km), Södertörn (1,707 km) and Syrakus (1,240 km).[43]

[43] The distance is calculated with the web service luftlinie.org. See Stephan Georg, Luftlinie.org, https://www.luftlinie.org/ (accessed 12 March 2018).

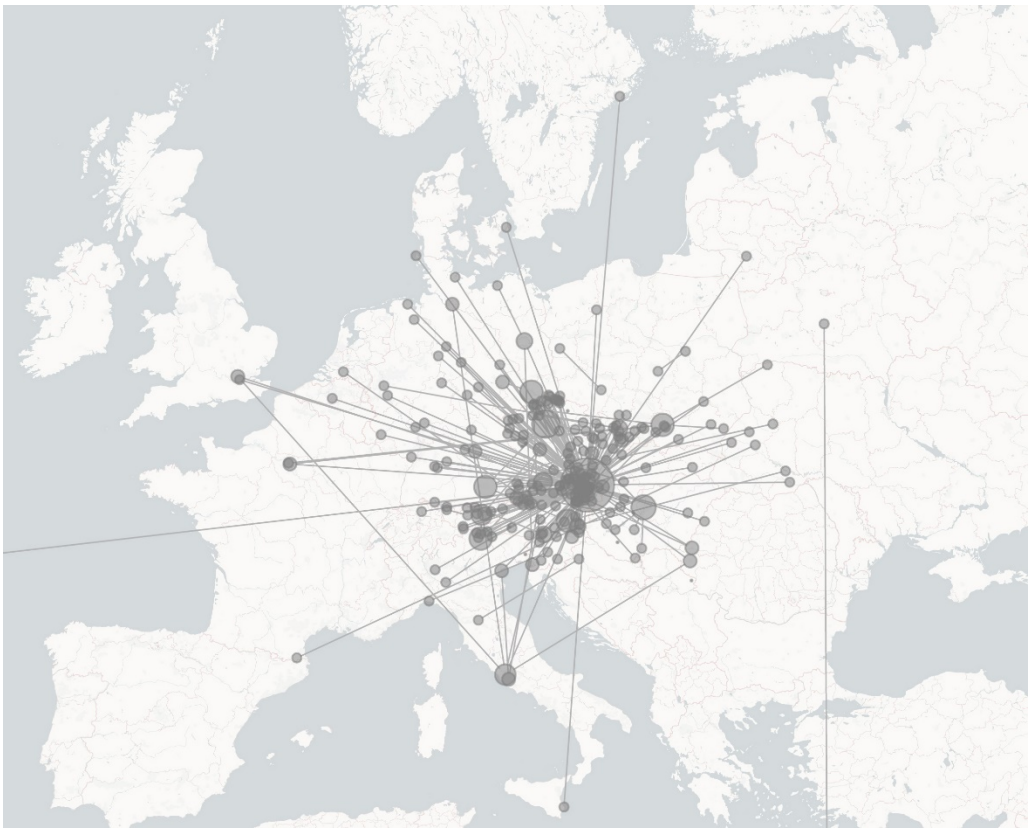


Figure 5. This map shows a similar migration network to *A Network Framework to Cultural History*. It is based on the birth and death places of artists in the ÖBL ("Künstlerhaus" data collection). Two edges lead beyond this map tile: that of Jehudo Epstein and Josef Urban (1872–1932). Epstein's emigration was mentioned before in this paper. The architect Urban fled 1908 after financial problems from Vienna to the US. In New York City he could establish himself professionally again.

But let us return to the biographies and actors of this network. The painter Eduard Ender (1822–1883), as one example picked out of the whole, was born in Rome and died in London. His biography in the ÖBL describes his life as following: "Pupil of his father Johann Nepomuk and the academy of fine arts Vienna, last educated in Paris. Mostly rejected from contemporary critics."[44] His migration path is easy to grasp in the cartographic representation because of its uniqueness in this dataset. But two steps between Rome and London as mentioned in his biography are missing. First of all, his time as a student of the Academy of Fine Arts in Vienna is implicit information. Following the annotation guidelines defined by the research group it therefore can't be

44 N. N., "Ender, Eduard, Maler", in *Österreichisches Biographisches Lexikon 1815–1950*, Bd. 1 (Lfg. 3, 1956), 247.

annotated in the text. Secondly, for completing his studies he went to Paris. This is given in the text, but it was annotated with the relation type "was educated in". So, it is clearly not displayed on this map. Other edges in this network which connect Vienna with distant places like Syrakus in the South of Italy are obvious indications for migration but likewise rare examples.[45]

The fact that Vienna, as the capital of the Habsburg Monarchy, is the central hub in this network is a characteristic result out of the sampling method. The GeoNames feature code "PPLC" stands for "capital of a political entity". Like other capital cities such as Prague, Budapest or Rome someone might think that they are the most prominent nodes. But by measuring the centrality value degree it becomes clearer that these places are not the only feature category within the top-10 of this network. Munich and Innsbruck, for example, both cities of regional importance follow shortly thereafter in the ranking.[46] Munich was not only important because of its relatively close distance to Vienna but through a similar set of cultural institutions. The Academy of Fine arts Munich was very popular under artists from all over the Habsburg Monarchy, the Glaspalast provided a generous exhibition platform and the artist association "Münchner Künstlergenossenschaft" offered possibilities comparable to the Künstlerhaus. Innsbruck however was the catchment area for people of the alpine regions, especially the southern and eastern parts of Tyrol and Vorarlberg in the West. Its importance lies in the fact that it has been the cultural and institutional centre of a transit region. To the south there were the art academies of Florence and Milano, to the north, as it was already mentioned, the academy of Munich and to the east the Academy of Fine Arts Vienna. Apart from Vienna, mobility between cities and villages is numerous but the reoccurrence of such paths is rather rare.

A subset of this dataset are places where a person both was born and has died. They are displayed as single points on the map. This case is not very typically for the members of the Künstlerhaus as the dataset shows. It regards just seven out of 506 persons. But in some of these cases it is still possible to find a migration story. The painter József Rippl-Rónai (1861–1927) is such a candidate. He was born and died in the village of Kaposvár in Hungary, but his ÖBL-biography is full of places which he visited in context of his work and study trips:

---

45 In this case it is the biography of the Orient painter Franz Xaver Kosler (1864–1905). He was a pupil of the painter Carl Leopold Müller and followed his teacher in the choosing of his topics. Several times he exhibited in exhibitions of the Künstlerhaus, in the Glaspalast Munich and in London. Finally, after a long-lasting illness he decided to move to the city of Syrakus where he died in December 1905. See Anonymus, "(Walter Kosler †.)", *Neues Wiener Tagblatt* 39, no. 355 (24 December 1905), 16.

46 GeoNames categorize both cities as feature category "seat of a first-order administrative division" (PPLA).

After graduating from grammar school, he was educator of the duke Zichy's family and started painting. Since 1884 he studied at the Academy of Fine Arts Munich and was the pupil of Herterich. From 1887 on he worked in Paris for Munkacsy. (...) Since 1892 he shared a studio with the sculptor Maillol in Neuilly. Around 1900 he returned to Hungary and lived some years in Budapest and since 1902 lived in Kapsovár.[47]

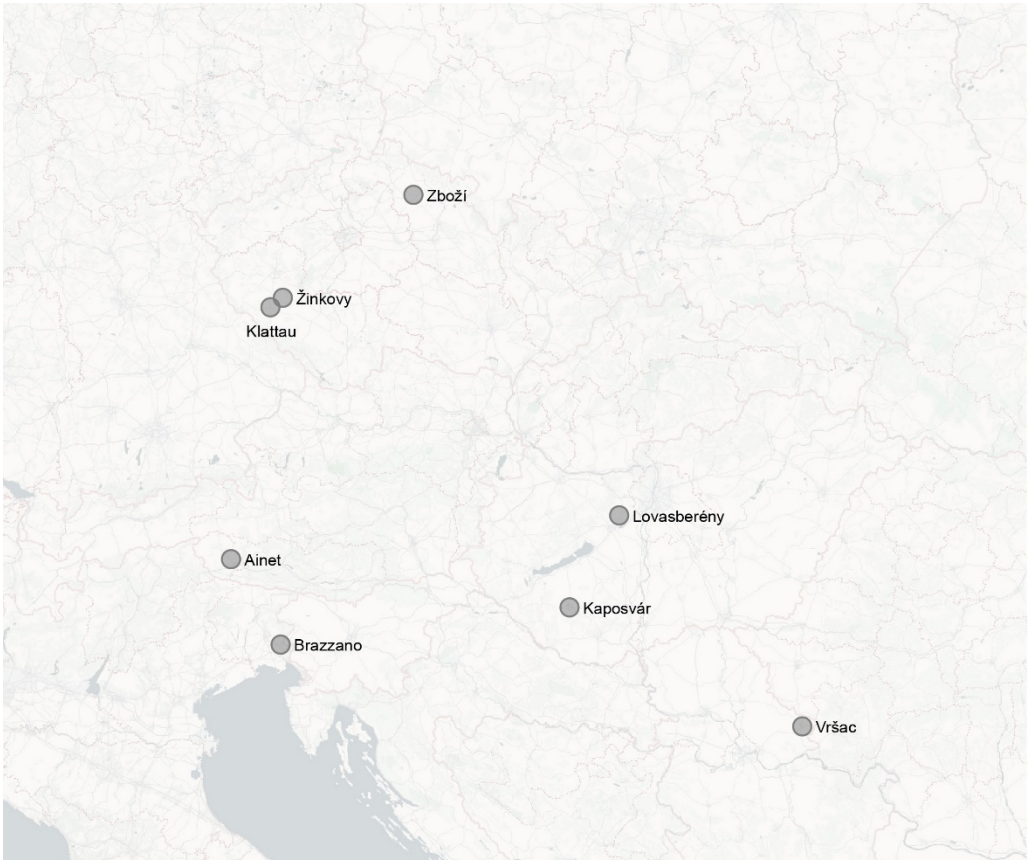


Figure 6. This map shows the places where artists were both born and have died. It is based on the data manually annotated in the "Künstlerhaus" data collection.

If these cases where the places of birth and death of one person are identical are visualized as a network in figure 6, every single point would represent an isolated node and therefore wouldn't be connected with any of the other places mentioned in the biography. So, when simply comparing a person's place of birth and place of death, obviously, lots of information on spatial mobility and migration occurring during the person's lifetime can be lost. This clearly shows

47 Benda Kálmán, "Rippl-Rónai, József, Maler und Graphiker", in *Österreichisches Biographisches Lexikon 1815–1950*, Bd. 9 (Lfg. 42, 1985), 174.

the usefulness of extending the analysis to further whereabouts as places of work or places of a journey.

4 Migration patterns and perspectives for data analyses

Based on the presented analyses of this article so far, it is possible to develop further research tracks. It can be summarized that for most of the members of the Künstlerhaus the city of Vienna is the most important place for education and work. Besides their place of birth, their spatial mobility in context of education and their travel activities, for work and exhibitions they often moved or returned to the capital of the Habsburg monarchy (figure 7). This explains the central position of Vienna in the person-place network.

For the following data analyses and identification of migration patterns, we used so called alluvial diagrams. Firstly, these diagrams are based on the biographical data, precisely on the persons' relations to places as well as to institutions which are mentioned within the biographical articles of the ÖBL. The institutions are also a key factor within the conceptual understanding of "Cosmobilities".

> In such a scenario [note: the historically shaped borders of East Central Europe are meant] it is no longer the implementation of the nation-state model but the dynamic and complex entanglements between regional, national, imperial and global levels that are put into the analytical focus. This, in turn, connects to our understanding, which views 'mobilities' as both infrastructural prerequisites for actors going beyond borders and influential factors in (re-)shaping and (re-)negotiating boundaries.[48]

Secondly, it was necessary to distinguish between those relations commonly associated with migration in a narrower sense (place: "was place of education", "was place of work"; institutions: "was student at", "was professor at") and those representing only short-time stays (place: "was place of a study trip") or weak ties to institutions (institutions: "was member at", "was delegate at"). The resulting diagram shows the migration paths step-wise and according to their relation to a certain geographical category (various toponyms). These categories can either stand for single places of the size of a capital city (GeoNames feature code "PPLC") such as Vienna or Prague, for larger cities (GeoNames feature code "PPLA") or several smaller places (GeoNames feature code "PPLA2", "PPLA3" and "PPLA4").[49] From the left to the right, a path can be read as a migration from one category or place to another, relating to the first four steps

48 Panter, Paulmann, and Szöllösi-Janze, "Mobility and Biography", 10.

49 In those cases, where it is not likely that the reader associates a place name with the right category of the alluvial diagram, the category is written afterwards and behind curved brackets.

of a person's migration process in a narrower sense, meaning longer-term spatial mobility.

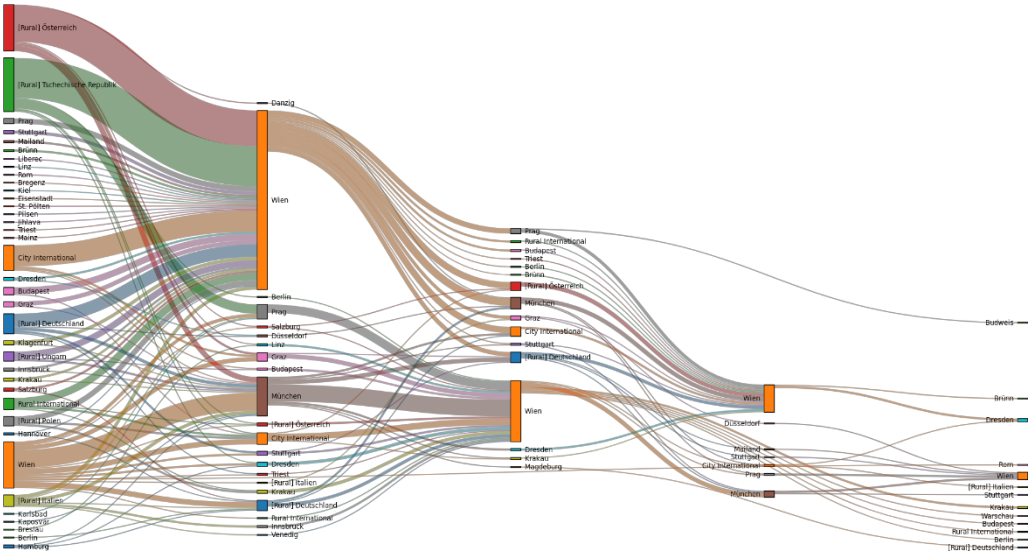


Figure 7. The alluvial diagram shows the migration steps in the order as they appear in the artists' biographies incorporated in the "Künstlerhaus" data collection. In this visualisation, nearly all relations to places and to institutions are considered to give the most complete picture of the persons' spatial mobility. Relations which represent only a short stay in context of a study trip or weak ties to institutions such as "was member of", "was delegate of" etc. have been excluded. All places below the size of the GeoNames feature code "PPLA" (seat of a first-order administrative division) are subsumed under the term "[rural] name of country". Cities which are neither connected to the neighbouring countries of Austria (Hungary, Czech Republic, Italy and Germany) nor have a distinctive seize in the sample are displayed as "City international" or "Rural international". For the sake of simplicity only the first four migration steps were taken into account. If a person has more than one step at the same place this relation is also excluded in this visualisation. (graphic of the authors)

4.1 Potentials of Automatic Relation Extraction

But what different questions could be answered if there wouldn't be the limitations of the time-consuming work of annotating artists' biographies manually? For this, it would be necessary to go beyond the manually collected data. In order to do so, the application of automatic information retrieval techniques would be necessary. Therefore, in APIS a system for Information Extraction has been developed. After the identification of Named Entities, first Entity Linking, then Relation Extraction is performed. The system uses open-source tools to derive structured semantic information from unstructured biographies, i.e. relations between Entities that have been identified and linked to Linked Open Data resources like GND and GeoNames. For the relation extraction, the open-source software *General Architecture for Text Engineering*

(GATE)[50] is currently tested. Regular expressions are run on the annotations in order to identify linguistic patterns that express a given relationship. If the application finds a string that corresponds to the pattern, it automatically assigns a new annotation to it, which describes the type of relation that holds between the entities in the given section of the text. In the ÖBL the name of the person is often omitted in order to avoid redundancy, which results in relations that cannot be annotated by other tools such as IEPY[51], as one of the entities in the relation (i.e. the person who performed the action) is missing in the text. Due to this, the regular-expression-based approach implemented in GATE is more applicable for ÖBL data. Another approach which is currently developed in APIS uses the manual annotations for the training of machine-learning algorithms. As it has been pointed out, both methods are still under development and will require more time for further training of the systems and the evaluation of the output.[52]

4.2 The subsamples of Tyrolean and Bohemian Artists

Nevertheless, for demonstrating the potentials of these methods, a comparison between two population groups and migration patterns based on subsamples of the manually annotated biographies of the "Künstlerhaus" data collection shall be made. To pick an example out of the almost 2,500 artists' biographies, the group of Tyrolean artists has been chosen. They are defined as persons who were born or have died within the boundaries of the federal province of Tyrol in the borders of 1910. The "Künstlerhaus" data collection contains 30 persons (21 relations "was place of birth"; 15 relations "was place of death"), to whom this specific characteristic is applicable. The distinction between places which can be found within the historical border of Tyrol and those which are located elsewhere is made with *HistoGIS* data, a project developed at the Austrian Centre for Digital Humanities.[53] In total 326 Tyrolean artists can be found in the ÖBL by using this method. The Tyrolean artists from the collection thus represent 9,20 % of the whole group. This group is of particular interest from an art historical point of view because of two reasons. Firstly, geographically Tyrol

50 *GATE*, https://gate.ac.uk/ (accessed 25 April 2018).

51 *IEPY*, https://github.com/machinalis/iepy (accessed 25 April 2018).

52 Matthias Schlögl and Katalin Lejtovicz, "A Prosopographical Information System (APIS)", in *BD-2017. Biographical Data in a Digital World 2017. Proceedings of the Second Conference on Biographical Data in a Digital World 2017, Linz, Austria, November 6-7, 2017*, eds. Antske Fokkens et al. (2018), 53–58, http://ceur-ws.org/Vol-2119/paper9.pdf (accessed 23 November 2018).

53 In *HistoGIS*, shapefiles, which can describe borders, rivers, shapes of cities or districts, are drawn on the basis of historical maps in order to answer geospatial questions. The application can distinguish for a certain time slice, if the marker, which always represents a certain place on a map, lies within the specific outline or not. On this basis, locations within the historical border of Tyrol can be identified. See *HistoGIS*, https://histogis.acdh.oeaw.ac.at/ (accessed 13 November 2018).

is positioned in the far west of the Habsburg monarchy. Secondly, Innsbruck as the regional capital didn't have its own art academy.[54] It thus can be assumed that for those persons interested in studying art, migrating to a larger city was inevitable. One question would be, which places are of importance for the Tyrolean artists? How far do they migrate into the east? With the end of the supremacy of academic styles and the appearance of the avant-gardes at the end of the 19th century, the networks and mobility of artists underwent significant changes. Deduced from that, how changed mobility for the generations of the artists born in the 1860s and later compared to the older generations?

A first analysis of the places of birth shows that – with the exception of Tony Grubhofer (1845–1935) and Leo Putz (1869–1940) – most of the Tyrolean artists in this small sample of 21 persons moved from one place to another at least once in their lifetime (figure 8). The diagram makes it clear that it is important to include more than one relation type and entity type. Otherwise the view on the data can lead to false assumptions. For example, the life and work of the artist Leo Putz can only be described as transnational.[55] He has been member not only of the Künstlerhaus but also of the German artist group "Die Scholle" and the Vienna Secession. Besides his stays in Munich and Vienna, he also travelled to Brazil ("City international") and lived there between 1929 and 1933. Tony Grubhofer on the other side was educated in Munich and Vienna, made study trips to foreign countries, worked for some time in Bolzano ("[Rural] Republic of Italy") and finally became director of the "Staatsgewerbeschule" in Innsbruck.[56] For both biographies, Munich plays a central role as a place for their education and work.

---

54 The highest educational institution for the art in Tyrol, which is listed in the *Hof- und Staatshandbuch der Österreichisch-Ungarischen Monarchie für 1900*, was the 'Staatsgewerbeschule' in Innsbruck. Prominent artists such as the architect Clemens Holzmeister were teaching there. The city of Hall in Tirol had a professional school for carving and carpentry. St. Ulrich in Gröden and Bozen, both situated in Southern Tyrol, had professional schools for woodworking. The professional school in Laas, also situated there and famous for its marble quarry, was specialised in stoneworking and sculpture. All these institutions can be found in the artists' biographies of the ÖBL.

55 Erich Egg, "Putz, Leo, Maler", in *Österreichisches Biographisches Lexikon 1815–1950*, Bd. 8 (Lfg. 39, 1982), 348.

56 N. N., "Grubhofer, Tony, Maler", in *Österreichisches Biographisches Lexikon 1815–1950*, Bd. 2 (Lfg. 6, 1957), 85.

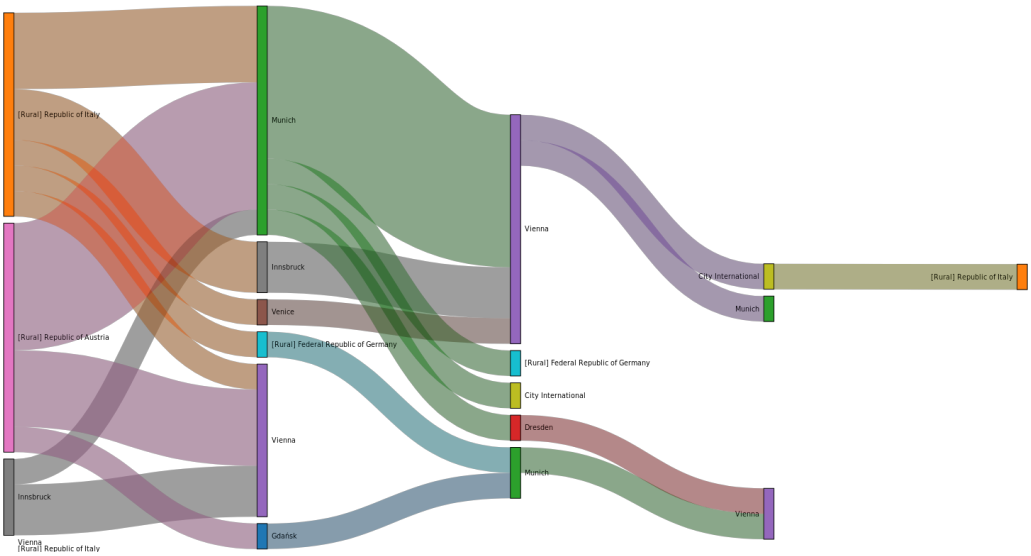


Figure 8. The alluvial diagram shows the subsample of persons born within the border of Tyrol. In this case all places of Southern Tyrol such as Bolzano, Ainet or Sarnthein are subsumed under the category "[Rural] Republic of Italy" (categorized according to modern borders). The places categorized as "[Rural] Republic of Austria" are places of birth in Northern (or Eastern) Tyrol such as Kitzbühl, Reutte in Tirol, Hall in Tirol or Schlitters. Nearly two thirds of this small subsample come from today's Austrian federal state Tyrol. One third was born in Southern Tyrol. It can be noted that artists who were educated in Innsbruck or Munich moved at least one step further compared to those who directly went to Vienna. Conversely, the other group could be characterised as being spatially less mobile. (graphic of the authors)

As a next step, this mobility pattern is extended by adding the group of artists who had died in Tyrol (figure 9). This aspect, on the other hand, additionally visualises artists who immigrated to Tyrol. In principle, the migration patterns of these 30 persons who were born or have died in Tyrol are quite similar to the 21 persons with their place of birth in Tyrol.

A second promising subsample can be found in those persons who were born or have died in the region of Bohemia.[57] In the "Künstlerhaus" data collection there are 52 persons which meet these criteria. They represent 16 % of all 325 Bohemian artists in the ÖBL. Besides Vienna, Prague is the most important node within the person-place network (figure 10). Historically seen, Bohemian art is deeply rooted in the tradition of Austrian art. Famous Austrian artists of Bohemian origin such as Franz Rumpler (born in Tachov) and Josef Matthias Trenkwald (born in Prague) were teaching at Austrian art academies

57 The persons were selected by using *HistoGIS* data one more time. See *HistoGIS*, https://histogis.acdh-dev.oeaw.ac.at/shapes/shape/detail/3604 (accessed 14 November 2018).

in the 19th century as well as many artists born in Austria such as Friedrich von Amerling or Josef von Führich were studying at the art academy in Prague.

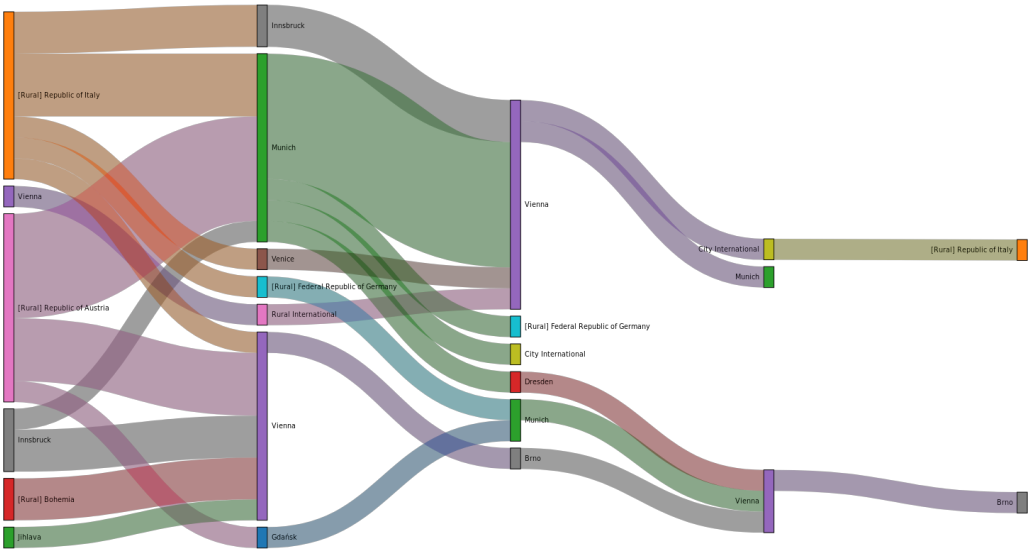


Figure 9. The alluvial diagram shows the subsample of persons who were born or have died within the border of Tyrol. Including the other relation type ("was place of death"), makes it possible that those artists who immigrated to Tyrol are also visualised (e.g. Ferdinand Kruis, Emanuel Stöckler or August Prokop). Emanuel Pendl (Meran to Vienna) and Karl von Blaas (Nauders to Vienna) belong to the group of emigrants. (graphic of the authors)

Artist associations situated in Vienna like the Künstlerhaus, the Secession or the Hagenbund followed this tradition in fostering these links through exhibiting Bohemian artists and vice-versa (e.g. in the S.V.U. Mánes). Art dealers and galleries built upon these networks their business beyond the borders.[58] In the year 1918, as the end of the Habsburg Monarchy was sealed, and the Republic of Austria was founded, the art historian Hans Tietze wrote an article in a journal for historical studies named *Österreich* about the future of Austrian art history. There, he addressed the question of the identification of Austrian artists aptly.

> In a certain sense, an artist belongs not only to the people from whom he descend, but also to the people who offered him the opportunity to develop his skills and in whose art life the trace of his days on earth lives on; thus Tilgner cannot be separated from Austrian art and Pettenkofen not from Hungarian art. Throughout the XIXth century, the rich interaction continues undiminished by immigrant or emigrant artists, by students who learned from their neighbours,

58 The galleries Miethke, Würthle and Neue Galerie were exhibiting Bohemian artists among others. The first secretary of the Vienna Secession Franz Hancke (1873–1909), to give one example, used this network to set up his own business 1907 in Breslau.

by masters who exerted influence through their work, and there the net of all these numerous threads should not have densified to inextricability.[59]

After illustrating these numerous ties between Vienna and the region of Bohemia, the question arises how Vienna is represented in the biographical data of this subsample. Are these artists mainly migrating from Bohemia to the capital city of the Habsburg monarchy? Which role plays Prague? Following the analyses of the group of Tyrolean artists, another question can be asked. Is it possible to identify a group of significant size which moved differently to the majority (e.g. like those artists from Southern Tyrol) or not?

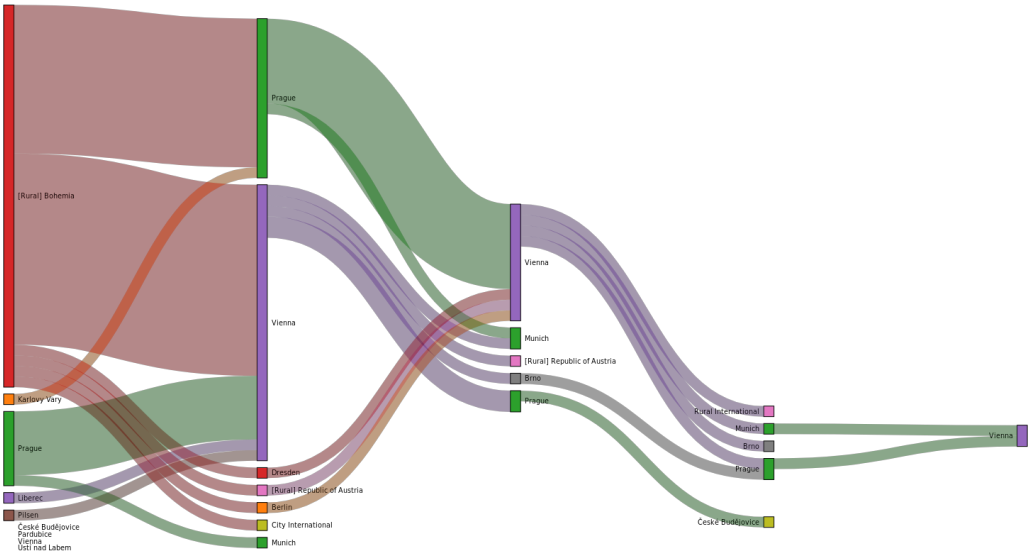


Figure 10. The alluvial diagram shows the subsample of persons who were born or have died in Bohemia. More than two thirds of this group were born on the countryside of Bohemia ("[Rural] Bohemia"). Besides Prague, artists of this subsample were born in the larger cities of Karlovy Vary, Liberec and Pilsen. Munich, as it was expected, played a much lesser role than in the subsample of Tyrolean artists. A larger share of this subsample migrated to Prague and stayed there or moved forward to Vienna in the next step. The other part of those artists born in rural places as well as artists born in Prague already migrated to Vienna in a first step, the vast majority of them remained there. From the first group, and this is an interesting fact, the data shows a migration to Vienna only as a third step. Some of them returned later once again to Prague. This pattern, of moving forth and back, can be noted as significant for the group of Bohemian artists. (graphic of the authors)

Deduced from Tietze's quote, it must be noted that biographical data can already be used for the identification of transnational actors such as Victor

59 Hans Tietze, "Das Problem der Österreichischen Kunstgeschichte", *Österreich. Zeitschrift für Geschichte* 1 (1918/1919), 59–60.

Tilgner or August Pettenkofen.[60] By including the analyses of migration patterns, researchers can distinguish not only between individual actors, but also between different types of spatial "mobilities". This has been demonstrated rudimentarily by the subsamples of the Tyrolean artists and the Bohemian artists. At this point it is possible to connect this research with the concept of "Cosmobilities". According to Panters understanding of this concept, Sarah Panter et al. write that it has to connect "local rootedness, transcultural orientations and global entanglements with insights on mobility" in order to overcome "one single master narrative".[61] In this point the biographical networks, maps and migration patterns presented in this article can be seen as an addition as well as a possible starting point for future (cos)mobility research.

5 Conclusion

Besides the fact that the methods for the retrieval of biographical data are still under development, recent achievements in this field and best practice examples have to be discussed.[62]

As it is mentioned above, the sample was compiled according to two criteria: Firstly, the biography in the ÖBL had to describe the life of an artist. Secondly, the membership in the Künstlerhaus had to be proven. Because of this sampling method, Vienna, compared to other cities, plays a significant role in the migration networks of these artists.

In the field of historical network research, however, first concepts for the application of biographical data are emerging. One approach would be to follow the idea of *A Network Framework to Cutural History*. To achieve this, the biographical data must be enhanced in order to examine the migration patterns in detail, both historically and statistically. In contrast to historical network research a wider range of historical sources are used to write the biographies. We therefore remind that by working with biographical data their genesis always has to be considered. Another possibility would be to use this data in a different application and enrich catalogue data as it is collected in exhibition databases (e.g. Hagenbund, Artl@s, Exhibitions of Modern European Paintings etc.). In this scenario, questions of cultural transfers as they are asked by spatial art history or social history of art could be answered.

---

60 This idea has already been put forward at the course of the presentation "Tracing transnationality through a biographical dictionary" at the conference ESSE 2018 (seminar "Transnational Biographies in Europe", Masaryk University Brno).

61 Panter, Paulmann, and Szöllösi-Janze, "Mobility and Biography", 10.

62 See Antske Fokkens, Niels Ockeloen, and Serge ter Braake, *BiographyNet*, http://www.biographynet.nl/ (accessed 27 April 2018); Jonas Kuhn and André Blessing, *Textual Emigration Analysis*, http://clarin01.ims.uni-stuttgart.de/tea/ (accessed 27 April 2018).

The automatic retrieval of information from the corpus of ÖBL biographies requires the testing of a variety of tools and methods. The success, however, depends strongly on the adaptation of corpus-specific characteristics such as the language used, or the content described in the biographies. It is obvious that those approaches will play an important role in the near future. Two subsamples of the "Künstlerhaus" data collection, namely the Tyrolean and Bohemian artists, were chosen to develop additional research tracks and to test what assumptions can be drawn from such data analyses. Both subsamples of the "Künstlerhaus" data collection represent 10 to 16 % of the overall share of these two population groups within the ÖBL, so patterns can already be identified. The actual data analyses will become possible at that moment all biographies have been annotated either manually or automatically.

Nevertheless, the aim of this paper was to show how biographical data can be generated and how methods of historical network analysis as well as historical migration analysis can be applied. In this way, the potential of this particular kind of resource should be demonstrated and become visible for future research within the realm of digital art history.

References

Anonymus. (Walter Kosler †.), in: Neues Wiener Tagblatt 39, no. 355 (24 December 1905): 16.

Bernád, Ágoston Zénó. "Das Österreichische Biographische Lexikon 1815–1950. Ein Nachschlagewerk im Wandel der Zeit". In *Hortus Amicorum. Köszöntőkötet Egyed Emese tiszteletére [Festschrift für E. Egyed]*, edited by Katalin Ágnes Bartha, Annamária Biró, Zsuzsa Demeter, and Gabriella-Nóra Tar, 457–467. Kolozsvár: Erdélyi Múzeum Egyesület, 2017.

Bernád, Ágoston Zénó, Christine Gruber, and Maximilian Kaiser, eds.. *Europa baut auf Biographien. Aspekte, Bausteine, Normen und Standards für eine europäische Biographik*. Wien: new academic press, 2017.

Ceserani, Giovanna, Giorgio Caviglia, Nicole Coleman, Thea de Armond, Sara Murray, and Molly Taylor-Polesky. "British Travelers in Eighteenth-Century Italy: The Grand Tour and the Profession of Architecture",*American Historical Forum* 122, no. 2 (April 2017): 425–450. DOI: 10.1093/ahr/122.2.425.

Cosmotool. https://cosmotool.de.dariah.eu/cosmotool/personsearch/ (accessed 23 November 2018).

Eisler, Max. Jehudo Epstein. Eine Tischrede, in: Menorah. Jüdisches Familienblatt für Wissenschaft, Kunst und Literatur 5, no. 3 (März 1927), 158–167.

Engelhart, Josef. Ein Wiener Maler Erzählt. Mein Leben und meine Modelle. Wien: Andermann, 1943.

Epstein, Jehudo. Mein Weg von Ost nach West. Erinnerungen. Stuttgart: J. Engelhorn & Nachfahren 1929.

Fassmann, Heinz. "Europäische Migration im 19. und 20. Jahrhundert". In Migrationen. Globale Entwicklungen seit 1850 (= Globalgeschichte und Entwicklungspolitik 6), edited by Albert Kraler, Karl Husa, Veronika Bilger and Irene Stacher, 32-53. Wien: Mandelbaum Verlag, 2007.

Fleckner, Uwe, Maike Steinkamp, and Hendrik Ziegler. "In die Welt geschickt. Künstlerische Mobilität vom Mittelalter bis in die Gegenwart". In *Der Künstler in der Fremde. Migration – Reise – Exil*, edited by Uwe Fleckner, Maike Steinkamp and Hendrik Ziegler, 1–23. Boston: de Gruyter, 2015.

Fokkens, Antske, Niels Ockeloen and Serge ter Braake. BiographyNet. http://www.biographynet.nl/ (accessed 27 April 2018).

General Architecture for Text Engineering (GATE). https://gate.ac.uk/ (accessed 25 April 2018).

GeoNames. http://www.geonames.org/ (accessed 23 November 2018).

Georg Stephen. Luftlinie. https://www.luftlinie.org/ (accessed 12 March 2018).

Gruber, Christine, and Roland Feigl. "Von der Karteikarte zum biografischen Informationsmanagementsystem. Neue Wege am Institut Österreichisches Biographisches Lexikon und biographische Dokumentation". In *Biografische Lexika im Internet. Internationale Tagung der „Sächsischen Biografie" in Dresden (30. und 31. Mai 2008)* (= Bausteine aus dem Institut für Sächsische Geschichte und Volkskunde 14), edited by Martina Schattkowsky and Frank Metasch, 55–75. Dresden: Thelem Universitätsverlag, 2009.

Hillbrand-Grill, Friederike. *Ein neues Forum der Personengeschichte* (= Österreichisches Biographisches Lexikon, Schriftenreihe 2). Wien, 1992.

HistoGIS. https://histogis.acdh.oeaw.ac.at/ (accessed 13 November 2018).

IEPY. https://github.com/machinalis/iepy (accessed 25 April 2018).

Integrated Authority File (GND). http://www.dnb.de/EN/Standardisierung/GND/gnd_node.html (accessed 23 November 2018).

Joyeux-Prunel, Béatrice, Catherine Dossin, and Sorin Adam Matei. "Spatial (Digital) History: A Total Art History?—The Artl@s Project". In Visual Resources. An International Journal of Documentation 29, no. 1–2 (March 12, 2013), 47–58. DOI: 10.1080/01973762.2013.761119.

Kaiser, Maximilian. Der Wiener Diskurs zur Avantgarde. Rekonstruktion und Analyse des Diskursnetzwerkes, Dissertation, Universität Wien, 2017.

Kienle, Miriam. "Digital Art History 'Beyond the Digitized Slide Library': An Interview with Johanna Drucker and Miriam Posner". In Artl@s Bulletin 6, no. 3 (November 30, 2017). (accessed 23 November 2018).

Kuhn, Jonas and André Blessing. Textual Emigration Analysis. http://clarin01.ims.uni-stuttgart.de/tea/ (accessed 27 April 2018).

Lerner, Jürgen. "Beziehungsmatrix". In Handbuch Netzwerkforschung, edited by Christian Stegbauer and Roger Häußling, 355–78. Wiesbaden: VS Verlag, 2010.

Panter, Sarah, Johannes Paulmann, and Margit Szöllösi-Janze. "Mobility and Biography: Methodological Challenges and Perspectives". In *Mobility and Biography*. (= Jahrbuch für Europäische Geschichte / European History Yearbook 16), edited by Sarah Panter, 1–14. Berlin and Boston: de Gruyter, 2016.

Rehbein, Malte. "Digitalisierung braucht Historiker/innen, die sie beherrschen, nicht beherrscht", H/Soz/Kult, 27 November 2015.. www.hsozkult.de/debate/id/diskussionen-2905. (accessed 18 April 2018).

Royal Collection Trust. https://www.royalcollection.org.uk/collection/themes/trails/the-prince-of-waless-tour-of-india-in-1875-6 (accessed 2 February 2018).

Schich, Maximilian, Chaoming Song, Yong-Yeol Ahn, Alexander Mirsky, Mauro Martino, Albert-László Barabási, and Dirk Helbing. "A network framework of cultural history" *Science 345*, no. 6196 (2014): 558–562. http://barabasi.com/f/491.pdf (accessed 23 November 2018).

Schlögl, Matthias, Ágoston Bernád, Maximilian Kaiser, Katalin Lejtovicz, and Peter Rumpolt. "Biographik in den Digital Humanities – Kritische Bestandsaufnahme und quantitative Analysemöglichkeiten am Beispiel des Österreichischen Biographischen Lexikons 1815–1950". In DHd 2018. Kritik der digitalen Vernunft. Konferenzabstracts. Universität zu Köln, 26. Februar bis 2. März 2018, edited by Georg Vogeler, 360–363. http://dhd2018.uni-koeln.de/wp-content/uploads/boa-DHd2018-web-ISBN.pdf (accessed 23 November 2018).

Schlögl, Matthias, and Katalin Lejtovicz. "A Prosopographical Information System (APIS)". In *BD-2017. Biographical Data in a Digital World 2017. Proceedings of the Second Conference on Biographical Data in a Digital World 2017. Linz, Austria, November 6-7, 2017*, edited by Antske Fokkens, Serge ter Braake, Ronald Sluijter, Paul Arthur, and Eveline Wandl-Vogt, 53–58. 2018. http://ceur-ws.org/Vol-2119/paper9.pdf (accessed 23 November 2018).

Seligmann, Adalbert Franz, ed. Leopold Carl Müller. Ein Künstlerleben in Briefen, Bildern und Dokumenten. Wien: Rikola Verlag 1922.

Servaes, Franz. "Jehudo Epstein". In Jüdische Künstler, edited by Martin Buber, 160–170. Berlin: Jüdischer Verlag, 1903.

Stone, Lawrence. "Prosopography". *Daedalus. Historical Studies Today* 100, no. 1 (1971): 46–79.

ter Braake, Serge, Antske Fokkens, Ronald Sluijter, Thierry Declerck, and Eveline Wandl-Vogt, eds. *BD2015. Biographical Data in a Digital World 2015. Proceedings of the First Conference on Biographical Data in a Digital World 2015, Amsterdam, The Netherlands, April 9, 2015*. 2015. http://ceur-ws.org/Vol-1399/. (accessed 23 November 2018).

Tietze Hans. "Das Problem der Österreichischen Kunstgeschichte". In Österreich. Zeitschrift für Geschichte, H. 1, 1918/1919, 52–63.

Wien Geschichte Wiki.
https://www.wien.gv.at/wiki/index.php?title=Josef_Anton_Engelhart.
(accessed 30 April 2018).

The following articles by Anne Purschwitz and Kimmo Elo are based on papers read at the

12th Historical Network Research Workshop

„Historische Netzwerkforschung. Kommunikation in Netzwerken – Netzwerke der Kommunikation. Thematische, methodische und theoretische Perspektiven historischer Netzwerkanalyse zwischen Geschichts- und Kommunikationswissenschaft“

convened by Matthias Bixler and Erik Koenen and held from 20-21 April 2018 at the University of Bremen.

ANNE PURSCHWITZ

Netzwerke des Wissens - Thematische und personelle Relationen innerhalb der halleschen Zeitungen und Zeitschriften der Aufklärungsepoche (1688-1818)

Journal of Historical Network Research x (2018) 109-142.

Keywords

Topic-Modelling, Journals, Enlightenment, Discourse, communicative networks

Abstract

Given that the amount of available digital sources is constantly growing, studies on the acquisition and transmission of knowledge in the public sphere tend to be methodologically selective in their analysis of historical discourses. These studies, as a rule, tend to offer a very partial image from an intellectual history perspective of the production, representation, reception, and evaluation of knowledge. A more productive approach would be to analyse the ways in which pre-modern knowledge production functioned based on the broadest possible sample of sources. The project will seek, with a computer-based methodology, to consistently compile and systematically analyse all 356 journals and periodicals published in Halle between 1688 and 1815, as well as to gather information on all the actors participating in the process (authors, editors, printers et cetera). The periodical press of the Enlightenment is particularly relevant for this type of approach. Newspapers and periodicals allowed for a broader circulation of the ideas of the educated and for faster reactions to news, controversies, and publications, as they became a widely accepted form of interpersonal

communication. They thus filled a public and institutional gap and provided an open and adaptable medium for various public discourses without thematic limitations. We qualitatively go beyond standard discourse analysis by using a wealth of text and network data digitally available to reconstruct the breadth and depth of all the types of discourses of the Enlightenment, the combined quantitative methods for text and network analysis will allow us to trace the impact of the many involved parties right into the structure of the discourses.

1 Einleitung*

Anhand einer Beispielstudie möchte unser Projekt: 'Gesellschaftliche Wissensproduktion in der Aufklärung – Text- und netzwerkanalytische Diskursrekonstruktion. Die halleschen Zeitungen und Zeitschriften (1688-1815)' dazu beitragen, methodische Grundlagen zu entwickeln, um Diskursthemen zu identifizieren, Diskursverläufe miteinander zu vergleichen und, wenn möglich, für die Aufklärungsepoche Diskursmodelle zu entwickeln, die zum einen die Verortung einzelner Kontroversen in einem Gesamtspektrum ermöglichen und zum anderen eine Antwort auf die Frage bieten können, welche Themen im Betrachtungszeitraum von Relevanz waren. Mit einer solch umfassenden Herangehensweise und durch die Heranziehung einer Vielzahl von Dokumenten sollen den Fragen nach der gesellschaftlichen Reichweite und den thematischen Schwerpunkten der Aufklärung neue Impulse gegeben werden. Der Fokus des Projektes liegt in der Entwicklung einer Methode, die in der Lage sein wird mediale Quellenbestände in Korrelation mit strukturellen Rahmenbedingungen auf eine bestimmte Frage und deren Entwicklung hin zu untersuchen. In einem zweiten Schritt sollen unter Einbeziehung von Metadaten thematische, personelle und institutionelle Netzwerke rekonstruiert, visualisiert und für eine inhaltliche Interpretation aufbereitet werden. Durch die Etablierung von Modellen und Kennzahlen für Diskurse wird es dann möglich sein, unterschiedliche Diskurse miteinander in Beziehung zu setzen und zu vergleichen. Zu diesem Zweck gilt es, innerhalb der für den Zeitraum 1688–1815 nahezu vollständig erschlossenen periodischen Presselandschaft der Stadt Halle diskursive Ordnungsmuster zu identifizieren, aber auch die entsprechenden Rahmenbedingungen zu

---

* **Acknowledgments**: This article is based on a paper read at the *12th Historical Network Research Workshop* „Historische Netzwerkforschung. Kommunikation in Netzwerken – Netzwerke der Kommunikation. Thematische, methodische und theoretische Perspektiven historischer Netzwerkanalyse zwischen Geschichts- und Kommunikationswissenschaft" convened by Matthias Bixler and Erik Koenen and held from 20-21 April 2018 at the University of Bremen.

Corresponding Author: Dr Anne Purschwitz, Chair of Economic and Social History / IZEA, Martin Luther University Halle-Wittenberg, anne.purschwitz@izea.uni-halle.de

analysieren, die die lokale Diskurslandschaft prägten. Das bedeutet, dass neben Texten auch die Ebene der personellen Akteure, der Institutionen und Medien entsprechend zu berücksichtigen ist. Erst auf diese integrative Weise lässt sich Wissensgeschichte im Sinne Sarasins als Gesellschaftsgeschichte definieren.[1]

Im Folgenden sollen die Möglichkeiten einer Kombination von Topic-Modeling und Netzwerkanalyse an einem ausgewählten Beispiel überprüft und der Mehrwert einer solchen methodischen Koppelung gegenüber einer singulären Anwendung des jeweiligen Verfahrens dargestellt werden.

2 Zielsetzung

Ziel des Projektes ist es, für ein Zentrum der Aufklärung in Deutschland Entwicklungsfaktoren, Dynamik und Funktionsweise von Wissenssystemen zu ergründen. Unser Projekt möchte die Frage beantworten, wie, wann und (möglichst auch) warum während der Aufklärungsepoche ein bestimmtes Wissen in Form einer öffentlichen Thematisierung gesammelt, verbreitet und genutzt beziehungsweise ebenso unterdrückt, vernachlässigt oder ignoriert wurde. Im Vordergrund steht zunächst die Aufgabe, inhaltliche Beziehungen zwischen Texten herauszuarbeiten, ohne sie in ihrer Gänze manuell lesen und bearbeiten zu müssen, um mithilfe dieser Bezüge thematische Verknüpfungen und daran anschließend Diskurse zu rekonstruieren.

Während der Aufklärung übernahm insbesondere die Presse stetig umfassendere Funktionen bei der Wissensvermittlung, was zur Generierung eines Mediums führte, das entsprechend schnell auf immer häufiger zu registrierende Neuerungen, Kontroversen und Publikationen reagieren und sich zugleich zu einer weithin akzeptierten Form der interpersonellen Kommunikation entwickeln konnte.

Die damit einhergehende Expansion und Veränderung der öffentlichen Kommunikation und des öffentlichen Raumes führten zugleich zu einer stärkeren Fragmentierung von Inhalten, Akteuren und Medien. Und dennoch war gerade die periodische Presse als Gruppenphänomen in der Lage die Kommunikationspartizipation zu erweitern; sie verband individuelle Kommunikationsstränge und galt zugleich als lokaler Spiegel der gelehrten Welt.[2] Ihre spezifischen Charakteristika ermöglichten jedem hinreichend Lesefähigen, am Wissensdiskurs teilzunehmen – sie bildeten eine, wenn nicht die entscheidende Arena für öffentliche Diskurse, aus diesem Grund stehen sie im Zentrum unseres Projekts.

---

1 Sarasin „Wissensgeschichte“, S. 171.

2 Gierl, „Kompilation“, S. 72.

Um ein möglichst umfassendes und dichtes Bild der Wissenskultur im Raum Halle während der Aufklärung zu erhalten, bietet es sich an, einen diskursgeschichtlichen Ansatz zu wählen, da nur so das gesamte Spektrum von Wissen und den damit verknüpften Akteuren Berücksichtigung erfahren kann. Anknüpfend an Jäger wird unter Diskurs ein Fluss von Wissen durch die Zeit verstanden,[3] wobei in der kritischen Diskursanalyse auch Ansätze der Interdiskursanalyse [4] Berücksichtigung erfahren. Die Anwendbarkeit der Diskursanalyse auf ein sehr umfangreiches Quellenkorpus wie periodisch erscheinende Presseerzeugnisse es bilden, stellt zugleich die größte methodische Schwierigkeit dar. Doch nur durch eine solche Betrachtung kann ein Blick auf die Medialität und Prozessualität von Wissen und auf Journale als intermediäre Vermittler von Wissen eröffnet werden.

Für die konkrete Herangehensweise des Projekts bedeutet das zunächst zu untersuchen, welche Themen in welchen Journalen, wann, mit welchen inhaltlichen Präferenzen und in welcher Intensität diskutiert wurden. Die mithilfe eines Topic-Modelling generierten Themen können jedoch erst durch zusätzliche Verfahren mit ergänzenden Informationen in Beziehung gesetzt werden, so zum Beispiel den an den Themen beteiligten Akteuren, Journalen, strukturellen Rahmenbedingungen et cetera. In dieser Hinsicht eröffnet gerade die Rekonstruktion von Netzwerken in Korrelation mit Themen und weiterführenden (Meta)Daten nahezu unbegrenzte Möglichkeiten. Da sie beispielsweise durch die Verbindung mit Informationen zu Publikum, fachlicher Verortung (der Akteure und/oder Journale), Mitarbeit von Kupferstechern, Verlegern, Druckern bis zum Preis und Verbreitungsgebiet der Journale die modellierten Themen in ihrer Ausprägung näher bestimmen, ihre Genese interpretativ unterstützen und einen detaillierten Vergleich zwischen unterschiedlichen Themen ermöglichen kann. Erst durch eine solche Verknüpfung wird ersichtlich (und möglichst auch visualisiert) werden, welche Themen wann, in welchen Medien, unter Einbeziehung welcher Akteure und vor dem Hintergrund welcher Rahmenbedingungen, im Vordergrund standen. Daraus können dann unterschiedliche Aspekte zu Fragen der Ähnlichkeit von Diskursen (wie beispielsweise Thema, Zeitpunkt, Intensität, Frames und Diskurskonstellationen), [5] der Vernetzung von Diskursen[6] oder der Definition des öffentlichen Raumes als entscheidender Voraussetzung für Wissenszirkulation[7] abgeleitet werden.

---

3 Vgl. Jäger; Jäger, Deutungskämpfe, S. 13-37.
4 Vgl. Link, „Diskurs".
5 Kleinen-von Königslöw, Arenen-Integration, S. 51-58.
6 Benkler et al., „Social Mobilization".
7 Dahlgren, „The Internet" und Papacharissi, „A privat sphere".

3 Datengrundlage

Die Grundlage für dieses Vorhaben bilden die 360 zwischen 1688 und 1815 in Halle erschienenen Zeitungen und Zeitschriften und die damit in Beziehung stehenden persönlichen Relationen der an ihnen, auf unterschiedlichen Ebenen, beteiligten Akteure. Aktuell sind 290 Zeitschriften vollständig durch ihre Inhaltsverzeichnisse erfasst und die für die Erschließung entwickelte Datenbank weist mehr als 30.000 Beitragsverzeichnungen auf; sie umfasst rund 1.800 Personeneinträge, von denen etwa 900 eindeutig identifizierbar sind.[8] Dieses grundlegende Korpus wird beständig ergänzt und soll letzten Endes alle Beiträge der halleschen Zeitschriften und (wenn auch mit methodischen Schwierigkeiten verbunden) Zeitungen enthalten. Die Artikel sind auf unterschiedlichen Ebenen und unter Nutzung von Mehrfachzuweisungen mit Personen verknüpft. Die Kombinationen zwischen Person und Zeitschrift können durchaus vielfältig sein. Berücksichtigung erfahren momentan Verleger-, Herausgeber-, Autoren- und Rezensentenbeziehungen. Die dafür erforderlichen Informationen stammen aus den Zeitschriften selbst, wurden zudem aber durch prosopographische Recherchen ergänzt. Die Fülle von Verknüpfungen zwischen Zeitschriften und Personen offenbart bereits, dass gerade die periodische Presse der Aufklärungsepoche als ein engmaschiges Netzwerk verstanden werden muss, in dem nicht nur einzelne Journale, meist durch personelle Querverbindungen, miteinander in Bezug standen. Vielmehr zeitigen persönliche Kontakte und Verflechtungen für die Wissensvermittlung und -generierung der Aufklärungsepoche eine große Relevanz und bewirken durchaus Wechselwirkungen mit den – beispielsweise in der periodischen Presse auffindbaren – Kontroversen. Durch die Rekonstruktion von multimodalen Netzwerken können, neben den 'klassischen' persönlichen und inhaltlichen Verknüpfungen, zudem beispielsweise die Adressierung von Texten sowie die Kommunikationsrichtung Berücksichtigung erfahren, gleichzeitig aber auch die Journale selbst als Akteure des Wissens in den Blick genommen werden.

Eine grundlegende Schwierigkeit beinhalten Journalbeiträge, die nur mit Kürzeln, Initialen oder Chiffren versehen sind. In diesen Fällen wird zunächst versucht, die Autoren zu identifizieren. Dafür wurde am Institut für Informatik der Martin-Luther-Universität Halle-Wittenberg ein Tool entwickelt, das in der Lage ist, automatische Abfragen an eine eigens dafür erstellte 'Pseudonymdatenbank' zu richten. Die darin enthaltenen

---

8 'Eindeutig identifizierbar' bedeutet in diesem Zusammenhang, dass eine Person mit Vornamen und Nachnamen bezeichnet werden kann, nicht zwingend beinhaltet sind an dieser Stelle Lebensdaten und eine Rekonstruktion ihrer Existenz, zumal zahlreiche Personen kaum Spuren hinterlassen haben.

Informationen entstammen zum Teil unterschiedlichen Pseudonymlexika,[9] aber ebenso den häufig am Ende eines Jahrgangs in den Periodika selbst verzeichneten Autoreninformationen und werden fortlaufend ergänzt.[10] Das Verfahren selbst sucht durch einen Abgleich zwischen den unterschiedlichen Datenbanken nicht bloß nach vollständigen Übereinstimmungen, sondern berücksichtigt auch Ähnlichkeiten und Wahrscheinlichkeiten, so zum Beispiel bei variierenden Namensschreibweisen. Ebenso können die teilweise in den Datenbanken verfügbaren Lebensdaten für eine genauere Zuordnung mit einbezogen werden. Nur im Idealfall, das heißt bei einer genauen Übereinstimmung von Nachnamen, mindestens zwei Vornamen und einer Kongruenz der Lebensdaten mit dem Zeitpunkt der Publikation wird ein Autor eindeutig identifiziert (ca. 50 Personen). In der Mehrzahl der Fälle kann hingegen ein Personenkreis bestimmt werden, der dann durch die (nicht-automatisierte) Bestimmung thematischer Schwerpunkte beziehungsweise eine Fachzuordnung der Autoren weiter eingegrenzt werden muss (bisher bei ca. 200 Personen).

Ein weiteres Problem der Periodika der Aufklärungsepoche sind die zahlreichen anonym erschienenen Beiträge. In diesen Fällen kann die Frage nach persönlichen Bezügen nicht beantwortet werden. Die Suche nach Beziehungen zwischen den Texten erhält hier einen deutlich höheren Stellenwert, denn auch die veröffentlichten Artikel, Rezensionen, Anmerkungen und Abhandlungen bilden ein Netzwerk; dies nicht länger nur auf Grundlage persönlicher Relationen zwischen Verlegern, Herausgebern oder Autoren, sondern vielmehr nehmen die Texte aufeinander Bezug oder thematisieren ähnliche Gegenstände in anderen Zusammenhängen und Medien. Diese Bezüge können mithilfe der Topic-Modellierung aufgefunden werden.

4 Topic-Modeling

Unter dem Schlagwort distant reading werden seit einigen Jahren neue Möglichkeiten diskutiert, um in der Menge der (zunehmend auch digital verfügbaren) Texte durch die Anwendung algorithmenbasierter Verfahren intertextuelle Zusammenhänge zu erkennen. Hierzu gehören beispielsweise Text Mining, Clustering oder Topic-Modeling.[11]

---

9 Bisher bearbeitet wurden zunächst: Weller, Lexicon pseudonymorum; Eymer, Pseudonymen-Lexikon; Holzmann, Pseudonymen-Lexikon und in Auszügen Parthey, Mitarbeiter.

10 Unveröffentlichte Bachelorarbeit: Mladenova, Automatisierung.

11 Vgl. dazu: Crane, „What Do You Do".

In Hinblick auf die automatische Sprachverarbeitung müssen dabei zwei Ebenen voneinander unterschieden werden, zum einen die sprachlich-stilistische und zum anderen die inhaltlich-thematische Analyse. Im Folgenden liegt der Fokus auf den Möglichkeiten einer inhaltlich-thematischen Strukturierung eines Textkorpus. Das dafür nutzbare Topic-Modeling bezeichnet eine quantitative Methode der Textanalyse, die auf ein Korpus von Dokumenten angewendet werden kann und dabei die Wortverteilung innerhalb dieser Textsammlung statistisch ermittelt. Ziel ist das Aufdecken 'verborgener' semantischer Strukturen, ohne dafür explizit über semantisches Wissen verfügen zu müssen.[12] Topic-Modeling folgt dabei der Hypothese, dass Wörter, die in demselben Kontext vorkommen, dazu tendieren, eine ähnliche Bedeutung zu besitzen. Dadurch könnten grundlegend wiederkehrende Themen, Motive und unter Hinzuziehung weiterer Daten auch Diskurse identifiziert werden. Die Topics (Themen) helfen somit Komplexität zu strukturieren.[13]

Zunächst erfolgt die Modellierung eines Standard-LDA-Topic-Models.[14] Dieses bietet im Unterschied zu Clustering-Verfahren[15] den Vorteil, dass die Wörter in einem Dokument einzeln den Topics zugeordnet werden können. Topic-Modelle betrachten ein Dokument somit als eine Mischung von Topics. Eine Textanalyse mithilfe von Topic-Modellen kann als Knowledge-Discovery-Prozess beschrieben werden und folgt den Schritten 1. Datenauswahl, 2. Datenvorverarbeitung und Datentransformation, 3. Datenanalyse und 4. Evaluation des Modells.[16] Die Topics selbst basieren auf der Verteilung von Wahrscheinlichkeiten von Wörtern, die jeweiligen Dokumente hingegen auf der Verteilung von Wahrscheinlichkeiten von Topics. Vorgegeben sind somit die in den Dokumenten enthaltenen Wörter, während die 'versteckten' Topics in einem iterativen Prozess aufgedeckt werden. Dieses rein mathematische Verfahren funktioniert sprachunabhängig und kann durch die Herstellung von inhaltlichen Zusammenhängen zwischen Wörtern und Dokumenten den Bedeutungsgehalt eines Textes erschließen. Um Topic-Modeling möglichst zielführend einsetzen zu können, wird zunächst ein (möglichst geschlossenes beziehungsweise eingrenzbares) Korpus benötigt. Im Unterschied zu zahlreichen anderen informatischen Analyseverfahren bedarf das Topic-Modeling jedoch keiner Metadaten, die die Topics inhaltlich vorab definieren.

---

12 Vgl. dazu: Lim et al., „Nonparametic"; Mimno und Blei, „Bayesian Checking"; Blei, „Probabilistic Topic Models"; Blei und Lafferty, „Dynamic Topic Models".

13 Zum Begriff ‚topic' vgl.: Jacobi et al. „Quantitative analysis".

14 Einen guten Überblick dazu bieten: Anthes, „Topic models" und Sievert und Shirley, „LDAvis".

15 Vgl. dazu bspw.: Stier et al. „Election Campaigning".

16 Vgl dazu ausführlich: Papilloud und Hinneburg, Qualitative Textanalyse, S. 28-36.

Das Haupthindernis in Hinblick auf historische, in Fraktur gedruckte Texte besteht in der Tatsache, dass diese nicht maschinenlesbar beziehungsweise erst durch mühsame Verfahren aufbereitet werden müssen, aus diesem Grund können Natural-Language-Prozesse nicht problemlos ansetzen. Digitalisierungsverfahren, die historische Texte in PDF- bzw. Bilddateien umwandeln, bieten hier zunächst keinen Mehrwert, da sie mit automatisierten Verfahren nicht lesbar sind.[17]

In unserem Projekt nutzen wir den, von Alexander Hinneburg an der Martin-Luther Universität Halle-Wittenberg entwickelten TopicExplorer[18]. TopicExplorer nutzt zurzeit die LDA-Implementierung von Mallet[19]. Die genutzte Parametrisierung folgt keinem festen Hyperparameter für LDA, sondern optimiert diese in regelmäßigen Abständen während des Trainings des Topic-Modells. Die Nutzung des TopicExplorers wird aktuell in einem permanenten Wechselspiel, an die Spezifika historischer Dokumente, und die damit verbundenen Schwierigkeiten, angepasst.

Da bisher nur für einen Teil der betreffenden Journale OCR-Volltexte vorliegen, erfolgten die ersten Topic-Modellierungen im Projekt ausschließlich anhand der Artikelüberschriften. Die zu untersuchende Haupteinheit der Zeitungen und Zeitschriften bilden somit die einzeln erfassten, vollständigen bibliographischen Angaben der in ihnen beinhalteten Artikel. In einem zweiten Schritt soll die Funktionalität dieses Vorgehens durch exemplarische Versuche mit Volltexten validiert werden. In den meisten Fällen sind die Titelinformationen jedoch durchaus umfangreich und beinhalten beispielsweise bei Rezensionen auch den Titel und Verfasser des rezensierten Werkes, in anderen Fällen, wie den biographischen Nachrichten, werden die Artikelüberschriften hingegen erst verständlich, wenn sie mit dem jeweiligen Zeitschriftentitel in Bezug gesetzt werden. Um diese methodische Schwierigkeit zunächst zu umgehen, wurde mit einem Datensatz gearbeitet, der aktuell sehr kurze Titel ausschließt. Die daraus resultierende ähnliche Dokumentenlänge kommt der Funktionalität der Topic-Modellierung entgegen. Schwieriger könnte sich die Situation gestalten, wenn die Volltexte

---

17 Die bisher zur Verfügung stehende OCR-Software erfordert aktuell noch zahlreiche Vorverarbeitungen, Korrekturen oder Trainings, sodass sie auf ein so heterogenes Korpus, wie es der Bestand unterschiedlicher Zeitschriften darstellt, kaum angewendet werden kann (vgl. Breuel, „The OCR"; Springmann, „OCR" und Stollwerk, „Machbarkeitsstudien").

18 Vgl. Hinneburg et al, „TopicExplorer"; Hinneburg et al., „Exploring Document Collections" und Röder et al., „Exploring the Space". Ausführlicher zum TopicExplorer und der dahinter stehenden Programmierung siehe: https://github.com/hinneburg/TopicExplorer.

19 Vgl. https://github.com/mimno/Mallet.

der Zeitungen und Zeitschriften in eine Topic-Modellierung einbezogen werden, da der Umfang der Beiträge in den halleschen Journalen große Schwankungen aufweist.[20]

Die Topic-Modellierung nutzt die verfügbaren bibliographischen Angaben (im Folgenden als Dokument bezeichnet) und ordnet möglichst jedes Wort (Token) eines Dokuments Themen zu, wobei die Ausgangsanzahl der Themen variiert werden kann und die Themen selbst Ähnlichkeiten zwischen den einzelnen Dokumenten erklären sollen. Im Resultat erhält man zunächst unabhängige Annahmen über die Daten, die aufgrund von Plausibilitätskriterien in Themengruppen gegliedert werden, wobei auch immer Mehrfachzuweisungen möglich und gewünscht sind.

Relativ aufwendig gestaltet sich die Vorverarbeitung der Dokumente. Satzzeichen und Stoppwörter, Einzelbuchstaben und Zahlen müssen entfernt werden. Mithilfe des TreeTagger[21] werden die Wörter der Dokumente in Wortarten klassifiziert (Part-Of-Speech tagging). Für historische Texte stehen zudem Tools zur Verfügung, die eine Vereinheitlichung und Anpassung der variierenden Schreibweisen an die aktuelle Rechtschreibung ermöglichen.[22] Zwar stellen abweichende Schreibweisen für das Topic-Modeling grundlegend keine Schwierigkeit dar, da nur die Zeichenfolge für die Analyse Relevanz hat; Probleme können jedoch bei der Lemmatisierung entstehen, sodass eine Normalisierung durchaus sinnvoll ist. Anschließend können die Wörter nach Wortarten für die Topic-Analyse sortiert werden. Substantive tragen dabei die meisten inhaltlichen Bedeutungsdetails eines Textes. Für die erste Überblicksanalyse erfolgte aus diesem Grund eine Konzentration auf Substantive, weitere Wortarten sollen in späteren Verfahren hinzugenommen werden. Da sehr häufige und sehr seltene Wörter die Genauigkeit von Topic-Modellen mindern, wurden niedrig-frequente Wörter entfernt und hochfrequente Begriffe einer detaillierten manuellen Selektion unterworfen. Durch die anschließende Lemmatisierung wurde das Vokabular für die Topic-Analyse so genau wie möglich bestimmt. Um die spätere Interpretation der generierten Topics zu erleichtern, erfolgte die Aufnahme von Metadaten zu jedem Dokument. Dazu zählen, Autor, Veröffentlichungszeitpunkt und Journaltitel. Diese Metadaten werden nicht in die Topic-Generierung einbezogen, stehen aber für die Interpretation der Topics in der späteren Visualisierung zur Verfügung, beziehungsweise können sie nach der

---

20 Vorstellbar wäre in diesem Zusammenhang, die innere Struktur langer Beiträge (Zwischenüberschriften) zu übernehmen oder jeweils auf der Ebene der Seiten zu arbeiten.

21 Vgl. Schmid, „Prohabilistic".

22 Beispielsweise das Tool CAB (Cascaded Analysis Broker) Historical Text, siehe: Jurish, „Canonicalizing".

Modellierung durch SQL-Abfragen an die Datenbank variabel abgefragt werden. Die Festlegung auf zunächst 30 Themen erfolgte nach mehreren Test-Modellierungen. Grundlegend bietet der genutzte TopicExplorer aber die komfortable Möglichkeit mehrere Themen zusammenzufassen und wieder zu trennen, so dass in Hinblick auf die Themenanzahl ein vom Bearbeiter später zu nutzender Spielraum besteht.

Als ein Problem erwies sich bei Betrachtung der ersten Ergebnisse die Sprache. Eine nicht geringe Anzahl von Dokumenten ist in lateinischer oder französischer Sprache verfasst. Diese wurden zwar gleichberechtigt in die Analyse einbezogen und entsprechend Themen zugeordnet, doch besteht noch keine Möglichkeit einer automatisierten Übersetzung. Daher wurden beispielsweise zwischen synonymen lateinischen Begriffen und deutschsprachigen Bezeichnungen keine Beziehungen hergestellt. Dies führte dazu, dass sich insbesondere die französischsprachigen Beiträge nahezu ausschließlich in eigenständigen Themen wiederfanden, ohne einen Bezug zu ähnlichen Themen deutschsprachiger Dokumente aufzuweisen. Diese Beobachtung lässt es als zwingend notwendig erscheinen, eine sprachliche Angleichung (Übersetzung) vorzulagern. In unserem Beispiel betrifft dies cirka ein Fünftel der Dokumente.

Eine für unseren Kontext sehr spezifische Schwierigkeit bilden die sowohl innerhalb der Dokumente als auch im Zusammenhang mit den bibliographischen Informationen zu den Zeitungen und Zeitschriften enthaltenen Personennamen und Ortsbezeichnungen. Insbesondere bei Rezensionen und auch in naturwissenschaftlichen Journalen stellen beispielsweise die namentlichen Bezüge zu Autoren von wissenschaftlichen Abhandlungen einen entscheidenden Orientierungspunkt dar, da anhand dieser Informationen bewertet werden kann, auf welche Gelehrte sich andere Publikationen bezogen. Bei der Bereinigung der Wortlisten wurde bewusst auf Initialen verzichtet, da dies in einem frühen Versuch zu Themenlisten geführt hatte, die zum Beispiel nur Einzelbuchstaben (S., P., W. et cetera) enthielten. Indessen erscheint es nicht zielführend, nur mit Nachnamen in die Analyse zu gehen, da in diesem Fall keine Eindeutigkeit erreicht werden kann. Andererseits werden zahlreiche Namen nicht als Namen erkannt, bzw. es erfolgt nicht in allen Fällen eine entsprechende Zuweisung zu Themen. Dies führte zu einem zusätzlichen vorbereitenden Element – der Normierung von Namensschreibweisen. Wobei es notwendig war, die Vornamen als feste Einheiten an die Nachnamen zu binden und auch die Form dieser Koppelung zu normieren.

Etwas schwieriger gestaltet sich die ähnlich gelagerte Fragestellung in Hinblick auf Ortsnamen. Diese finden sich ebenfalls häufig in Dokumenten, weisen aber in vielen Fällen keine analoge Relevanz für die Themengenerierung auf, wie dies bei den Personennamen angenommen

werden muss. Häufig beziehen sich Ortsbezeichnungen, insbesondere in Rezensionszeitschriften, auf den Verlagsort – diese Information ist zwar nicht ohne Interesse, bildet aber nur in Ausnahmefällen eine entscheidende Komponente für die Themenmodellierung. Die dadurch provozierte Häufung führte in ersten Tests zu einer Verzerrung der Themenlisten. So fanden sich in einem Thema auffallend viele Ortsnamen (Berlin, Paris, Petersburg, England, Brüssel, Venedig et cetera), deren Beziehung durch die anderen dem Thema zugeordneten Begriffe nicht ersichtlich wurde. In dieser Hinsicht müssen noch weitere Versuche erfolgen, um zu testen, wie sinnvoll es sein kann, auf Ortsnamen vollständig zu verzichten, bzw. um zu entscheiden, in welchen Fällen diese für die Topic-Modellierung unerlässlich sind. Dies ist beispielsweise bei Reisebeschreibungen, in Verbindung mit der Frage, welche Weltteile als besonders interessant und erforschenswert galten, wie auch in Bezug auf Staatenbeschreibungen offensichtlich.

Die bisher skizzierten Schwierigkeiten, Probleme und weiter zu entwickelnden Anforderungen sollen jedoch keineswegs verdecken, dass bereits beeindruckende Ergebnisse erzielt werden konnten.

4.1 Der TopicExplorer in der praktischen Anwendung[23]

Abbildung 1 gibt einen Eindruck von der Nutzeroberfläche des TopicExplorers;[24] in der unteren Leiste ist ein Teil der (in ihrer Anzahl vorgegebenen) 30 Themen zu erkennen.

Die Topics werden als Wortlisten dargestellt, deren Wörter nach absteigender Wahrscheinlichkeit sortiert sind, sodass die häufigsten Wörter in einem Topic oben stehen. Um das gesamte Topic-Spektrum von links nach rechts ohne inhaltliche Brüche durchsehen zu können, werden die Topics linear nach wortbasierter Ähnlichkeit geordnet. Die berechnete Themenreihenfolge wird als Regenbogenskala dargestellt, wobei die Farbe als visuelle Hash-Funktion dient, da eine Ähnlichkeit der Farbe auf eine Ähnlichkeit zwischen den Themen hinweist. In der oberen Hälfte hingegen sind die, einem Thema zugeordneten, Dokumente aufgelistet.

---

23 Ausführlich zur Handhabung des TopicExplorers: https://blogs.urz.uni-halle.de/hallejournals/kategorie/methoden/topic-modelling/.

24 Der Prototyp des TopicExplorers zu den halleschen Journalen kann unter dem Link: http://topicexplorer.informatik.uni-halle.de:8080/HALJOUR_0_te/# eingesehen werden.

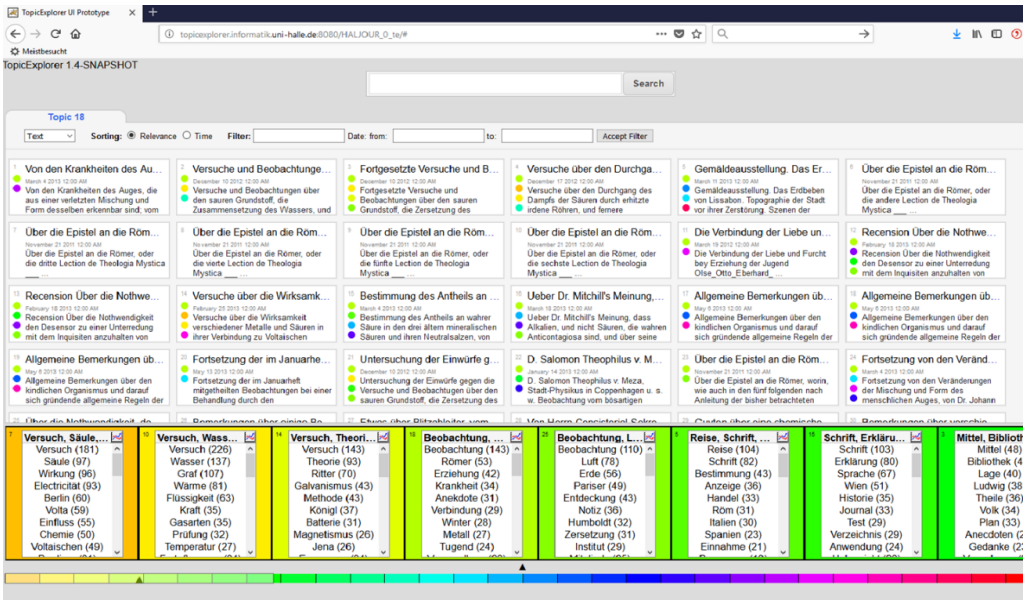


Abbildung 1: Screenshot TopicExplorer zu den halleschen Journalen.

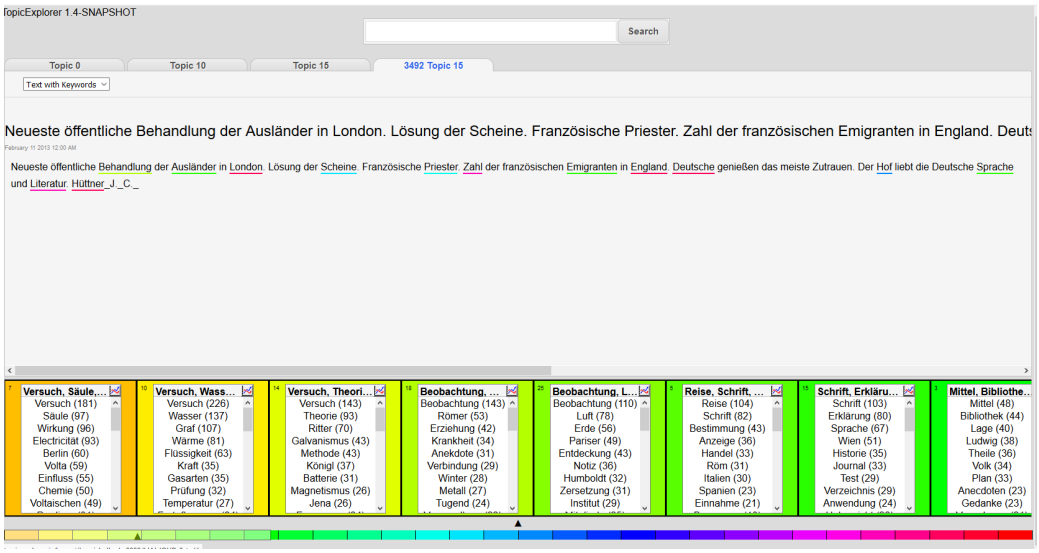


Abbildung 2: Screenshot TopicExplorer zu den halleschen Journalen.

Durch die farbigen Punkte neben den Dokumenten wird die jeweilige Themenzuweisung (je Dokument die vier häufigsten Themen, basierend auf der Zuordnung von Token eines Dokuments zu Themen) verdeutlicht. Die Oberfläche des TopicExplorer ermöglicht es nahe beieinanderliegende Themen zusammenzuführen, sie bietet eine Suchfunktion mit Auto-Vervollständigen, unterschiedliche zeitliche Filterfunktionen und die Option, für jedes einzelne Dokument die Zuordnung von Wörtern zu Themen sichtbar zu machen (Abbildung 2). Im TopicExplorer selbst kann die zeitliche Entwicklung jedes Topic visualisiert werden.

4.2 Ergebnisse des Topic-Modeling

Betont werden muss an dieser Stelle ausdrücklich, dass die generierten Themen zunächst keinerlei Interpretation beinhalten, diese muss erst durch den Bearbeiter, beziehungsweise durch das Hinzuziehen ergänzender Informationen geleistet werden.[25] Einige der generierten Themen weisen bereits auf den ersten Blick eine relativ hohe Verständlichkeit auf. Doch nicht alle abgebildeten Themen offenbaren eine so hohe Klarheit. Um sie erschließen, verorten und interpretieren zu können, werden zwingend zusätzliche Informationen benötigt. Die dafür erforderlichen Metadaten können gleichzeitig dazu dienen, die Struktur und Stabilität der generierten Topics zu überprüfen. Für eine adäquate Darstellung und Einbeziehung dieser Metadaten, erscheint die Netzwerkanalyse als das vielversprechendste Instrument.

Der TopicExplorer speichert alle Daten der Texte und des Topic-Modells in einer relationalen Datenbank. Für weiterführende Analysen und Datenexporte kann auf diese Datenbank, unabhängig vom TopicExplorer, mithilfe von SQL-Abfragen zugegriffen werde.[26] So kann beispielsweise für jedes Thema hinterfragt werden, welche Autoren und/oder Journale zu welchem Zeitpunkt involviert waren. Diese Abfragen stellen gewichtete Beziehungen zwischen Thema und Journal, Thema und Autor und Thema, Autor und Journal her. Dadurch kann gezeigt werden, wie sich Themen in Relation zu/mit Zeitschriften oder Akteuren ausbilden. Des Gleichen besteht die Möglichkeit, die zeitliche Entwicklung aller – oder einzelner Themen – in Relation zu den Journalen zu setzen; damit kann beispielsweise die Frage beantwortet werden, ob bestimmte Zeitschriften Themenkonjunkturen bedingen. Dies erscheint

---

25 Vgl. zur Problematik der Interpretation von Topics: Chang et al., „Reading Tea Leaves“; Sievert und Shirley, „LDAvis“; Kataria et al., „Context senxitive“; Riddell, „Text Analysis“; Tangherlini und Leonard, „Traweling in the Sea“.

26 Eine Auswahl gängiger SQL-Statements findet sich unter: https://github.com/hinneburg/TopicExplorer/blob/master/sql-analytics.md.

aktuell insbesondere in Hinblick auf die naturwissenschaftlichen Zeitschriften als sehr naheliegend.

5. Topic-Modeling und Netzwerkanalyse (experimentelles Beispiel)

Da sich unser Projekt noch in einem Frühstadium befindet, sollen im Folgenden insbesondere die Visionen einer Kombination zwischen Topic-Modeling und Netzwerkanalyse an einem exemplarischen Beispiel in ersten Ansätzen aufgezeigt werden. Grundlage dafür bildet das Thema 4. Alle folgenden Darstellungen gehorchen dabei keinem vorangestellten Narrativ, dass durch die voraussetzungsfreie Anwendung einer Topic-Modellierung gerade verhindert werden soll, sondern stellen – wie auch die in Beziehung gesetzten Vergleiche – ausschließlich aus dem Projekt und den genutzten Quellen hervorgegangene Interpretationen dar.

Dem Thema sind knapp 1.000 Dokumente zugeordnet, die keinesfalls alle gleichermaßen in eine Netzwerkanalyse einbezogen werden sollen und können. Vor diesem Hintergrund erfolgt über eine entsprechende SQL-Abfrage, die die Metadaten der Topic-Modellierung berücksichtigt, eine Hierarchisierung der involvierten Journale und Autoren. Dafür wird die Summe aller Dokumente mit einer Tokenzuordnung zu Thema 4 in Relation zu Journalen und Autoren abgefragt. Für die Gesamtmenge dieser Dokumente wird dann die durchschnittliche Anzahl der dem Thema zugeordneten Token berechnet. Diese Zahl dient im Folgenden zugleich für die Gewichtung der Relationen innerhalb der Netzwerkanalyse. In die spätere Untersuchung werden nur durchschnittliche Werte >1 einbezogen. Durch entsprechende SQL-Abfragen können unterschiedliche Listen generiert werden. Zum einen eine nach absteigender Relevanz sortierte Liste zu den an Thema 4 beteiligten Journalen (Tabelle 1) und eine gleichermaßen strukturierte Aufzählung der involvierten Autoren (Tabelle 2). Damit wird zum Ausdruck gebracht, welche Autoren/Journale in welcher Intensität mit dem Thema verknüpft sind.

Durchschnittliche Anzahl Token der summierten Dokumente	Journal_Titel (Kurztitel)
3,67	Magdeburg-halberstädtische Blätter
2,00	Magazin für die Religion
1,70	London und Paris
1,50	Beyträge zur Beförderung der Erkenntnislehre
1,50	Merkwürdige Krankengeschichten
1,33	Allgemeine Geschichte der europäischen Staaten
1,33	Auserlesene und noch nie gedruckte Gedichte
1,33	Compendiöse Bibliothek der schönen Wissenschaften
1,20	Neue Reisebemerkungen
1,19	Magazin für die neue Historie und Geographie
1,16	(Fortgesetzte) Betrachtungen über die Werke Gottes in der Natur
1,13	Journal der Physik
1,13	Der Andächtige
1,13	Magazin der deutschen Critik
1,09	Der Naturforscher
1,09	Deutsche Bibliothek der schönen Wissenschaften
1,06	Allgemeine historische Bibliothek
1,06	Annalen der Physik

Tabelle 1: Thema 4. Durchschnittliche Anzahl Token der summierten Dokumente je Journal über den Gesamtzeitraum (Ausschnitt).

Durchschnittliche Anzahl Token der summierten Dokumente	Name_Autor
2,00	Campbell, H.
2,00	Dohm, Christian Wilhelm
2,00	Gilbert, Ludwig Wilhelm
2,00	Hensler, Philipp
2,00	Monroe, Donald
2,00	O.
2,00	Porteau
2,00	Rambach, Friedrich Eberhard
2,00	Schloiser, Johann Heinrich
1,84	Hüttner, J. C.
1,78	Aubigny gen. Engelbrunner, N.
1,33	Berthollet
1,22	Hostig, K. G.

Tabelle 2: Thema 4. Durchschnittliche Anzahl Token der summierten Dokumente je Autor über den Gesamtzeitraum (Ausschnitt).

Bereits eine solche Darstellung kann, mit Vorwissen über die fachliche Zuordnung von Autoren und Journalen, Interpretationsansätze bieten, so handelt es sich offensichtlich um ein Thema mit dem geographischen Schwerpunkt England, mit dem eine Vielzahl thematisch unterschiedlicher Journale in Beziehung standen. Die Inhalte reichen von Geschichte (*Allgemeine Geschichte der Europäischen Staaten*, *Allgemeine historische Bibliothek*), Unterhaltung (*London und Paris*), Literatur (*Auserlesene und noch nie gedruckte Gedichte*, *Magazin der deutschen Critik*, *Deutsche Bibliothek der schönen Wissenschaften*) über Naturwissenschaften (*Journal der Physik*, *Der Naturforscher*), bis hin zu Theologie (*Magazin für die Religion*, *Betrachtungen über die Werke Gottes in der Natur*). Der Schwerpunkt liegt jedoch auf einer allgemeinen Wissensvermittlung über die aktuellen politischen, wissenschaftlichen und sozialen Verhältnisse in England (*Magdeburg-halberstädtische Blätter*, *Neue Reisebemerkungen*, *Magazin für die neue Historie und Geographie*). Mit einer solchen Überblicksdarstellung kann jedoch, auch unter Hinzuziehung der Metadaten zu Autoren und Journalen, noch nicht beantwortet werden, ob alle Bereiche im zeitlichen Verlauf in gleicher Intensität abgedeckt wurden, beziehungsweise welchen inhaltlichen Verschiebungen das Thema unterlag.

Ausgehend von den Möglichkeiten des TopicExplorers kann die Entwicklung des Themas über den Gesamtzeitraum betrachtet werden (Abbildung 3). Diese Visualisierung bietet zudem die Option, für jeden beliebigen Zeitpunkt die vier jeweils dominanten Token anzeigen zu lassen.

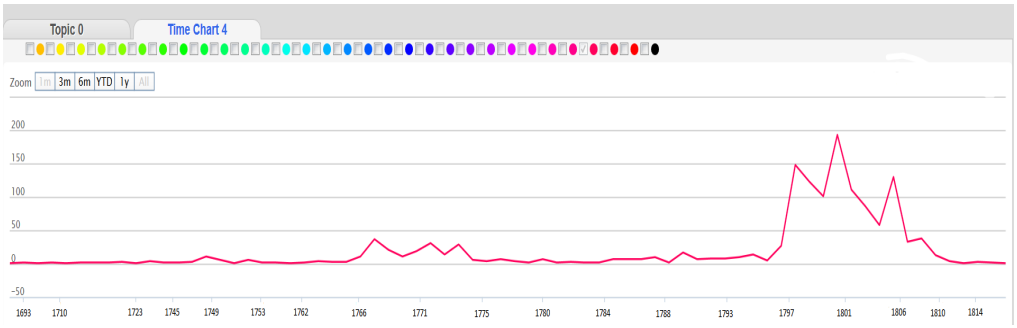


Abbildung 3: Zeitliche Entwicklung von Thema 4. Darstellung im TopicExplorer.

An dieser Stelle kann somit die grundlegende Entwicklung eines Themas, einhergehend mit begrifflichen Verschiebungen und Schwerpunktlegungen reflektiert werden. Erklärungen dafür, warum es beispielsweise zu einer Zunahme der Beiträge kam, beziehungsweise welche Faktoren die erkennbaren Schwankungen verursachen, bietet der TopicExplorer nicht.

Um diese Fragen zu beantworten, werden die Informationen der generierten Listen durch eine Netzwerkanalyse miteinander in Beziehung

gesetzt. An dieser Stelle in die Netzwerkanalyse integriert werden Metadaten zu den beteiligten Journalen (Fachzuordnung, Erscheinungsweise, Erscheinungszeitraum, Herausgeber, Drucker, Verleger et cetera) und Autoren (persönliche Beziehungen, Beruf, Aufenthaltsort, Publikationen, Fachgebiet et cetera). Für die sich anschließende Darstellung liegt der Schwerpunkt auf der fachlichen Zuordnung der Journale und Autoren, um damit inhaltliche und thematische Verschiebungen innerhalb der Entwicklung des Themas aufzeigen zu können. Infolge dieser Ergänzung von Daten entsteht ein multimodales Netzwerk, dessen, für die Wissensrekonstruktuion elementare Akteurstypen: Thema, Autor und Journal darstellen. Die entsprechenden gerichteten Relationen bringen zum Ausdruck, welche Journale bzw. Akteure sich an dem entsprechenden Thema beteiligen und welcher Fachrichtung sie zugeordnet werden können – Mehrfachzuweisungen sind dabei möglich. Die Gewichtung der Relation zwischen Journal/Autor und Thema erfolgt anhand der SQL-Abfragen in der Datenbank des TopicExplorers, die in diesem Fall mit zeitlichen Grenzen versehen sind.

Mithilfe dieser Daten kann quasi das Ego-Netzwerk von Thema 4 im Zeitverlauf rekonstruiert werden, wobei die zeitliche Dimension zum einen an die Erscheinungsdauer der Journale und zum anderen an das Publikationsdatum der zugeordneten Beiträge gebunden wird. Thema 4 stellt das ‚Ego' dar, während Journale und Autoren die ‚Alteri' bilden. Mit dieser Konstellation kann das Netzwerk um Thema 4 zunächst isoliert von anderen Themen in seiner Entwicklung und Ausprägung untersucht werden. In Hinblick auf das behandelte Thema ergibt sich ein Zeitraum zwischen 1689 und 1817, der für eine vergleichende Analyse in vier mathematisch etwa gleich große Abschnitte eingeteilt wird. Anhand einer solchen Strukturierung, die zunächst nicht an herausragenden Ereignissen, sondern vielmehr an einer gleichmäßigen Verteilung über den Gesamtzeitraum orientiert ist, können thematische, mediale und personelle Schwerpunktverschiebungen erforscht werden. Eine generelle Problematik liegt, insbesondere für die ersten Jahrzehnte des Betrachtungszeitraums, in der häufig fehlenden Zuordnung von Beiträgen zu Autoren, da eine namentliche Kennzeichnung von Artikeln erst im Lauf des 18. Jahrhunderts eine zunehmende Verbreitung erfuhr.

Die Netzwerkvisualisierung selbst erfolgt mit Gephi[27] unter Verwendung des Fruchterman Reingold Algorithmus[28]. Für den ersten Abschnitt zwischen 1689 und 1725 finden sich sieben Zeitschriften, die sich in etwa gleicher Intensität an Thema 4 beteiligten (Abbildung 4). Gruppiert man die involvierten Journale nach ihrer inhaltlichen Ausrichtung, so findet sich eine –

---

27 https://gephi.org/ und https://github.com/gephi/gephi. Version 0.9.2.

28 Vgl. Fruchterman und Reingold, „Graph drawing".

dem grundlegenden Schwerpunkt der halleschen Journale dieser Zeit entsprechende – Konzentration auf Rezensionszeitschriften mit theologischer und staatswissenschaftlicher Ausrichtung.

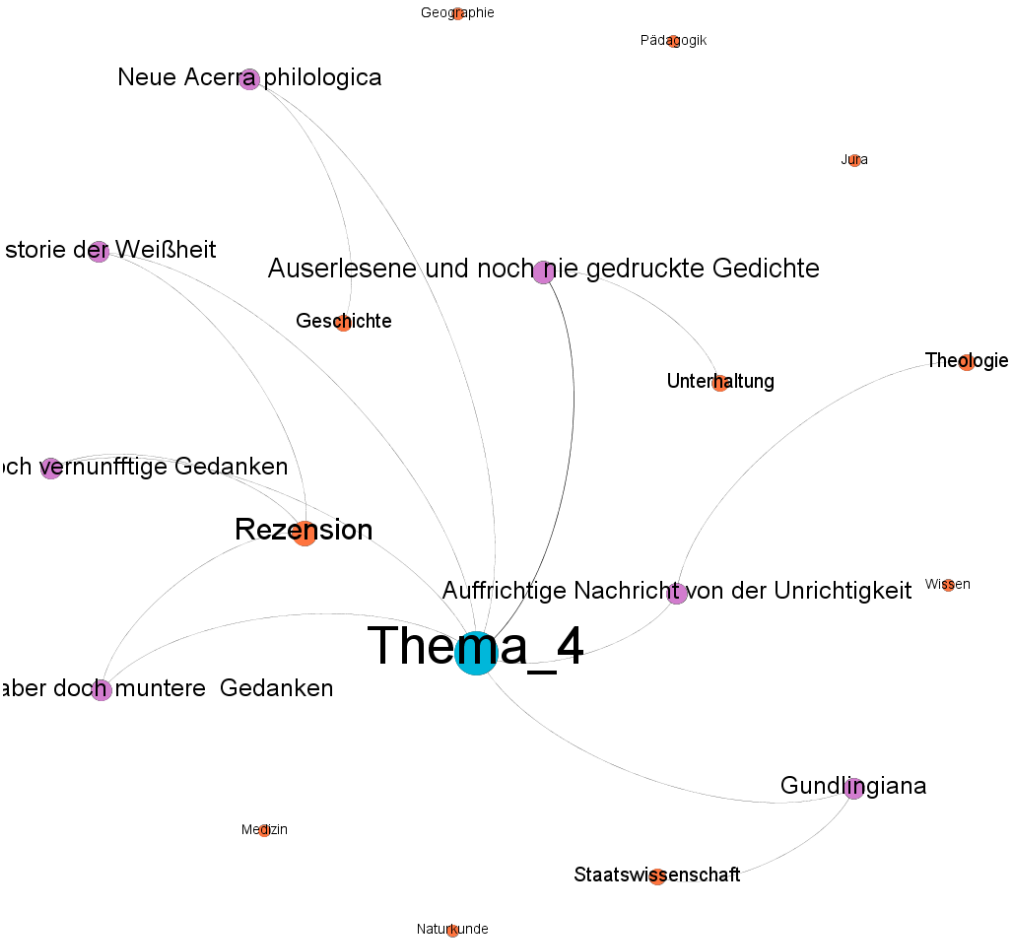


Abbildung 4: Thema 4 Netzwerk zwischen 1689 und 1725. Violet = Journale, rot = Fachzuordnung, Knotengröße und Schriftgröße = Gewichteter Grad.

Keiner der entsprechenden Beiträge ist dabei mit einem Autorennamen gekennzeichnet. Gleichzeitig verläuft das Thema sehr flach, das bedeutet die Diskursintensität des Themas ist nur gering ausgeprägt (Abbildung 3), eine hervorstechende Auseinandersetzung mit England ist somit noch nicht existent, sondern vielmehr Nebenprodukt derjenigen Journale Halles, die es sich zur Aufgabe gemacht haben, eine große Bandbreite von Themen, Gegenständen oder Publikationen zu betrachten. Aus diesem Grund findet sich das Thema, meist relativ isoliert in den zeitgenössisch herausragenden halleschen Journalen. So beispielsweise in den von Christian Thomasius

herausgegebenen Zeitschriften *Freymüthige jedoch vernunfft- und gesetzmäßige Gedanken über allerhand fürnehmliche aber neue Bücher (1689-1690)* und den *Ernsthaffte aber doch munter und vernünfftige Thomaische Gedancken und Erinnerungen über allerhand außerlesene juristische Händel (1720-1721)*.[29] Die staatswissenschaftliche Dimension wird vorrangig durch die von Nicolaus Hieronymus Gundling herausgegebenen *Gundlingiana*[30] *(1715-1732)* und die *Neue Acerra philologica*[31] *(1715-1723)* abgedeckt. An dieser Stelle gilt es im Projekt Maße zu entwickeln, die Aussagen darüber ermöglichen, ab welcher ‚Beteiligungsintensität' und ‚Akteursvielfalt' von einem Diskurs gesprochen werden kann. Zudem erforderlich ist die Etablierung unterschiedlicher Diskurstypen (beispielsweise Laien- oder Expertendiskurs).

Erst in der zweiten Periode zwischen 1726 und 1756 lassen sich einzelne Autoren identifizieren. Auffällig ist, dass die grundlegende Themenentwicklung in ihren Schwerpunkten den allgemeinen Tendenzen des halleschen Zeitschriftenmarktes, der zu diesem Zeitpunkt unterhaltenden Schriften einen größer werdenden Raum eröffnete, folgt. Dieser Beobachtung widersprechend sticht die pädagogische Konnotation des Themas heraus, die sich in dieser Intensität nicht im Gesamtspektrum der halleschen Journale wiederfindet (Abbildung 5).

Diese thematische Konzentration wird hauptsächlich von der pädagogischen Zeitschrift *Altes und Neues von Schulsachen (1752-1755)*, herausgegeben von Johann Gottlieb Biederman getragen. An diesem Beispiel kann belegt werden, dass, da die grundlegende Auseinandersetzung mit dem englischen Thema noch sehr gering ausfiel, bereits ein Journal eine wesentliche Verschiebung in der inhaltlichen Ausrichtung bedingen konnte ohne dabei zwingend eine Kontroverse initiieren zu müssen. Im Unterschied dazu beschäftigten sich die literarisch-kulturelle Publikumszeitschrift *Bemühungen zur Beförderung der Critik und des guten Geschmacks (1743-1747)*, herausgegeben von Christlob Mylius und Johann Andreas Cramer und die von Friedrich Eberhard Rambach veröffentlichte theologische Fachzeitschrift *Beyträge zur*

---

29 Ein weiteres Journal Thomasius' das diesen Themenbereich berührt ist die *Historie der Weißheit und Thorheit (1693)*.

30 Vollständiger Titel: *Gundlingiana: darinnen allerhand zur Jurisprudentz, Philosophie, Historie, Critic, Litteratur und übrigen Gelehrsamkeit gehörige Sachen abgehandelt werden*.

31 Vollständiger Titel: *Neue Acerra philologica oder Gründliche Nachrichten aus der Philologie, und denen römischen und griechischen Antiquitaeten: darinnen die schweresten Stellen aller auctorum classicorum der studierenden Jugend zum Besten in einer angenehmen Erzehlung kürtzlich und gründlich erkläret werden*.

Beförderung der Erkänntniß (1754-1759)[32] mit durchaus variierenden Elementen des englischen Themas.

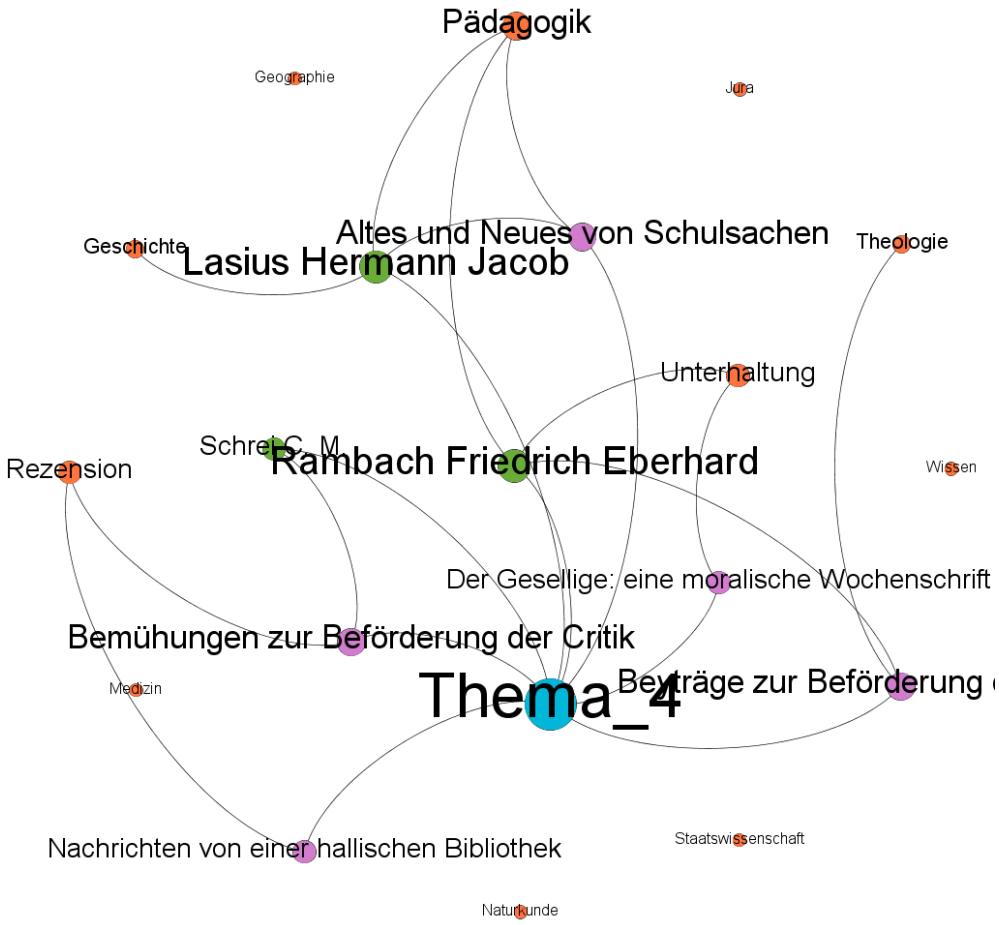


Abbildung 5: Thema 4 Netzwerk zwischen 1726 und 1756. Violet = Journale, rot = Fachzuordnung, grün = Autoren, Knotengröße und Schriftgröße = Gewichteter Grad.

---

32 Vollständiger Titel: *Beyträge zur Beförderung der Erkäntniß der Wahrheit zur Gottseligkeit aus den Schriften ausländischer und bewährter Gottesgelehrten*, als Fortsetzung der in Leipzig und Greifswald 1750 erscheinenden *Samlung auserlesener Abhandlungen ausländischer Gottesgelehrten, zur Unterweisung des Verstandes und Besserung des Herzens*.

Auch für die hervorstechenden Autoren kann konstatiert werden, dass aufgrund der noch eingeschränkten Breite des Themas einzelne Personen schnell zu einer relativen Verzerrung führen können. Sowohl Rambach, als auch Lasius beteiligten sich nur punktuell an einer Darstellung englischer Themen, da jedoch kaum weitere Autoren namentlich greifbar gemacht werden können, erscheinen die betreffenden Personen prägender als sie es tatsächlich waren. Um diesen, einem Mangel an Daten geschuldeten Verzerrungen zu begegnen, muss im Projektverlauf in Hinblick auf die Netzwerkrekonstruktion eine Lösung für den Umgang mit anonym erscheinenden Beiträgen angestrebt werden.

In der dritten Phase zwischen 1757 und 1787 weitete sich, entsprechend der allgemeinen Expansion des halleschen Zeitschriftenmarktes, die Anzahl der involvierten Journale und Autoren aus, wodurch zugleich inhaltliche Interessenverschiebungen bedingt wurden. Geschichte und Geographie rückten, gerade bei der Betrachtung ausländischer Staaten zunehmend in den Mittelpunkt, wie auch Fragen des Rechts. Zugleich offenbart sich, dass personelle Verknüpfungen, auch über Staatsgrenzen hinweg vor allem von Naturwissenschaften und Theologie getragen wurden. In diesem Bereich bestand relativ früh ein intensives Kommunikationsbedürfnis, das die gelehrten Briefwechsel endgültig in den Hintergrund drängte und zu einer personalisierten Wissenschaftspraxis in den naturwissenschaftlichen und theologischen Journalen führte. Eine Vielzahl herausragender Naturwissenschaftler publizierte Forschungsergebnisse in naturwissenschaftlichen und medizinischen Zeitschriften und auch Theologen nutzten die Möglichkeit, durch Publikationen in Journalen ihre Überzeugungen und Überlegungen einer größeren – auch weniger gebildeten – Öffentlichkeit zur Verfügung zu stellen. Parallel gewährleisteten Kompilationen, Zitationen, Rezensionen und Übersetzungen in beiden Bereichen eine relativ schnelle Verbreitung. Die hervorstechenden Autoren sind aufgrund dessen, mehrheitlich in Naturwissenschaften und Theologie verankert (Abbildung 6).

Generell ist der Stellenwert theologischer Betrachtungen und Abhandlungen interessant. Diese wurden zum einen von einflussreichen Autoren getragen, andererseits spiegelt sich hier der in Halle, auch in der publizistischen Praxis, lange Bestand habende theologische Schwerpunkt wider. Dieser führte nahezu im gesamten Betrachtungszeitraum zu einer beeindruckenden Stabilität theologischer Journale und Themen. Einige von ihnen setzten sich dabei offensichtlich auch mit theologischen Fragen in anderen räumlichen Kontexten auseinander. Zu ihnen gehörten beispielsweise das *Magazin für die Religion (1780-1781)* herausgegeben von Johann Salomo

Semler, die *Betrachtungen über die Werke Gottes (1772)*[33] veröffentlicht durch Christoph Christian Sturm oder der ebenfalls von Sturm, gemeinsam mit Jakob Friedrich Feddersen, herausgegebene *Der Andächtige (1773-1774).* Gleichbleibend finden sich Rezensionen in allen fachlichen Bereichen. Diese thematischen Schwerpunkte wurden zwischen 1757 und 1787 von einer Vielzahl von Journalen getragen, wie auch die generelle Zunahme von periodisch erscheinenden Schriften eine Verzerrung des Gesamtbildes, wie im zweiten Zeitabschnitt, zunehmend verhinderte.

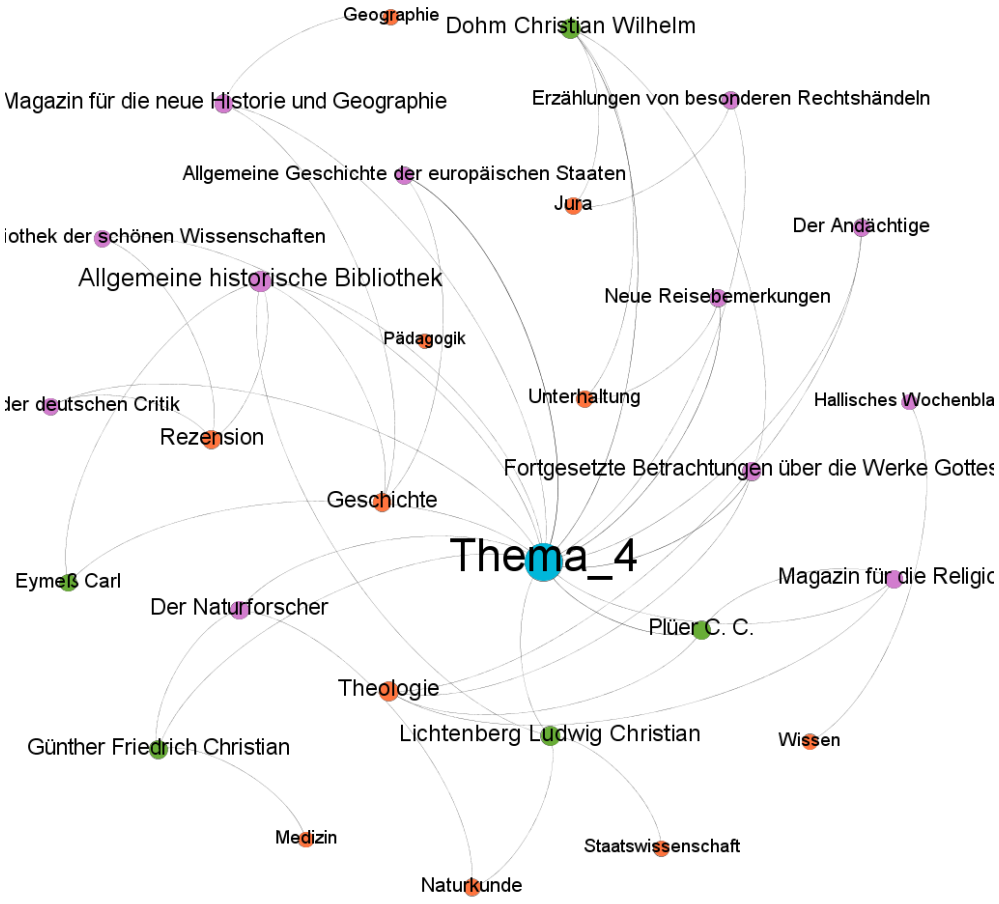


Abbildung 6: Thema 4 Netzwerk zwischen 1757 und 1787. Violet = Journale, rot = Fachzuordnung, grün = Autoren, Knotengröße und Schriftgröße = Gewichteter Grad.

---

[33] Vollständiger Titel: *Betrachtungen über die Werke Gottes im Reiche der Natur und der Vorsehung auf alle Tage des Jahres.*

Für die letzte Periode zwischen 1788 und 1818 ist ein neuerlicher Anstieg der Beiträge zu verzeichnen, wobei die inhaltliche Vielfalt gegenüber zwei klaren, wenn auch divergierenden, Interessenschwerpunkten abnahm (Abbildung 7).

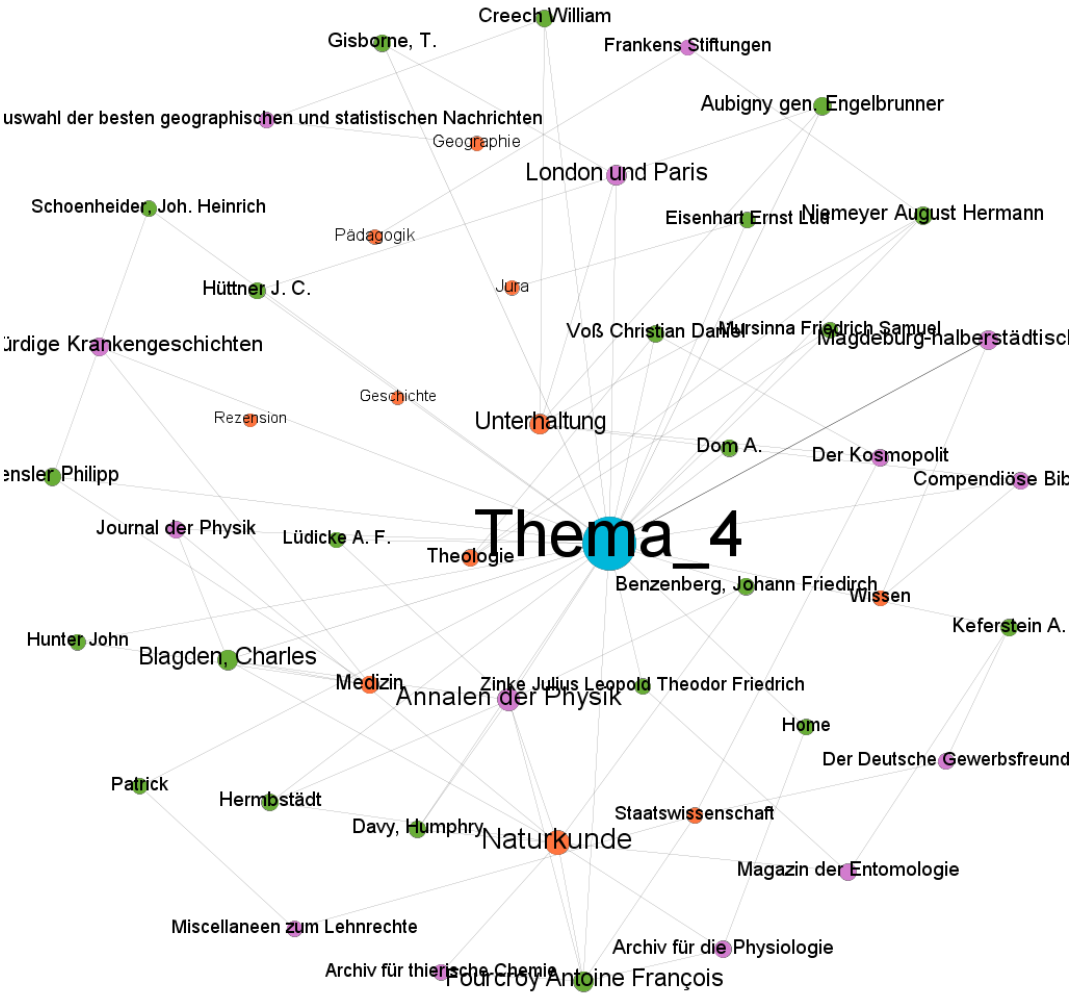


Abbildung 7: Thema 4 Netzwerk zwischen 1788 und 1818. Violet = Journale, rot = Fachzuordnung, grün = Autoren, Knotengröße und Schriftgröße = Gewichteter Grad.

Diese zwei Hauptinteressengebiete bilden die Unterhaltung, in ihrem Schwerpunkt getragen vom Journal *London und Paris*, das über alle interessant erscheinenden Moden, kulturellen Ereignisse, Literatur, bildende Künste et cetera aus den beiden Metropolen Paris und London informierte. Ebenso berichteten kosmopolitisch ausgerichtete Zeitschriften und die allgemeine Tagespresse mit wechselnder Intensität aus und über London. Der diesem Bereich gegenüberliegende Schwerpunkt bestand in naturwissenschaftlichen

Abhandlungen. Qualitativ hochwertige naturwissenschaftliche Journale (insbesondere die *Annalen der Physik*) prägten seit 1790 den halleschen Markt. Es gelang zahlreichen medizinisch-naturwissenschaftlichen Zeitschriften sich in Halle zu etablieren, wobei betont werden muss, dass sich in diesen Fällen der Verbreitungsraum weit über die Grenzen Halles erstreckte. So weisen das *Magazin für Entomologie,* das *Archiv für Physiologie,* das *Journal der Physik* oder die *Merkwürdigen Krankengeschichten* eine intensive inhaltliche und personelle Vernetzung weit über die Grenzen der halleschen Universität auf. Der sich in der zeitlichen Darstellung des TopicExplorers offenbarende massive Anstieg seit 1797 (Abbildung 3) wurde damit zu großen Teilen von einer etwa gleich gewichteten Mischung aus neu begründeten naturwissenschaftlichen und unterhaltenden periodischen Schriften getragen. Im Unterschied dazu können kaum noch prägende Autoren ausgemacht werden. Andererseits war es unter Umständen das Verdienst von Günther, Rambacher, Dohm und anderen ein breites Interesse an England geweckt zu haben, das in den Jahren nach 1800 zwar zu einer zunehmenden Konzentration auf zwei Themenschwerpunkte führte, dabei jedoch von einem beeindruckenden Personenkreis getragen wurde.

Dieses beispielhafte Netzwerk soll im weiteren Projektverlauf um zusätzliche Informationen ergänzt werden, wie auch die zeitliche Einteilung entlang des im TopicExplorers generierten Verlaufs oder herausragenden historischen Ereignissen variierend strukturiert werden wird.

5 Ausblick

Dieser kurze Einblick in die Zielsetzung des Projekts sollte an einem kleinen Ausschnitt beleuchten, worin die Chancen und Möglichkeiten einer Kombination von Topic-Modeling und Netzwerkanalyse liegen. Keiner der beiden Ansätze allein würde eine solche Fülle von Daten berücksichtigen können und zudem flexibel angepasste Interpretationen und Vergleiche ermöglichen. Wendet man die dargestellte Herangehensweise auf die Gesamtheit der generierten 30 Topics und ihrer Repräsentanz in den halleschen Journalen an und nimmt zunächst noch keine weitere Unterscheidung in unterschiedliche zeitliche Phasen vor – kann ein Blick auf das Gesamtnetzwerk zwischen 1689 und 1824 bereits erste Eindrücke vermitteln. Berücksichtigt wurden in der folgenden Darstellung je Topic die zehn relevantesten Journale (Abbildung 8). Die struktuelle Vernetzung zwischen einzelnen Themen kann dabei über Netzwerkdichte und durchschnittlichen Grad betrachtet werden.

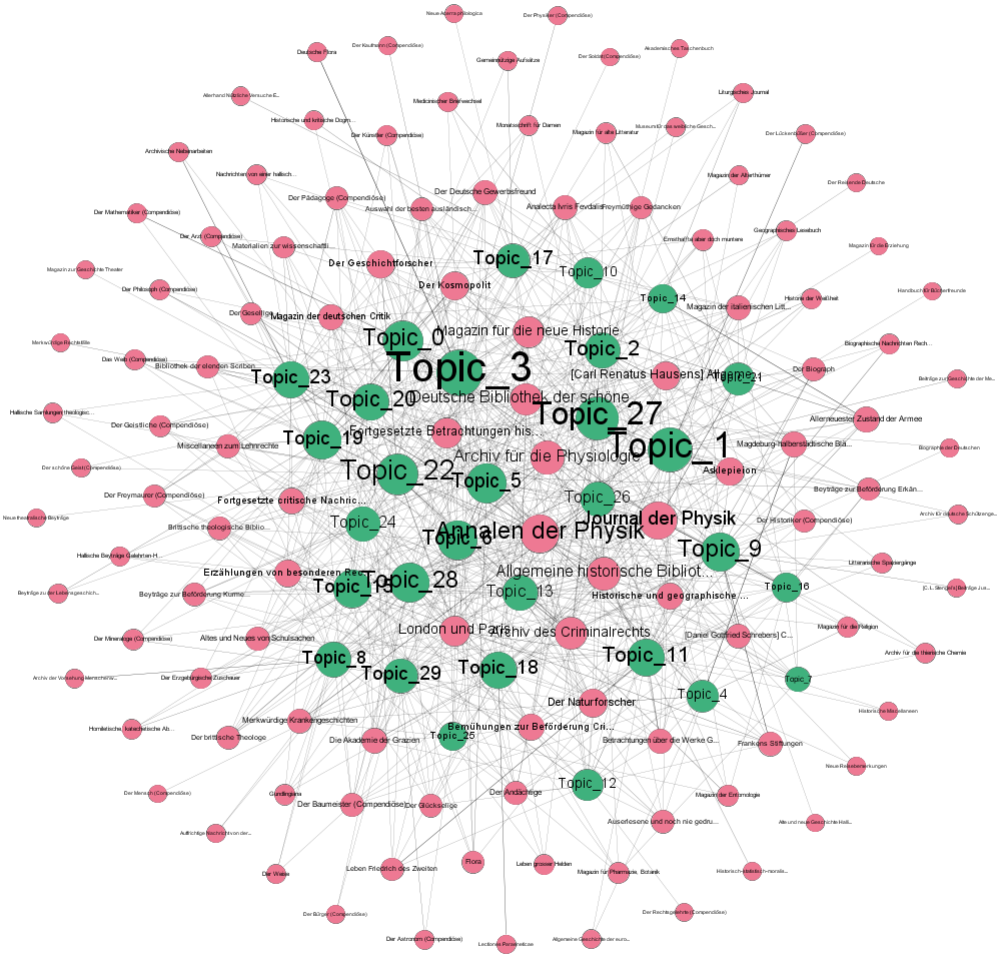


Abbildung 8: Topics mit Journalen (1689-1817). Grün = Topic, pink = Journal, Knotengröße = gewichteter Grad, Schriftgröße = Betweenness Centrality. Netzwerkdurchmesser = 4, mittlerer gewichteter Grad = 70,2.

Als besonders zentrale Themen und damit unter Umständen (Wissens)Diskurse, über den Gesamtzeitraum, erweisen sich Topic 3 mit Rezensionen und Beiträgen zu politischen und staatswissenschaftlichen Schwerpunkten, Topic 1 aktuelle Nachrichten zu Politik und vor allem regionaler Wirtschaft beinhaltend und Topic 27 mit eher beschreibenden Artikeln zu Geschichte und Verfassung unterschiedlicher Länder. Andere Themen, die häufig nur punktuell auftreten, beziehungsweise relativ spät in den Journalen der Aufklärungsepoche Raum erhalten, finden sich eher am Rand des Netzwerks. So beispielsweise Topic 7 mit den Kernpunkten elektrische Versuche und der Reflektion über die Erkenntnisse Alessandro

Voltas oder Topic 17, dessen Hauptaugenmerk auf Beschreibungen deutscher Städte, insbesondere ihrer Universitätskultur liegt. Analog dazu können relativ einfach Journale ausgemacht werden, die, da sie eine Vielzahl von Themen bedienen, in zentralen Positionen des Netzwerks gefunden werden können. Dazu zählten zum Beispiel die *Allgemeine historische Bibliothek* (1767-1771), die *Annalen der Physik* (1799-1824) oder die *Deutsche Bibliothek der schönen Wissenschaften* (1767-1770). Diese Periodika stechen dabei nicht zwingend mit einem besonders langen Erscheinungszeitraum hervor, sondern vielmehr mit ihrer Fähigkeit unterschiedlichste Themenbereiche umfangreich zu bedienen. Nicht ohne Grund finden sich unter ihnen zahlreiche Rezensionszeitschriften, ohne expliziten fachlichen Schwerpunkt. Betont werden muss an dieser Stelle noch einmal, dass es sich bei dieser Darstellung aktuell noch um eine Momentaufnahme handelt, da noch nicht alle halleschen Journale des betreffenden Zeitraums in die Topic-Modellierung einbezogen werden konnten.

Nichts desto trotz kann eine Unterscheidung in (zunächst) vier, der zeitlichen Differenzierung von Thema 4 entsprechende Abschnitte eine differenziertere Sichtweise auf die vorliegenden Daten und die Tendenzen der thematischen Entwicklung des halleschen Zeitschriftenmarktes ermöglichen (Abbildung 9-12), um detailliertere Informationen über Verschiebungen, Brüche und Entwicklungen in den Interessenschwerpunkten der halleschen Journale generieren, beziehungsweise einen Vergleich gestatten zu können.

Für den Zeitraum 1689-1725 (Abbildung 9) wird offensichtlich, dass sich der Zeitschriftenmarkt in Halle noch in den Kinderschuhen befand und nur ein sehr dünnes Wissensnetzwerk etablieren konnte. Acht Journale finden sich in diesem Zeitabschnitt wieder, berücksichtigt werden muss dabei eine gewisse Verzerrung, da nur die jeweils für ein Topic relevantesten Zeitschriften in die Visualisierung inbegriffen sind. Journale, die sich in geringerer Intensität an Themen beteiligen, werden zunächst noch ausgeschlossen. Deutlich wird bei den zentralen Journalen dieser Phase aber ihre Positionierung zwischen unterschiedlichen Themen. In Hinblick auf die Themen selbst ist erkennbar, dass eine Vielzahl von Themen noch nicht angesprochen wird, während alle relevanten Themen theologisch geprägt erscheinen. Im zweiten Zeitabschnitt zwischen 1726 und 1756 (Abbildung 10) ist eine zentrale Zeitschrift erkennbar, die insbesondere durch ihre intensive Vernetzung zu einer Vielzahl unterschiedlicher Themen heraussticht, während die *Gundlingiana*, mit einer thematischen Verbindung zu vier Topics, zwar mit der Intensität ihrer Relationen beeindruckt, sich aber dennoch eher am Rand des Netzwerkes findet. Inhaltlich rücken Themen zu Staatswissenschaften, Geschichte und Geographie (Topic 27) und Rezensionsthemen mit staatswissenschaftlichen Schwerpunkten (Topic 15) in den Mittelpunkt.

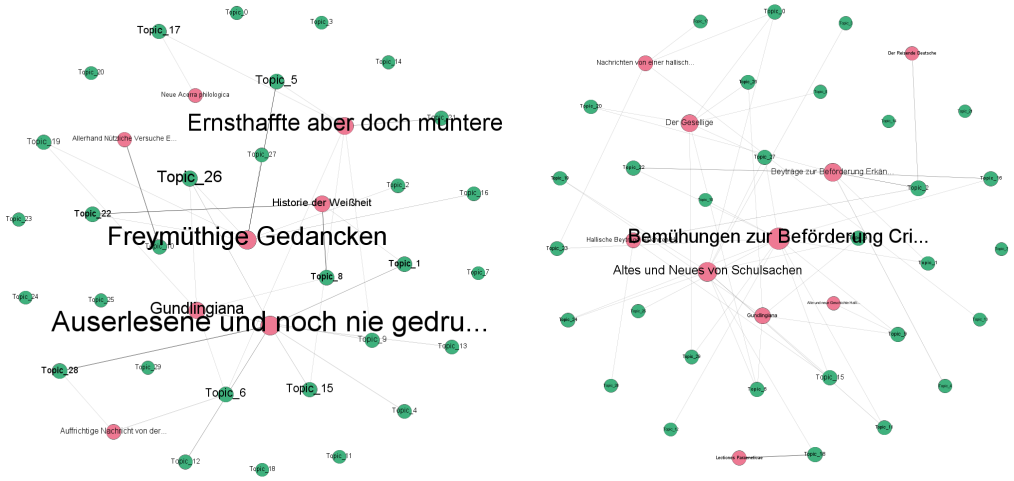


Abbildung 9 und 10: Themen und Journale 1689-1725 und 1726-1756. Grün = Topic, pink = Journal, Knotengröße = gewichteter Grad, Schriftgröße = Betweenness Centrality. Abbildung 9: Netzwerkdurchmesser = 6, mittlerer gewichteter Grad = 11,4. Abbildung 10: Netzwerkdurchmesser = 7, mittlerer gewichteter Grad = 16,6.

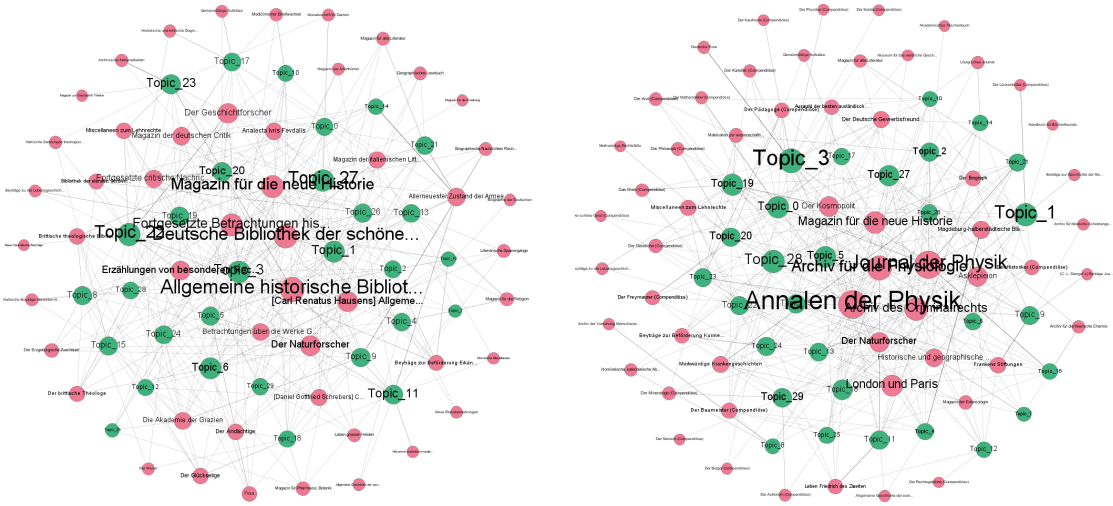


Abbildung 11 und 12: Themen und Journale 1757-1787 und 1788-1817. Grün = Topic, pink = Journal, Knotengröße = gewichteter Grad, Schriftgröße = Betweenness Centrality. Abbildung 11: Netzwerkdurchmesser = 4, mittlerer gewichteter Grad = 59,2. Abbildung 12: Netzwerkdurchmesser = 4, mittlerer gewichteter Grad = 60,4.

Für den dritten Abschnitt zwischen 1757-1787 (Abbildung 11) muss das Kantengewicht relativiert dargestellt werden, denn das Netzwerk verdichtet sich auf allen Ebenen, obwohl noch einzelne hervorragende Journale auszumachen sind, scheint der hallesche Zeitschriftenmarkt eine Vielfalt angenommen zu haben, die unterschiedlichsten periodischen Publikationen eine Existenz bieten konnte. Auf den ‚ersten Plätzen' (nach Betweenness Centrality) der Journale finden sich zwei allgemeine Rezensionsorgane (*Allgemeine historische Bibliothek*, Betweenness 260.9 und *Deutsche Bibliothek der schönen Wissenschaften*, Betweenness 254.2) und mit schon deutlicherem Abstand eine historische Zeitschrift (*Magazin für die neue Historie*, Betweenness 216.6). Zu diesem Zeitpunkt scheinen nahezu alle Themen bedient zu werden, wenn auch eine staatswissenschaftliche (Topic 27 und Topic 3) und eine literarisch-kulturelle (Topic 20 und Topic 6) Fokussierung auffällt.

In der letzten Phase zwischen 1788-1817 (Abbildung 12) dünnt sich das Netzwerk deutlich aus, verstärkt tritt zudem eine Differenzierung zwischen Themen und Journalen hervor, sodass von der relativen Ausgewogenheit insbesondere zwischen den Themen der vorangegangenen Phase nicht mehr gesprochen werden kann. Naturwissenschaftliche Journale drängen ins Zentrum des Netzwerkes und weisen einen deutlichen Vorsprung, in Hinblick auf ihre Zentralität, vor Zeitschriften mit anderen fachlichen Schwerpunkten auf. Ähnlich, wenn auch nicht in einer solchen Intensität und Vielzahl präsent sind unterhaltende Journale. Insbesonder theologische Themen und Journale werden an den Rand des Netzwerkes gedrängt. In diesem letzten Zeitabschnitt sind alle Themen integriert, das heißt, dass im Zeitverlauf trotz vielfältiger thematischer 'Neuerungen' und Erweiterungen keine Topics wieder vollkommen aus dem Netzwerk herausfallen. Auch wenn Sie deutlich seltener aufgegriffen wurden, blieben sie grundlegend bestehen. Die Themen dieser Phase werden dominiert von Landesbeschreibungen unterschiedlicher fachlicher Ausrichtung, physikalischen Abhandlungen aber auch (über)regionalen Nachrichten.

Die Gegenüberstellung der beiden letzten Zeitschnitte erweckt dabei den Eindruck eines relativ klaren Übergangs von historisch-literarisch-kulturellen Themen hin zu einer klaren Dominanz der Naturwissenschaften am Ende des 17. Jahrhunderts. Gerade diese klar erscheinenden Konzentrationsprozesse auf bestimmte Themen werden jedoch immer wieder durchbrochen von Zeiträumen, die durch eine Vielzahl von Themen und Journalen geprägt waren. Diese Beobachtung legt die Überlegung nahe, dass nach Abschluss eines relativ intensiv thematisierten Gegenstandes, innerhalb der periodischen Presse die Suche nach neuen, das Publikum und die Autoren interessierenden, Fragestellungen begann. Gerade hier bietet die Flexibiltät der Netzwerkanalyse gute Ansatzpunkte. Betrachtet man den ‚Zwischenzeitraum' 1777 bis 1797, also die jeweils 10 Jahre vor und nach dem obigen Schnitt ergibt sich ein ausgewogenes Bild (Abbilung 13). Bereits die Visualisierung lässt

erkennen, dass die Themen sich stärker aneinander annähern und von einer deutlich höheren Anzahl von Journalen ‚umgeben' werden. Für diesen Zeitraum erhöht sich die Anzahl der relevanten Themen deutlich. Im Vergleich weisen nur elf Themen für den Zeitraum 1757-1787 eine Betweenness >100 auf, im anschließenden Abschnitt 1788-1817 sind dies vierzehn Themen. Im Zwischenzeitraum jedoch finden sich neunzehn Themen mit einer solchen Betweenness im Netzwerk wieder, was auf eine relative ‚Ausgewogenheit' des thematischen Spektrums in diesem Zeitabschnitt hinweist.

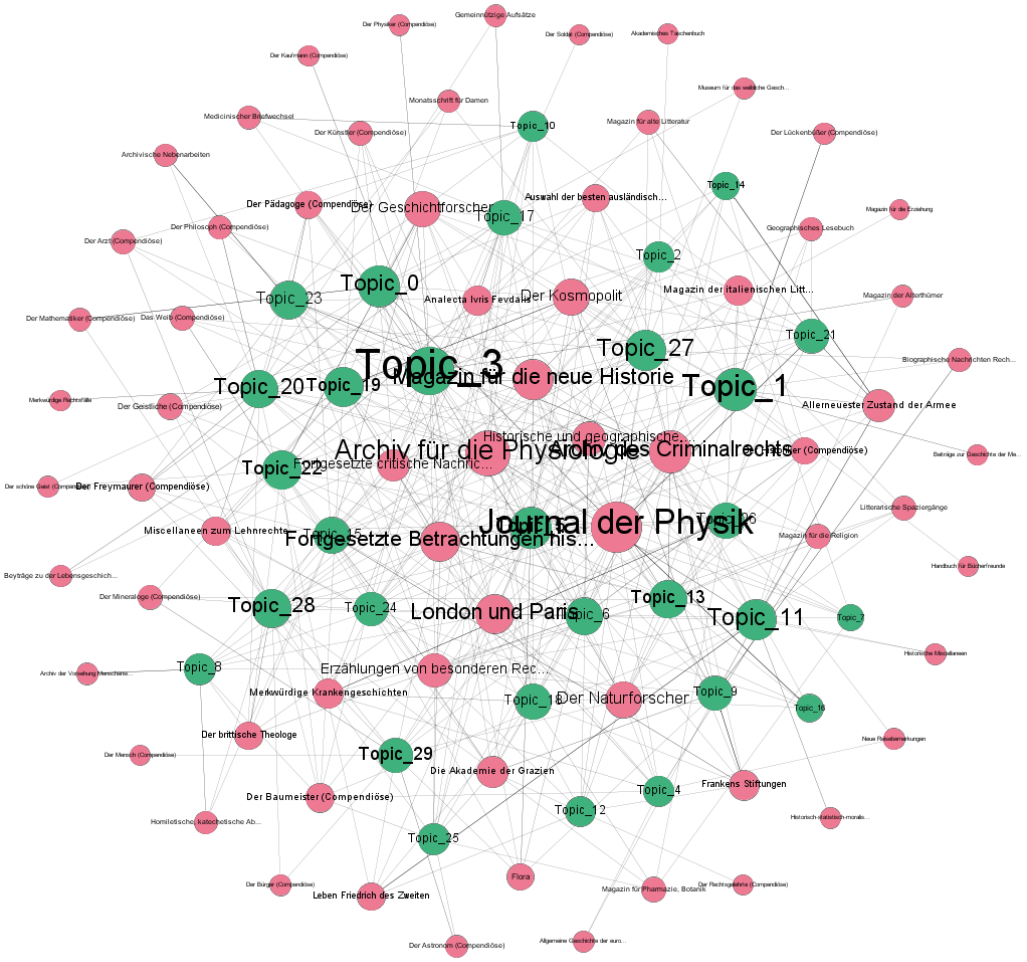


Abbildung 13: Themen und Journale 1777-1797. Grün = Topic, pink = Journal, Knotengröße = gewichteter Grad, Schriftgröße = Betweenness Centrality. Abbildung 11: Netzwerkdurchmesser = 4, mittlerer gewichteter Grad = 56,7.

Die in unseren ersten Versuchen generierten Themen geben Hinweise darauf, worüber, wann und in welchem Umfang in den halleschen Zeitungen und Zeitschriften gesprochen wurde. Ersichtlich wird aber auch, dass Topic-Modelle Themen hervorbringen können, die nach Themen aussehen, sich aber bei einer näheren Betrachtung eher als Relationen anderer Art entpuppen, so zum Beispiel geographische oder territoriale Eigennamen. Gleichzeitig wurden auch Gruppen generiert, die zunächst keinen semantischen oder thematischen Zusammenhang erkennen lassen, während andere, erwartete Themen (zum Beispiel das Thema 'Aufklärung') zu fehlen scheinen. Doch auch in diesen Fällen werden die Daten nicht mehr nur selektiv als 'Beispielspender' genutzt, sondern in allen ihren Widersprüchlichkeiten und zunächst auch einmal mangelnden Plausibilitäten ernst genommen. Für die spätere Analyse der erhaltenen Themen muss darauf Wert gelegt werden, dass alle Ergebnisse des Topic-Modeling einbezogen werden, nicht nur die positiv erscheinenden. In sich anschließenden Schritten führt dies dazu, sowohl die Wortlisten konstanter zu pflegen als auch die Voraussetzungen der Modellierung an die spezifischen Bedürfnisse des Nutzers und seiner Texte anzupassen. Einfluss auf die Ergebnisse haben zweifelsohne die Korpusgröße und -zusammenstellung und ebenso die Parameter des Topic-Modells (zum Beispiel die Topic-Anzahl, die Segmentlänge et cetera). Ein weiteres Problem stellen die zahlreichen erforderlichen Vorarbeiten dar. Diese beschränken sich nicht nur auf die Bereinigung der Wortlisten, sondern sollten für weitere Versuche ebenso Übersetzungen beinhalten. Hier erscheint es als durchaus sinnvoll und wünschenswert, zumindest einige dieser Vorarbeiten direkt in den TopicExplorer zu integrieren, um seine grundlegende Anwendbarkeit auf historische Dokumente weiter zu verbessern.

Als durchaus belastbar hat sich die Einbeziehung von Metadaten in die Datengrundlage des TopicExplorers erwiesen. SQL-Abfragen ermöglichen eine Filterung der mit den Topics in Beziehung stehenden Dokumente und die Überführung in eine Netzwerkanalyse, die zudem auf gewichtete Beziehungen anhand der Topic-Modellierung aufbauen kann. Perspektivisch sollen jedoch nicht nur diese Informationen in die Netwerkanalyse Eingang finden. Eine große Relevanz kommt den persönlichen Beziehungen zwischen den Akteuren zu, wobei Akteure in diesem Zusammenhang nicht länger nur als Autoren verstanden werden dürfen (wie im aktuellen Beispiel geschehen). Beabsichtigt ist vielmehr, die persönlichen Relationen einer Vielzahl von Akteuren in die Netzwerkmodellierung einzubeziehen, dazu gehören beispielsweise Drucker, Verleger oder Kolporteure. Die Betreffenden sollen in unserem Projekt dabei nicht länger über ihren sozialen Status oder ihre familiäre Herkunft, sondern vielmehr über ihr Wissen, die Möglichkeiten dieses zu generieren und andere Akteure an ihm partizipieren zu lassen definiert werden, dadurch rückt ihr

symbolisches Kapital, verstärkt durch die Theorie der Wissensnetzwerke,[34] in den Vordergrund. Denn die Journale der Aufklärung müssen zwingend als öffentliches Netzwerk verstanden werden, in dem sich inhaltliche und strukturelle (persönliche) Verbindungen überlagerten und ergänzten. Der von ihnen dadurch, teilweise neu definierte öffentliche Raum war die entscheidende Voraussetzung für die Zirkulation von Wissen während dieser Epoche.

Literatur

Anthes, Gary. 2010. „Topic models vs. unstructured data." In Commun. ACM 53 (12), 16–18. DOI: 10.1145/1859204.1859210.

Benkler, Yochai; Roberts, Hal; Faris, Robert; Solow-Niederman, Alicia; Etling, Bruce. 2015. „Social Mobilization and the Networked Public Sphere: Mapping the SOPA-PIPA Debate." In Political Communication 32 (4), 594–624.

Blei, David M. 2012. „Probabilistic Topic Models." In Commun. ACM 55 (4), 77–84. DOI: 10.1145/2133806.2133826.

Blei, David M.; Lafferty, John D. 2006. „Dynamic Topic Models." In Proceedings of the 23rd International Conference on Machine Learning. New York, ACM (ICML '06), 113–120.

Breuel, Thomas M. (2008). „TheOCRopusOpenSourceOCRSystem." In Document Recognition and Retrieval XV ed. Berrin A. Yanikoglu; Kathrin berkner, San Jose, 68150F-68150F-15.

Crane, Gregory. 2006. „What Do You Do with a Million Books." D-Lib Magazin (12).
Online: http://www.dlib.org/dlib/march06/crane/03crane.html, zuletzt geprüft am 19.12.2017.

Chang, Jonathan; Gerrish, Sean; Wang, Chong; Boyd-Graber, Jordan L.; Blei, David M. 2009. „Reading Tea Leaves: How Humans Interpret Topic Models." In Advances in Neural Information Processing Systems 22: Curran Associates, ed. Y. Bengio, D. Schuurmans, John D. Lafferty, C. K. I. Williams und A. Culotta, 288–296. Red Hock, NY: Curran.

Dahlgren, P. 2005. „The Internet, Public Spheres, and Political Communication. Dispersion and deliberation". In Political Communication 22, 147–162.

Eymer, Wilfried. 2004. Eymers Pseudonymen-Lexikon. Realnamen und Pseudonyme in der deutschen Literatur. Wiesbaden: Marix-Verlag.

Fruchterman, Thomas M. J.; Reingold, Edward M. 1991. „Graph drawing by force-directed placement." In Softw: Pract. Exper. 21 (11), 1129–1164.

[34] Vgl. Henningsson und Hanseth, „The essential".

Gierl, Martin. 2001. „Kompilation und die Produktion von Wissen im 18. Jahrhundert." In Die Praktiken der Gelehrsamkeit in der frühen Neuzeit, ed. Helmut Zedelmaier und Martin Mulsow, 63–94.

Henningsson, Stefan; Hanseth, Ole. 2011. „The essential dynamics of information infrastructures." In International Conference on Information Systems ICIS; 2011.3.

Hinneburg, Alexander; Preiss, Rico; Schröder, Renè. 2012. „TopicExplorer: Exploring Document Collections with Topic Models." In Machine Learning and Knowledge Discovery in Databases. European Conference, ECML PKDD 2012, ed. Peter A. Flach, 838-841, Bristol, UK.

Hinneburg, Alexander; Rosner, Frank; Pessler, Stefan; Oberländer, Christian. 2014. „Exploring Document Collections with Topic Frames." In Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, ed. Jianzhong Li. 2084-2086. New York.

Holzmann, Michael. 1906. Deutsches Pseudonymen-Lexikon. Wien und Leipzig: Akadem. Verlag.

Jacobi, Carina; van Atteveldt, Wouter; Welbers, Kasper. 2016. „Quantitative analysis of large amounts of journalistic texts using topic modelling." In Digital Journalism 4 (1), 89–106.

Jäger, Margarete; Jäger, Siegfried. 2007. Deutungskämpfe. Theorie und Praxis Kritischer Diskursanalyse. Wiesbaden: VS Verlag für Sozialwissenschaften.

Jurish, Bryan. 2016. „Canonicalizing Historical Text with CAB. Searching Linguistic Patterns in Large Text Corpora for Digital Humanities" Research July, 2016. Online verfügbar unter http://kaskade.dwds.de/~moocow/mirror/pubs/esu2016-jurish-cab.pdf, zuletzt geprüft am 04.01.2018.

Kataria, Saubrah; Mitra, Prasenjit; Caragae, Cornelia; Giles, C. Lee. 2011. „Context sensitive topic models for author influence in document networks." In Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, IJCAI-11. Barcelona, Catalonia, Spain, 16 - 22 July 2011 ed. Toby Walsh, 2247-2280, Menlo Park, Calif.: AAAI Press/Internat.

Kleinen-von Königslöw, Katharina. 2010. Die Arenen-Integration nationaler Öffentlichkeiten. Der Fall der wiedervereinten deutschen Öffentlichkeit. Wiesbaden: VS Verlag für Sozialwissenschaften.

Lim, Kar Wai; Buntine, Wray; Chen, Changyou; Du Lan. 2016. „Nonparametric Bayesian topic modelling with the hierarchical Pitman–Yor processes." In International Journal of Approximate Reasoning 78 (Supplement C), 172–191. DOI: 10.1016/j.ijar.2016.07.007.

Link, Jürgen. 2013. „Diskurs, Interdiskurs, Kollektivsymbolik." In Zeitschrift für Diskursforschung (1). Online verfügbar unter http://www.content-select.com/index.php?id=bib_view&doi=10.3262/ZFD1301007. zuletzt geprüft am 03.05.2018.

Mimno, David; Blei, David. 2011. „Bayesian Checking for Topic Models.“ In Proceedings of the Conference on Empirical Methods in Natural Language Processing. Stroudsburg, 227-237. PA, USA: Association for Computational Linguistics (EMNLP ’11).

Mladenova, Yana. 2016. Automatisierung der Identifikation von Autoren mit Pseudonymlexika am Beispiel der Halleschen Zeitschriften von 1688-1815. Bachelorarbeit. Martin-Luther Universität, Halle. Informatik. (Unveröffentlicht).

Papacharissi, Zizi. 2010. A private sphere. Democracy in a digital age. Cambridge: Polity (Digital media and society series).

Papilloud, Christian; Hinneburg, Alexander (2018): Qualitative Textanalyse mit Topic-Modellen. Eine Einführung für Sozialwissenschaftler. Wiesbaden: Springer Fachmedien Wiesbaden (Studienskripten zur Soziologie).

Parthey, Gustav. 1973. Die Mitarbeiter an Friedrich Nicolai’s Allgemeiner Deutscher Bibliothek nach ihren Namen und Zeichen in zwei Registern geordnet. Ein Beitrag zur deutschen Literaturgeschichte. Hildesheim: Gerstenberg.

Riddell, Allen: „Text Analysis with Topic Models for the Humanities and Social Sciences.“ Online verfügbar unter https://de.dariah.eu/tatom/, zuletzt geprüft am 20.06.2017.

Röder, Michael; Both, Andreas; Hinneburg, Alexander. 2015. „Exploring the Space of Topic Coherence Measures.“ In Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, 399-408. New York, NY, USA: ACM (WSDM ’15).

Sarasin, Philipp. 2011. „Was ist Wissensgeschichte?“ In Internationales Archiv für Sozialgeschichte der deutschen Literatur : IASL 36 (1), 159–172.

Schmid, Helmut. 1994 „Probabilistic Part-of-Speech Tagging Using Decision Trees.“ In International Conference on New Methods in Language Processing, Manchester UK, 44–49.

Sievert, Carson; Shirley, Kenneth. 2014. „LDAvis. A method for visualizing and interpreting topics.“ In Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces, 63-70. Baltimore, Maryland, USA: Association for Computational Linguistics.

Springmann, Uwe. 2016. „OCR für alte Drucke“. In Informatik Spektrum 39 (6), 459-462.

Stier, Sebastian; Bleier, Arnim; Lietz, Haiko; Strohmaier, Markus. 2018. „Election Campaigning on Social Media: Politicians, Audiences, and the Mediation of Political Communication on Facebook and Twitter.“ In Political Communication 35 (1), 50–74.

Stollwerk, Christoph. 2016. „Machbarkeitsstudien zu Einsatzmöglichkeiten von OCR-Software im Bereich „Alter Drucke“ zur Vorbereitung einer vollständigen Digitalisierung deutscher Druckerzeugnisse zwischen 1500 und 1930.“ In DARIAH-DE working paper 6.

Tangherlini, Timothy R.; Leonard, Peter. 2013. „Trawling in the Sea of the Great Unread: Sub-corpus topic modeling and Humanities research." In Poetics 41 (6), 725–749. DOI: 10.1016/j.poetic.2013.08.002.

Weller, Emil. 1886. Lexicon pseudonymorum. Wörterbuch der Pseudonymen aller Zeiten und Völker oder Verzeichnis jener Autoren, die sich falscher Namen bedienten. Regenburg: Coppenrath.

ELO, KIMMO

Geospatial Social Networks of East German Opposition (1975-1989/90)

Journal of Historical Network Research 2 (2018) 143-165.

Keywords

GDR, opposition, regional connectedness, Roland Jahn

Abstract

During the last two decades single photographs and photograph corpora have gained in popularity as sources for historical research. In addition to their important function as carriers of the past, photographs also contain valuable information about past social relations. However, to utilise this information a researcher needs a more structured dataset, a photograph corpus containing rich metadata, which allows us to explore and analyse contextual information stored in alphanumeric form. My paper will exemplify how photography corpora could be used as a source for network analysis seeking to explore, reconstruct and visualise hidden historical social networks. The empirical case of my paper revolves around regional and interregional networks of East German dissident movement. The main empirical material explored for network analysis and visualisations consists of a large enriched photograph corpus on East German dissident movement maintained by Robert Havemann Foundation in Berlin. Based on this corpus my paper will explore the structure and dynamics of regional and interregional networks of East German opposition. The results introduce evidence that regional connectedness based on personal mobility among the East

German dissidents both changes and increases over time, thus resulting in continuously evolving patterns of social interaction. Further, the analysis of Roland Jahn's geospatial networks evidences the usefulness and power of historical network analysis when it comes to tackling changes in patterns of social interaction.

1 Introduction*

Sometimes a single event can trigger people to take the step that turns a loose group of friends into a tightly connected, political network. For the topic of this article, the regional networks of the East German opposition, the death of young political activist Matthias Domaschk on 12 April 1981 in Gera in a pretrial detention of the East German security service (*Ministerium für Staatssicherheit*, abbr. the Stasi) after 13 hours of continuous interrogations, was that trigger. On 16 April 1981 Domaschk was buried in Jena in the presence of almost three hundred friends and collaborators, among them also Roland Jahn, a close friend of Domaschk.[1] For Jahn and his close friends the death of Domaschk was a turning point, the triggering event to get radical and start creating an underground platform for political opposition.

In March 1983 Jahn and his collaborators succeeded in establishing *Jenaer Friedensgemeinschaft* (Jena Peace Community) as an organised form of political opposition. The Jena Peace Community rapidly became one of the most important forums of organised political activism in the system-immanent competition between the Jena dissident community and the ruling Socialist Unity Party (*Sozialistische Einheitspartei Deutschlands*, SED). The community was also one of the first major opposition communities established *outside* the protective walls of the Evangelical church. Founders of the Jena Peace Community were disillusioned with the reluctant resistance of the Evangelical church against the state repression and, hence, sought to establish a new, independent platform under the umbrella of the European Peace movement. Another reason was the fact that the *Junge Gemeinde* (Young Congregation), from which the Jena Peace Community emerged, was heavily infiltrated and

---

* Acknowledgments**: This article is based on a paper read at the 12th Historical Network Research** Workshop „Historische Netzwerkforschung. Kommunikation in Netzwerken – Netzwerke der Kommunikation. Thematische, methodische und theoretische Perspektiven historischer Netzwerkanalyse zwischen Geschichts- und Kommunikationswissenschaft" convened by Matthias Bixler and Erik Koenen and held from 20-21 April 2018 at the University of Bremen.

Corresponding Author: Dr Kimmo Elo, Centre for Parliamentary Research, University of Turku, kimmo.elo@utu.fi

[1] "Biografie Matthias Domaschk", edited by Bundeszentrale für politische Bildung and Robert-Havemann-Gesellschaft e.V., last modification: October 2017. [Online: www.jugendopposition.de/145407. Last visited: 27 June 2018].

observed by the Stasi after the expulsion of Wolf Biermann[2] in 1976. These circumstances resulted in a growing demand for a new political platform residing outside the church.[3]

Taking the Jena dissident scene as its starting point this article has two main objectives. The first objective is to analyse the structure and dynamics of the geospatial social networks based on selected key figures in the Jena dissident community. By exploring the networks of these dissidents, we are able to discover how oppositional thoughts spread not just among people, but also between geographical regions. Within these networks we focus on changes in Roland Jahn's social networks and seek to evidence, how changes in his social networks correlate with changes in his personal biography. The second objective is to evidence the usability of digitised historical photography corpora as a source for network re-construction. In this respect we will also tackle questions related to the quality of meta-data and to data preparation process. Since historians and social scientist interested in benefitting from the growing availability of digitised, non-native digital materials, we should also pay (more) attention to problems emerging from the digitisation process itself, i.e. problems related to poor image quality, missing or incomplete meta-data or false positives. Since quantity cannot replace quality, scholars in Digital Humanities exploiting methods of distant reading should pay specific attention to source criticism and data quality in order to avoid biased interpretations.

The bulk of studies on the GDR opposition and resistance have thus far been dominated by research focusing on certain phenomena or periods of time from the perspective of historical research. Undoubtedly, these studies have enriched and improved our understanding of this complex phenomenon. Complex, because we need to keep in mind the pitfalls hiding it the source material. Hence, on the one hand, it is a well-known fact, that the Stasi never fully understood the informal logic, thinking or motives behind its most important internal enemy, and as a result, the Stasi apparatus produced a lot of false interpretations and incorrect assessments. On the other hand, due to the permanent risk of leaks, opposition and dissident groups were extremely careful when it came to sharing written materials or other documents.[4] However, although we possess

---

2 Wolf Biermann was a famous East German song-writer, whose nonconformist views alarmed the East German establishment already in the early 1960s. He was banned from performing several times and in December 1965 he was denounced as a "class traitor" and blacklisted his performances and publications. In 1976 Biermann was on an officically authored tour in West Germany as he was strip of his citizenship. (See https://wolf-biermann.de/vita.) Reactions on the Biermann expulsion are well documented in the Stasi's assessments of the situation, see Suckut 2009; Bispinck 2012.

3 Veen et al. 2000, 106; Scheer 1999, 201, 231.

4 Veen et al. 2000, 13.

a reasonably reliable historical understanding of the dissident movement scene in the GDR and – thanks to several well-prepared reference books and biographies – also know "who was who" in the GDR opposition movement, we lack a systematic understanding of the structures and dynamics of the social networks behind and underlying the East German opposition. This might be due to the fact that the mainstream of studies focusing on the GDR opposition has been interested in events, epochs, developments or single persons from a chronological, historical perspective, and to the best of my knowledge, no single study has thus far made an attempt to obtain *relational* data from the sources and study complex relational patterns – affiliations, interactions or collocations – "behind the scene".

The structure of this article is as follows. In the first section we present our primary material and discuss questions related to our method. The second section contains the main analysis, which builds on network visualisations. We end the article with concluding remarks in which we sum up the most important results and their implications.

2 Opposition, Resistance and State Repression in the GDR: A Historical Framework

One of the main tasks of the East German security apparatus was to underpin all attempts to build a system of opposition or resistance against the party dictatorship in the GDR. In accordance with this agenda, real or suspected members of the political opposition, dissidents and resistance groups were systematically haunted by the state security.[5] However, like in other socialist countries in Eastern Central Europe, also in East Germany the opposition was not a monolithic bloc, but instead, its internal structure and setting, goal-setting and means were under a continuous, dynamic evolution during the existence of the GDR from 1945/49 to 1989/90.[6] This evolution is, at least in retrospect, understandable, even natural, as it follows the general political development in the GDR. During the early years of the GDR (1949-1953) the main goal was a rapid reunification of Germany, supported by attempts to foster the creation of a pluralist party system and resistance against the "block policy" of the SED. Between 1953 and 1961 – the failed uprising in June 1953 and the construction of the Berlin wall in August 1961 – the history of the opposition is almost identical with the "republic flee" from the GDR in Western Germany. This phase

---

5 Weber 1999, 130.

6 Neubert (1998, 29-33) differentiates between three forms of political dissidence: opposition (Opposition), resistance (Widerstand) and political protest (politischer Widerspruch).

of "voting with the feet" marks a period of mass resistance against the SED regime, abruptly ended by the building of the Berlin Wall and followed by a period of economic and political stabilisation, which greatly undermined the role and status of the political resistance and opposition.[7]

The temporal focus of this article is the period between 1975 and 1989/90. This period was preceded by several political events shaping also the political space of the opposition and resistance groups. First of all, the *Conference on Security and Cooperation in Europe* (CSCE) held in Helsinki in August 1975 resulted in the second half of the 1970s in growing tensions within the Soviet empire, including the GDR. The East German party leadership was increasingly concerned with the destabilising impact of the CSCE on its power and sought to underpin all opposition activities drawing from the CSCE.[8] The numerous repressive actions by the state leadership against the dissidents and opposition groups – the most prominent example of this being the expulsion of Wolf Biermann in 1976 – evidence the growing fear among the party leadership of the destabilising effect of dissisident activities on the East German dictatorship. The same strategy – targeted actions and sanctions against a single visible member of the opposition – was also used in the 1980s against the peace movement that was questioning the GDR's self-image as a "peace state". Overall, the main aim of the repression was to scatter the resistance and opposition by eliminating their leading personalities.[9]

Considering the history of the East German opposition, Jena was, together with Berlin region, one of the most important regions what comes to the structure, means, motives and dynamics of the opposition groups in the GDR. During the whole history of the GDR, Jena was one of the cities where the discrepancy between democracy and dictatorship resulted over-proportionally often in open conflicts. As a result of both the "1968" in Western Europe and especially the "Prague spring", Jena rapidly became *the* region for political opposition in the GDR. It was also called the secret capital of the GDR opposition reflecting the complex domestic conflict between the state apparatus, church and opposition in the GDR.[10] For example, a political thesis paper published in December 1970 entitled "*Sozialismus in der DDR - Anspruch und Wirklichkeit*" (Socialism in the GDR - Ideal and Reality) triggered the Stasi's countermeasures

---

7 Veen et al. 2000, 8-9.

8 E.g. Schroeder 1998, 233ff.; Gieseke 2008.

9 Veen et al. 2000, 27-29.

10 See http://www.bstu.bund.de/DE/InDerRegion/Gera/Regionalgeschichten/Aktion-Gegenschlag/20130507_jena_aktion-gegenschlag-kowalczuk-vortrag.html [online. Last visited: 27 June 2018].

(operation "Anarchist") against the Jena opposition group around Jochen Anton Friedel and Reinhard Fuhrmann.[11]

The Jena Peace Community established after Matthias Domaschk's funeral was a dissident platform of short duration, but of long-lasting impact. Of short duration, because already in the spring of 1983 the security authorities were alarmed about the developments in Jena and decided to destroy the Jena Peace Community once and for all. The operation "*Gegenschlag*" (Counter-strike) in May 1983 was a targeted operation against the core group of the Jena. Circa 40 persons, Jahn included, were expelled, causing an almost complete destruction of one of the most active opposition groups in the GDR. Of long-lasting impact, because the counter-strike of the Stasi did not achieve its main goal, andbetween 1983 and 1989 Jena remained an unsettled city and one of the most important places for the political opposition.[12]

3 Method and Material

The structure and dynamics between the issues, events and people underlying historical phenomena are the key interest areas of historical research. Together with the expansion of digitised materials and advances in the availability of digital research tools have opened up totally new possibilities for the application of formal methods to historical empirical material in order to model and analyse patterns of social interaction underlying certain phenomena. Among these methods, historical network analysis (HNR) - rooted in social network analysis (SNA) - enjoys a growing popularity among historians and historically oriented social scientists interested in examining social or institutional relations or interactions of past communities in an aggregated historical context[13]:

With SNA, we are only interested in individuals as part of a much bigger whole. In fact, one advantage to the technique is that SNA helps us view an entire community and figure out which individuals we should be truly interested in and which ones were perhaps less significant. When we study past relationships systematically as SNA allows us to do, the method will prevent us from misunderstanding the function of an individual's relationships or exaggerating the distinctiveness of those relations.[14]

Although network research itself dates back to the late 19th century, it was the development of powerful personal computers and easy-to-use software allowing powerful graphical presentations, that is, visualizations of networks

---

11 Veen et al. 2000, 190-192.

12 Neubert 1998, 488. The operation "Counter-strike" is documented in: BStU 2013.

13 E.g. Bearman et al. 2012; Brughmans 2013; Epple 2017. For a more detailed discussion about historical network research, see http://historicalnetworkresearch.org/.

14 Morrissey 2015, 69–70.

that has increased the interest in network research among historians. In recent years, network visualizations have gained a central position in network research. This mostly, because network visualizations are an effective form of presenting complex relationships in an intuitive and (quite) easy-to-understand form. Additionally, different visualization layouts offer new possibilities for highlighting network-related attributes or visualising a node's relevance in regard to its close neighborhood or the complete network.[15]

This article exploits an experimental approach to explore geospatial social networks of the East German dissident community with data obtained from a digitised enriched photograph corpus. The primary source consist of a heterogeneous photograph corpus maintained by the Robert Havemann Society in Berlin as part of its archive of the East German Opposition and consisting of ca. 60.000 digitised photographs[16], exemplified in Figure 1. The digitised corpus has a relatively rich meta-data providing information about the date the photograph was taken and the photographer, a descriptive title, keywords (also including regional/geographical information), and information about the persons to which the photograph is related.

The primary material used in this article is a filtered dataset selected by applying two selection criteria. First, all photographs taken between 1975 and 1990 were selected. Second, based on previous studies we created a list of the most prominent figures of the East German dissident scene. We then used this list to filter the most frequently occurring persons in our database and to select the ten (10) most prominent persons: Matthias Domaschk, Jürgen Fuchs[17], Roland Jahn, Robert Havemann[18], Katja Havemann[19], Bettina Wegner[20], Carlo

15 For more, see e.g. Scott 2013; Schultz-Jones 2009; During and Stark 2011.

16 See further: https://www.havemann-gesellschaft.de/archiv-der-ddr-opposition/bildarchiv/ [online. Last visited: 10 October 2018].

17 Jürgen Fuchs studied in Jena, moved to East Berlin in 1975 and was expelled to West Berlin in 1977 where he engaged himself in the peace movement.

18 Robert Havemann (1910-1982) was a committed socialist, a former resistance fighter against the Nazi regime and - until 1964 - a professor at the Humbolt University in Berlin. By 1956 he was increasingly in opposition to the East German political leadership and lived out his last years in a house arrest.

19 Katja Havemann married Robert Havemann in 1974 and supported her husband in his dangerous work as a political writer. She lived in Berlin/Grüneheide and engaged herself in the opposition movement, especially in the Neues Forum, in the latter half of the 1980s.

20 Bettina Wegner was an East German song-writer and lyricist. In 1983 she was threatened with prison and forced to leave the GDR for West Berlin.

Jordan[21], Gerd Poppe[22], Bärbel Bohley[23], and Tom Sello[24]. The final primary dataset contains only records of photographs having at least one of the key figures mentioned in its meta-data.

Foto: Albrecht/Kleindienst, Quelle: Robert-Havemann-Gesellschaft
Jena, 19.5.1983, Jenaer Friedensgemeinschaft beteiligt sich mit eigenen Transparenten an offizieller Demonstration anläßlich des Pfingsttreffens der FDJ

Figure 1: Example photograph from the material corpus (source: *Bildarchiv Robert-Havemann-Gesellschaft*, signature RHG_Fo_HAB_11308).

---

21 Carlo Jordan joined the protest group against the expulsion of Wolf Biermann in 1976. Thereafter he was continuously on the radar of the Stasi and was forced to break up his academic studies due to political reasons. In the 1980s Jordan engaged himself in the environmental movement.

22 Gerd Poppe was a political activist who fought for human rights in the GDR. He was also actively engaged in the publication and dissemination of several illegal underground publication (Samisdat). Poppe was subject to the Stasi's intensive observation and repressive activities.

23 Bärbel Bohley was an East German opposition figure and artist. She was one of the co-founders of the Initiative for Peace and Human Rights (1985) and of Neues Forum (1989).

24 Tom Sello engaged himself in several dissident groups in the GDR, especially in the 1980s. He also wrote for several underground Samisdat-publications and was repeatedly attacked by the Stasi.

During the data extraction process some problems with the material already became evident. The biggest problem probably were the inconsistencies and pure typos in referencing the people in the pictures. Since the meta-data has been created afterwards, persons not accurately recognised were marked as "unknown". Other problem were misspelled or incomplete names. Our solution was to create a so-called alias table consisting of all unique person name entries from the database in one column and the correct form in another column. The original, uncorrected data had over 900 unique person references, in the cleaned data this number was slightly under 600, thus illustrating the bias caused by incorrect and misspelled person references.

Our cleaned and filtered database contained 841 records (photographs) with 171 unique person references. On average, each record had 2.5 person references within the range from 1 to 15. This database was then processed further to extract geographical information available either in the title or in the keywords field. First, we processed the keyword entries and collected all recognised geographical names (e.g. Berlin, Jena, Bad Frankenhausen). We also stored sub-regional information like street names or city districts. Second, all the records not containing identified geolocations were processed manually in order to find out whether the geolocation could be identified by indirect information like known places, buildings etc. In total, almost 3/4 of the photographs included in our analysis could be connected with a geographical location.

Once the data preparation process was finished, the data table was imported to the statistical package R for further processing. The network data creation was carried out as follows. First, we created a person-to-person link list by connecting each person mentioned in one record with all other persons mentioned in the same record. In other words, we treated each photograph as a fully connected small world network. We also attached the original record signature as well as the geographical (names of places) and temporal data (year) to each link list item. The geographical names were also used in order to obtain geocodes, that is coordinates (latitude and longitude), to be used in geospatial visualisations. The resulting dataset was used to reconstruct temporal social networks. Second, we created a geospatial network data where two geographical places were connected if the person had been photographed in both places. For example, if Roland Jahn was photographed in Berlin and Jena, these two cities were connected. At this point we also added temporal information to the link list data in order to be able to analyse network dynamics over time. All these steps were carried out with the R scripting language in the RStudio environment. The network data was then exported from R and

imported in Visone[25] for network visualisations and analysis. It should be noted here that Visone offers the possibility to use geocodes in order to visualise connections between geographical locations.

4 Analysis and Results

We have split our analysis into two main parts. We will first turn to geospatial networks presenting interregional connections and analyse how regions were connected through social mobility of the members of the East German opposition community, and second, exemplify the impact of social change on social networks by analysing Roland Jahn's geospatial social networks prior and after his expulsion in 1983.

Our analysis is based on graphical network visualisation. A (network) graph is, according to graph theory, a set of dots (nodes, vertices) and connecting lines (edges) between the nodes.[26] Both the nodes and the edges can be enriched by attaching both quantitative and qualitative attributes to them, and these attributes can be used in the analysis. As described in the previous section we have added geocodes and time related attributes to our edge data, whereas personal names and city names are used as node attributes. The main purpose of using attributes, however, is to ensure that the formal network model, the network topology, describes its real-world counterpart as exactly as possible and, thus, provides valid, reliable and appropriate information concerning the research question. Network analysis is based on the assumption that both the units (nodes) selected for the analysis and the connections (edges) between these units are significant when it comes to understanding and explaining the phenomenon the network is connected to. From this perspective, network analysis is expected to provide quantitative information about dependencies between variables in the data, thus improving our understanding of the phenomenon subject to analysis.[27]

4.1 Regional Connectedness

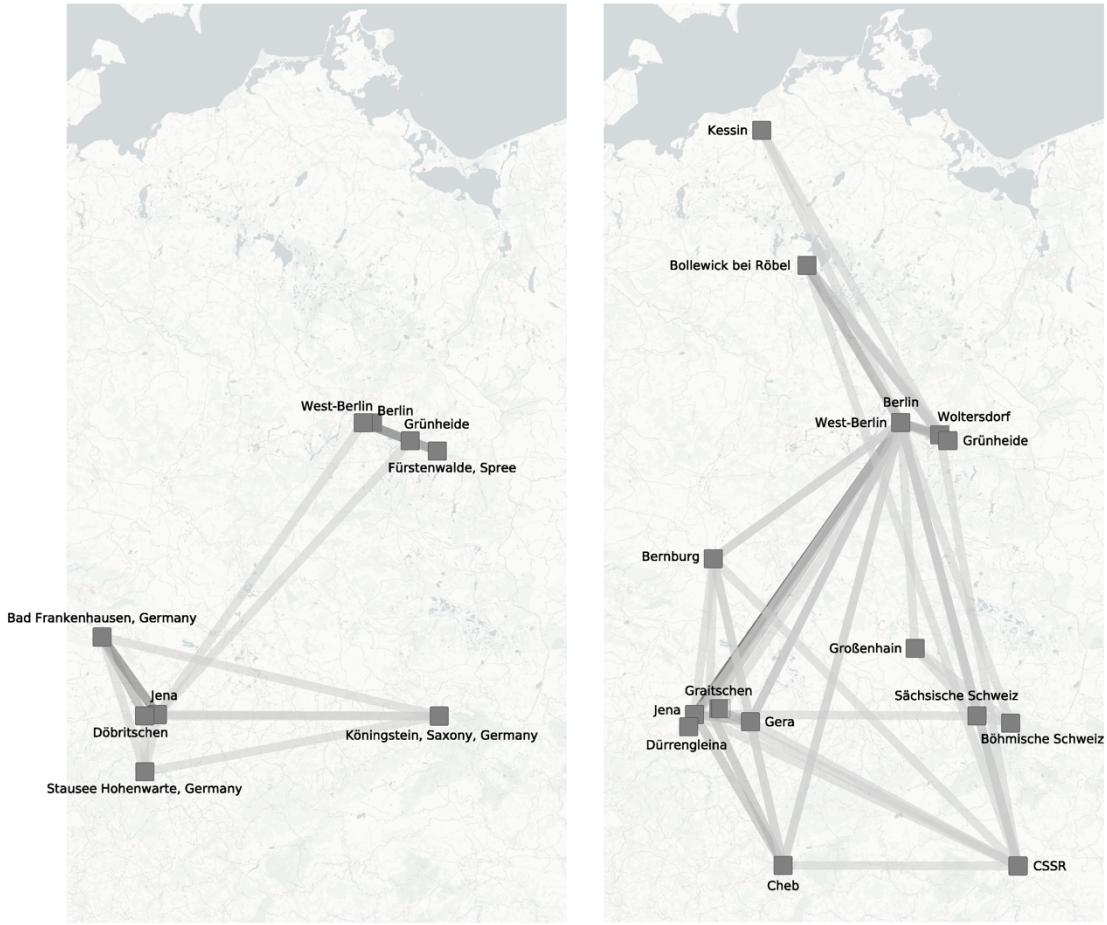
Figure 2 tackles the question of geospatial dynamics of the East German opposition by visualising interregional connections based on geographical references in the database andusing geocodes to position the nodes in the networks. The graph is made with Visone, which allows the use of coordinate data and online map resources (OSM database) in order to generate the background map.

---

25 Visone is an powerful, yet easy to use open-source software for network analysis and visualisation. Homepage: https://visone.org

26 For a good introduction to graph theory, see Ruohonen 2013.

27 Morrissey 2015-



Figures 2a and 2b: Person to person via location networks: (a, left) 1975-1979, (b, right) 1980-1984

Each map covers the same geographical area (11.0-15.0 longitude, 50.0-54.5 latitude), but the networks present the geographical connections in three, equally-sized time widows: 1) 1975-1979, 2) 1980-1984, and 3) 1985-1990. The link colour is based on interpolation from light grey to black and is proportional to the number of geographical co-occurrences, thus helping us to identify the most important geographical connections.

We consider the fact that the resulting time frames are rather different what comes to the number of photographs (1975-79: 383, 1980-84: 604, 1985-1990: 1157) quite natural and unproblematic when put in the historical context. Oppositional activities in the GDR increased during the whole of the 1980s, but

the trend was extremely strong during the latter half of the 1980s. Our figures support this general understanding, as the number of photographs is roughly double of that of the first half on the 1980s.

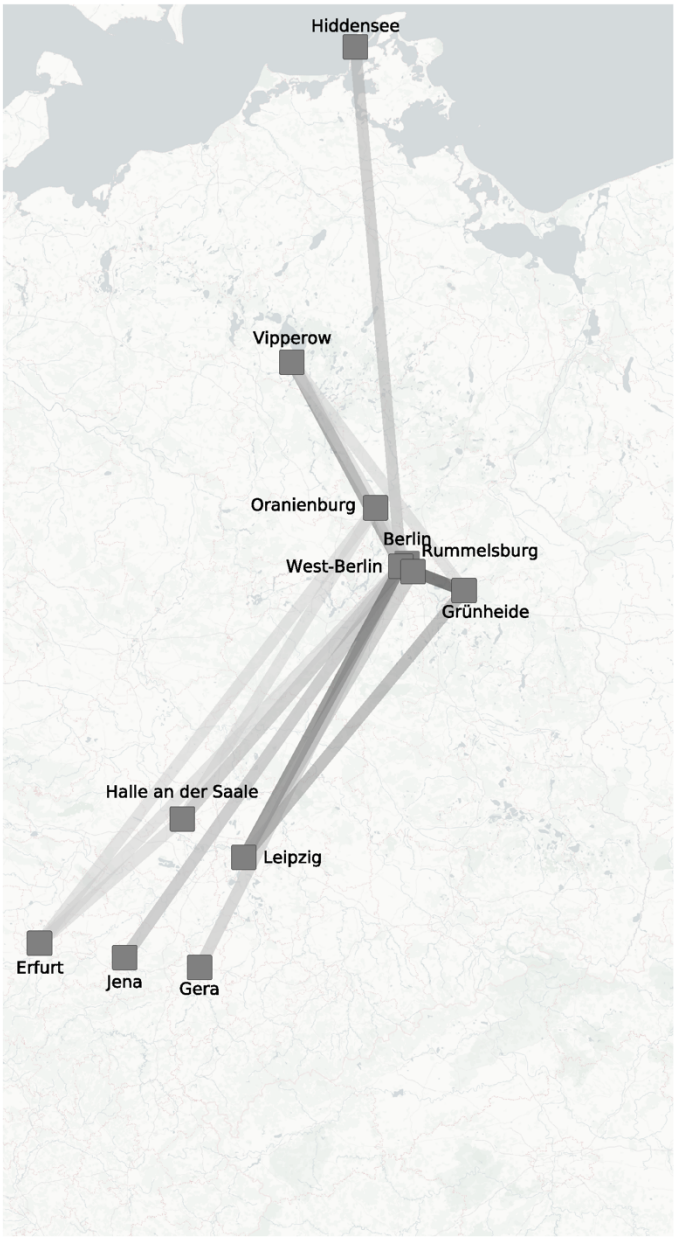


Figure 2c: Person to person via location networks: (c) 1985-1990.

As regard to the regional connectedness, the clearest change over time occurs between Jena and Berlin. Until 1984, Jena remains the most important region, thus confirming previous studies stressing the importance of the Jena region for the East German dissident community. From 1985 onwards, Berlin gains in importance and becomes the most frequently referenced region in our data. This development is well in line with the overall course of events during the second half of the 1980s. Also, many of the photographs documenting the late 1980s have references to the *Umweltbibliothek* in East Berlin. The *Umweltbibliothek* was founded in 1986 in the cellar rooms of the *Zionsgemeinde* in East Berlin and it rapidly became one of the central communities of the East German dissident movement.

Another aspect worth mentioning is the regional mobility among the dissidents. As the geospatial visualisations evidence, in the first period (1975-1979) main activities seem to have revolved around the Jena region with some "excursions" to Berlin. The towns Fürstenwalde (Spree) and Grüneheide were closely connected to Robert Havemann, who was sentenced to house arrest in 1976. Most of the photographs referring to Havemann document his life under the house arrest and people visiting him.

Compared to the late 1970s, the regional mobility of the East German dissidents seems to have increased quite a lot during the first half of the 1980s, although Jena and Berlin still remained the most important regions. The regional mobility in the last period (1985-1990) is well in line with the overall developments in the GDR. For example, the rise of Leipzig to one of the key regions evidences not only the establishment of new opposition groups in other regions of the GDR, but also their emancipation from the evangelical church.[28] One important factor explaining the regional mobility documented in our database is the establishment of new political platforms and movements like *Neues Forum*, *Demokratischer Aufbruch* or *Demokratie jetzt*, offering an interregional, GDR-wide platform for the leading personalities of the new East German opposition movement.[29]

Overall, our regional connectedness analysis clearly shows how the political erosion of the GDR increased the room of manoeuvring of different, independent opposition groups. The regional hotspots dominating our database are well-known in the literature, but our analysis evidenced that same persons were actively engaged in different regions. This, in turn, supports the argument

---

28 Dietrich and Jander 1999.

29 Knabe 1999.

that the actual "hard core" of the East German opposition was rather small in number.[30]

4.2 Dynamics of Geospatial Social Networks: the Case of Roland Jahn

In order to exemplify the geospatial dynamics of the social networks related to the East German opposition, we decided to take a closer look at how the social networks of Roland Jahn, a prominent figure in Jena, changed over time. Roland Jahn was a close friend of Matthias Domaschk and an active member in the Jena dissident community. Jahn himself had been on the Stasi's radar already since the mid-1970s when he engaged himself as a young university student in protest actions against the expulsion of Wolf Biermann. As a result, Jahn was ex-matriculated from the University of Jena in 1977.[31] In 1980 Jahn got involved with the Polish "Solidarność" movement and was arrested several times by the security authorities. Later, in 1982 Jahn was condemned to 18 months' imprisonment, but was freed in February 1983 thanks to international protests against his imprisonment. Shortly thereafter, in May 1983, Jahn was among those 40 persons who were expelled to West Germany in the Stasi's operation "*Gegenschlag*". However, also after his expulsion Jahn continued to support the East German opposition movement.

Against this background, we assumed it to be possible to identify a structural change in Jahn's social networks before and after his expulsion both in regard to personal connections and to geospatial structure. Our assumption leaned on previous studies that have shown that the actual "hard core" of the Jena community was rather small in number.[32]

Figure 3 visualises the whole social network we could reconstruct from our database for the time period 1975-1990 and is the person-to-person network data created as described in section 3. The resulting graph is quite dense, but it does appear to have a relatively clear structure. There are a total of 315 nodes representing persons and 1510 edges representing co-occurrences of different persons.[33]

---

30 Neubert 1998; Veen et al. 2000.

31 On resistance and repression at the University of Jena, see Fritsch & Nöckel 2006 and Lenski 2017.

32 Scheer 1999.

33 Thanks to Visone's capabilities to handle multiple edges between the same two nodes the Figure 3 correctly visualises also connections between two persons having been photographed together in two different geographical location. In such a case there exists multiple connections between the two persons, each of them representing a unique location.

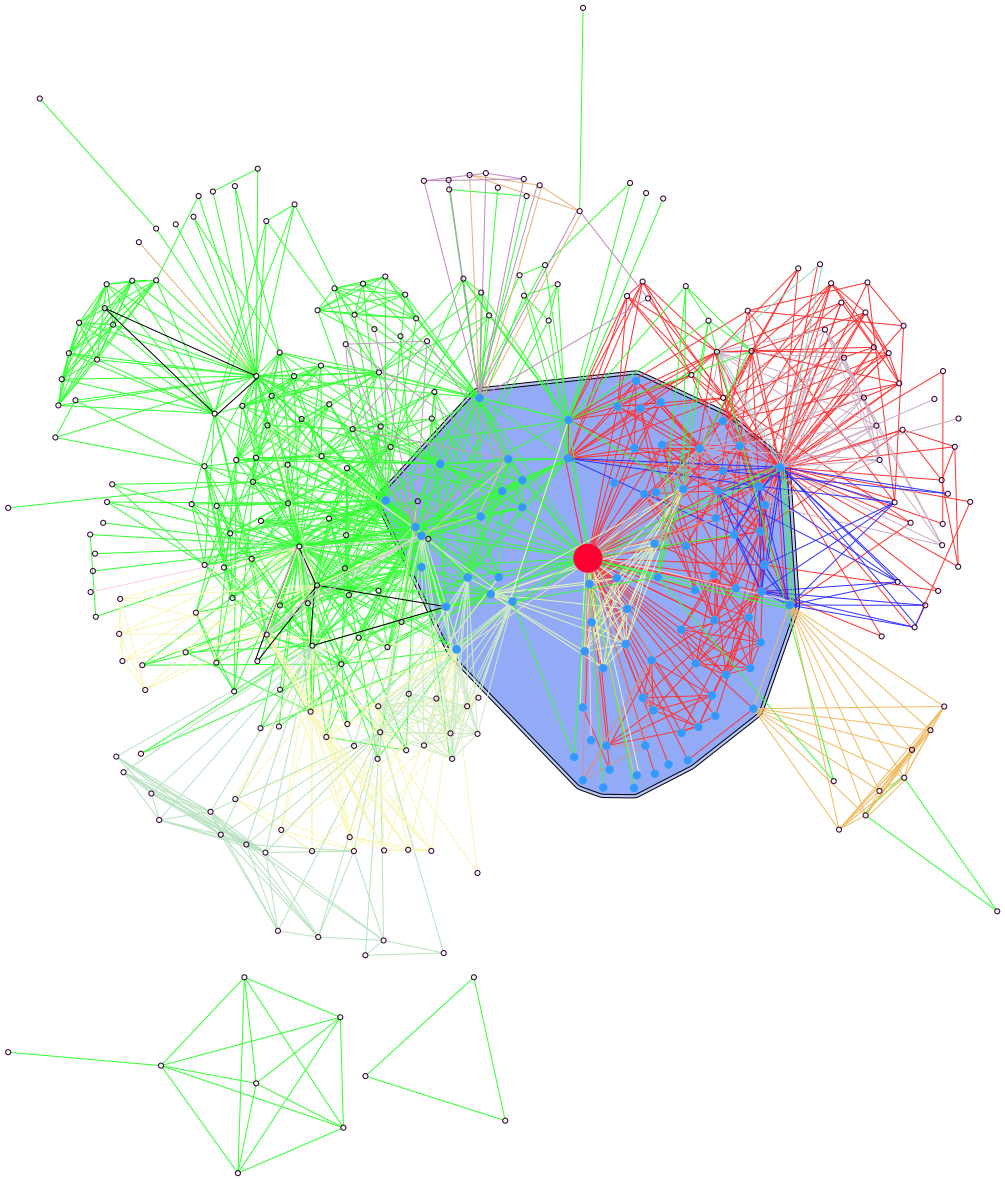


Figure 3: Roland Jahn's social network within the whole network structure between 1975-1990. Red node: Roland Jahn. Blue nodes: Jahn's direct neighbours. Layout: stress minimisation. Red edges = JENA, Green edges = BERLIN AREA, other colours = diverse geographical connections.

We have applied stress minimisation layout in ordeg to visualise sub-communities in the graph structure and to highlight the most important nodes by collecting them tn the centre of the graph. Further, we have added some visual effects helpful for understanding Roland Jahn's position and status in the network structure. Jahn's node is marked with red and members of his ego

network are marked wit blue, whereas nodes and edges not belonging to Jahn's ego network are greyed out by adding transparency. Edges coloured in red are connections located in Jena, whereas Berlin-based connections are coloured in green

Jahn's position in the opposition network an be described as that of a mediator, a position which also strongly related to his expulsion. He stays between the two main clusters in the graph and has a dense network of ego connections in both clusters. Jahn's ego network consists of 84 persons, a relatively high number evidencing his broad activities within this community. A closer look into the underlying network data reveals, that Jahn's Berlin connections (coloured in green in Figure 3) were mainly established after his expulsion.

In order to analyse the impact of Jahn's expulsion in 1983 on his social networks we divided Jahn's networks into two temporal networks, one reconstructing his social network prior to, and another after his expulsion. Our underlying assumption was that Jahn's expulsion marked an interruption, a break even in his oppositional activities. Based on this assumption we expected Jahn's social networks to reflect this interruption both in regard to its geospatial and to its co-occurrence structure.

The results of this zooming in are visualised in Figure 4. Since Jahn's ego network is close to what network analysis knows as small world, we have applied a specific layout called "quadrilateral Simmelian backbone layout". This layout tries to tackle the so-called hairball problem caused when force-directed layouts are applied to small-world graphs with low average distance and high density or conductance.[34] As we can see, the resulting visualisation is not just morr readable, but it also allows us to identify nodes deeply embedded within the community without losing the general understanding of the connectedness of the whole network. In other words, the layout helps us to untangle hairballs created by force-directed layouts.

Once again, extra visual effects are used to highlight elements important for our analysis and interpretation. An edge between two nodes indicates that the two persons represented by the connected nodes are referenced to in the same photograph. Additionally, each edge has an attribute containing the geographical information about the connection. Thus, each connection is a geographical linkage as well, making the graph to visualise geospatially grounded social networks.

---

[34] See further Nocaj et al. 2015.

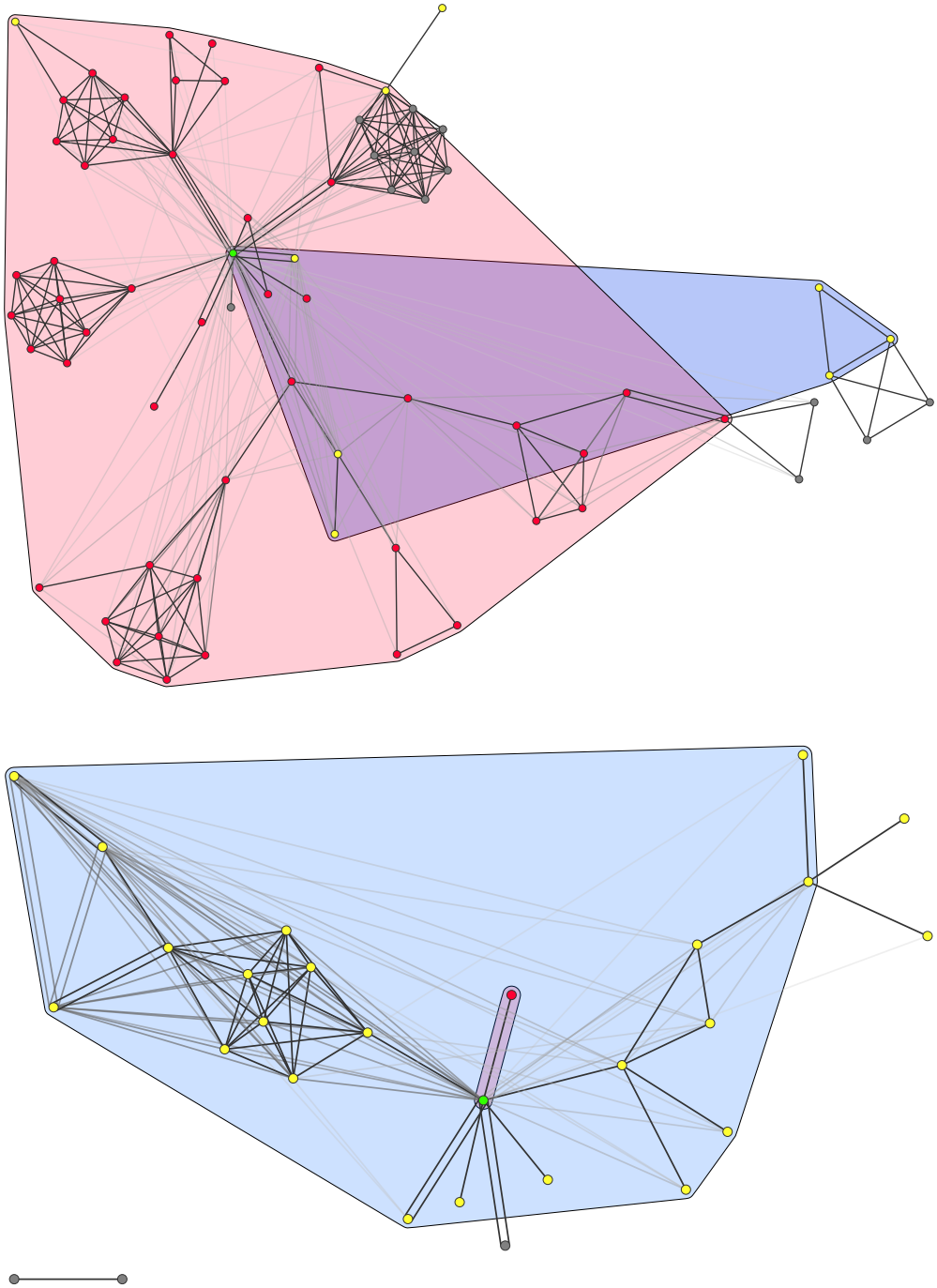


Figure 4: Roland Jahn's geospatial social network (a, top) prior to (1975-1983) and (b, bottom) after 1984-1990. Berlin connections in yellow, Jena connections in red. Layout: backbone (quadrilateral Simmelian).

eISSN 2535-8863
DOI: 10.25517/jhnr.v2i1.45

Both visualisations use the same colouring schema. Roland Jahn's node is coloured in green, connections re-constructed by photographs taken in Berlin (both West and East) are coloured in yellow, and connections re-constructed by photographs taken in Jena are coloured in red. In addition, we have created group nodes for both main regions, Berlin and Jena, in order to better highlight the geospatial social connections. These group nodes are visible as partly overlapping areas, coloured in red (Jena) and blue (Berlin).

The visualisations give support to the assumption that Jahn's expulsion resulted in significant changes in his social networks. Considering, first, the geographical dimension, we can see a clear shift from the Jena region to Berlin. Photographs taken prior to Jahn's expulsion are dominantly taken in the Jena region, whereas a majority of the photographs taken after his expulsion have been taken in the Berlin region.

A similar significant change occurred in Jahn's social connectedness. In fact, only 11 out of 69 persons from Jahn's ego network prior to his expulsion can be found in his social network after 1983. These 11 persons are visualised as larger white nodes in Figure 5. This is an interesting finding, since Jahn illegally visited the GDR also after his expulsion. One possible interpretation is, that Jahn tried to protect his former collaborators by not showing up with them in situations where the risk of being photographed was real. This interpretation gain support from previous studies focusing on opposition movements in other dictatorships[35].

But there is also an alternative (or better: complementary) explanation for this minimal overlap between the two networks. Many of the names connected to Jahn's post-expulsion photographs are not directly linked with the Jena region. Instead, these findings provide support for previous studies which have shown that after his expulsion Jahn became a key mediator between the East German opposition movement and Western politicians and journalists.[36]A closer look at selected names of these 11 persons gives support for both interpretations: Peter Rösch was one of Jahn's collaborators in the hard core of the Jena opposition. Bodo Sturhann, Gerold Hildebrand, Ulrike and Gerd Poppe, Reinhard Weißhuhn, and Carsten Hahn, in turn, were activist in different opposition groups especially in the Berlin region.[37] However, all of them started their engagement in different dissident communities already in the 1970s or the early 1980s and were involved in the same activities – such as

---

35 E.g. Düring and Beer 2011.

36 See especially Veen et al. 2000, 189-190; Scheer 1999; Praschl 2011; http://www.bstu.bund.de/DE/BundesbeauftragterUndBehoerde/Bundesbeauftragter/_node.html [online. Last visited: 10 October 2018].

37 Neubert 1998, 432.

protests against Biermann's expulsion – as Jahn. Hence, Jahn seems to have maintained and exploited these old connections also after his expulsion. Jürgen Fuchs, in turn, was expelled to West Berlin already in 1977, but he remained involved in the East German dissident community until 1990, especially via Lutz Rathenow, who was also continuously present in Jahn's ego networks. Another important linkage between West and East was Petra Kelly, a Green activist involved in the peace movement in the 1980s.

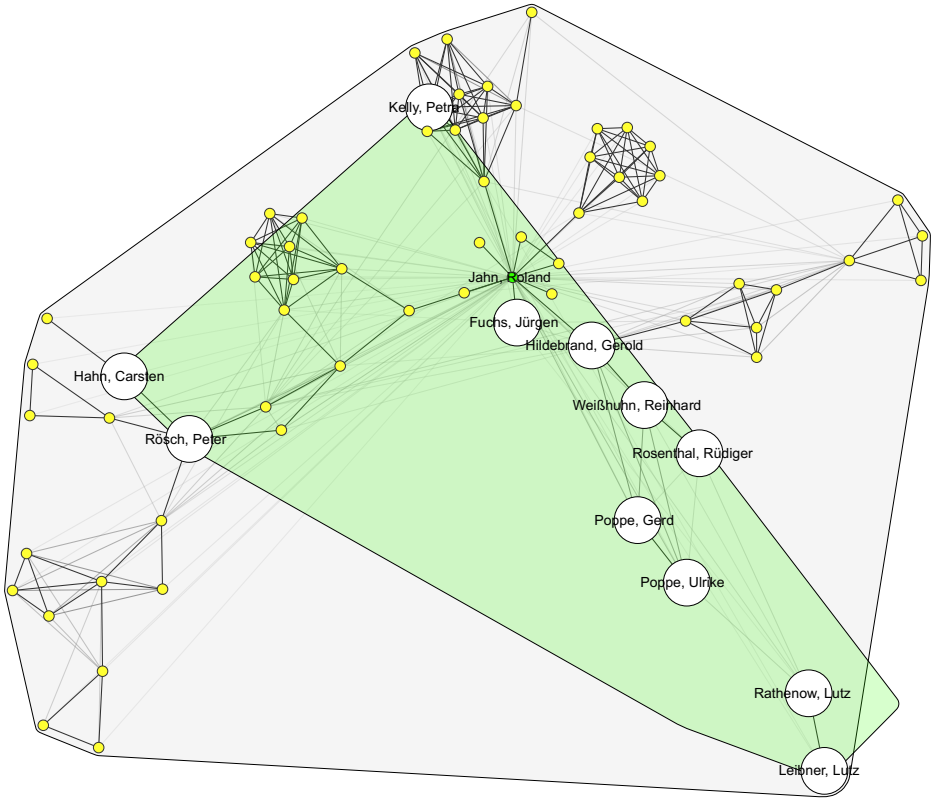


Figure 5: Change in Roland Jahn's geospatial social network after his expulsion from the GDR. Layout: backbone (quadrilateral Simmelian).

5 Discussion

This article sought to exemplify how a digitised photograph corpus can be exploited for network analysis focusing on the structure and dynamics of geospatial social networks. Our analysis shows how the regional connectedness based on personal mobility among the East German dissidents both changes and increases over time, thus resulting in continuously evolving patterns of social interaction. Further, the analysis of Roland Jahn's geospatial networks evidences the usefulness and power of historical network analysis when it comes to

tackling changes in patterns of social interaction. In all communist countries, the GDR included, dissident movements were torn between the search forpublicity and visibility, and permanent repression. The numerous expulsions of leading persons of the East German opposition movement underline the importance of social networks, also in the eyes of the Stasi. Consequently, the GDR opposition did its best to protect its social networks, while at the same time documenting its political actions and supporters by photographs. Photographs were also used as illustrations in underground magazines like "die andere" or "Umweltblätter".

Focusing on one crucial tipping point of the Jena opposition, the years 1982-1983, we followed the assumption that major changes in a person's situation would result in changes in geospatial social networks as well. Although this might sound self-evident, but the opposite direction provides an interesting methodological conclusion for exploratory research: exploring major changes in the social network structures could help us to identify turning points thus far lacking a plausible explanation or having escaped scholars attention.

There are two important, theoretical aspects relevant for understanding social cohesion and closeness among the East German dissident community. First, as already pointed out, previous studies have shown that the East German dissident movement revolved around a rather small number of core members. Our empirical analysis of Roland Jahn's social networks seems to confirm this. Speaking more generally, dissident communities can be assumed to be, at least to a great extent, emotional communities bringing similarly minded people together. And as previous research has evidenced, smaller networks tend to produce a higher level of emotional closeness.[38]

Second, geographical closeness – in this article in the form of regional connectedness – seems to support social closeness by strengthening other existing ties between the individuals.[39] Considering the space–time dimension, we can conclude that emotional and geographical closeness reduced the time needed for communication. Hence, mobile dissidents helped to overcome – or at least to reduce – the geographical distance by acting as transmission lines between regions. This rather theoretical consideration on the importance of geographical closeness gains support from Figure 3, where Roland Jahn acts as mediator connecting geographically rather distant regions. Naturalle, Jahn was not the only mobile dissident, but his case well illustrates the point. According to the structure of the network, Jahn seem to have had contact to regional "hubs", that is the core members in regional opposition communities. These

---

38 Roberts and Dunbar 2011.

39 Yin and Shaw 2015.

regional hubs had, in turn, their own local social networks (visible in Figure 3 as nodes connected to nodes in Jahn's ego network but not to Jahn himself), resulting in a sub-network with an inner and outer layer, where the personal *non-tie* between a person in the outer layer and the ego is just as important as a tie itself.[40] It seems that these geographical links are "weak links" dividing the network structure into an "inner" and "outer" layer, whereas the ego-related links are "strong links" indicating also a stronger emotional closeness, in our case caused by the engagement in the opposition movement.[41] Our analysis also seems to confirm the finding of Roberts and Dunbar (2011) that the importance of "weak ties" lies in their function to provide "access to a greater variety of information, ideas, and experience than the stronger ties at the inner layers of the network".

Despite the rather promising analytical results discussed above, the analysis presented in this article has certain limitations the reader should be aware of. First, like any analysis, also our analysis is highly dependent on the data used. Although the photograph corpus used in this paper is quite large, neither the corpus itself nor its meta-data can be considered perfect or complete. Hence, there is a modest risk that the data itself can cause bias and, thus, compromise the results. We have tried to avoid over-interpretations, as well as under-interpretations, and sought to reflect our findings against previous studies in order to critically consider the relevance, validity and reliability of the results.

Second, this paper's focus has been on geospatial social networks obtained from photographs and, thus, can only be used to explore and analyse personal and regional connections and ties among a certain part of the East German dissident movement. One positive result is, that the networks re-constructed and analysed in this article fit quite well with the historical facts about the East German opposition. Another positive result is, that theoretical considerations on the role and impact of geographical and emotional closeness seem to explain – at least to a great extent – our findings. However, we cannot estimate how much data is missing or whether a bigger sample would have affected our results differently. Network analysis is known to be quite sensitive to missing data, a challenge relevant to all historians interested in applying network analysis. A proper understanding of the corpus serving as the primary source is indispensable, as well as understanding that historical network analysis is a research method, not an explanatory theory, and the results are always subject to context-related interpretation.

---

40 Morrissey 2015, 72.

41 For "weak" and "strong" ties, see especially Granovetter 1973.

References

Bearman, P. S., Moody, J., and Faris, R. (2012). Networks and History. *Complexity*, 8: 61–71.

Bispinck, H., editor (2012). *Die DDR im Blick der Stasi 1977: Die geheimen Berichte an die SED-Führung*. Vandenhoeck & Ruprecht, Göttingen.

Brughmans, T. (2013). Thinking through networks: A review of formal network methods in archaeology. *Journal of Archaeological Method and Theory*, 20(4):623–662.

BStU (2013). *Aktion "Gegenschlag": Die Zerschlagung der Jenaer Opposition 1983*. BStU, Gera.

Dietrich, C. and Jander, M. (1999). Die Ausweitung zum Massenprotest in Sachsen und Thüringen. In Kuhrt, E., Buck, H. F., and Holzweißig, G., editors, *Opposition in der DDR von den 70er Jahren bis zum Zusammenbruch der SED-Herrschaft*, Am Ende des realen Sozialismus, Bd. 3, pages 737–786. Leske+Budrich, Opladen.

During, M. and Stark, M. (2011). Historical Network Analysis. In *Encyclopedia of Social Networks*, Sage Publishing, London.

Düring, M. and Beer, S. (2011). Hilfe für jüdische Verfolgte im Nationalsozialismus. *Medaon - Magazin für jüdisches Leben in Forschung und Bildung*, 9.

Epple, R. (2017). Flucht und Unterstützung: Die Auskunftsstelle für Flüchtlinge in Zürich und ihre Netzwerke. *Journal of Historical Network Research*, 1(1):106–145.

Fritsch, W. & Nöckel, W. (2006). *Opposition und Repression an der Universität Jena (1956-1968). Eine Dokumentation*. Verlag Jena 1800, Jena.

Gieseke, J. (2008). East German Espionage in the Era of Detente. *Journal of Strategic Studies*, 31(3):395–424.

Granovetter, M. (1973). The Strength of Weak Ties. *American Journal of Sociology*, 78(6):1360–1380.

Knabe, H. (1999). Der lange Weg zur Opposition – unabhängige politische Bestrebungen 1983 bis 1988. In Kuhrt, E., Buck, H. F., and Holzweißig, G., editors, *Opposition in der DDR von den 70er Jahren bis zum Zusammenbruch der SED-Herrschaft*, Am Ende des realen Sozialismus, Bd. 3, pages 139–170. Leske+Budrich, Opladen.

Lenski, K. (2017). *Die Staatssicherheit an der Friedrich-Schiller-Universität Jena*. Campus, Frankfurt/Main.

Morrissey, R. M. (2015). Archives of connection. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 48(2):67–79.

Neubert, E. (1998). *Geschichte der Opposition in der DDR 1949-1989*. Ch. Links, Berlin.

Nocaj, A., Ortmann, M., and Brandes, U. (2015). Untangling the hairballs of multi-centered, small-world online social media networks. *Journal of Graph Algorithms and Applications*, 19(2):595–618.

Praschl, G. (2011). *Roland Jahn. Ein Rebell als Behördenchef*. Ch. Links, Berlin.

Roberts, S. G. B. and Dunbar, R. I. M. (2011). Communication in social networks: Effects of kinship, network size, and emotional closeness. *Personal Relationships*, 18(3):439 – 452.

Ruohonen, K. (2013). *Graph theory*. Technical report, Tampere University of Technology.

Scheer, U. (1999). *Vision und Wirklichkeit: die Opposition in Jena in den siebziger und achtziger Jahren*. Ch. Links, Berlin.

Schroeder, K. (1998). *Der SED-Staat. Geschichte und Strukturen der DDR*. Bayerische Landeszentrale für politische Bildungsarbeit, München.

Schultz-Jones, B. (2009). Examining information behavior through social networks: An interdisciplinary review. *Journal of Documentation*, 66(4):592–631.

Scott, J. P. (2013). *Social Network Analysis*. Sage Publishing, London, 3 edition.

Suckut, S., editor (2009). *Die DDR im Blick der Stasi 1976: Die geheimen Berichte an die SED-Führung*. Vandenhoeck & Ruprecht, Göttingen.

Veen, H.-J., Eisenfeld, P., Kloth, M., Knabe, H., Maser, P., Neubert, E., and Wilke, M., editors (2000). *Lexikon Opposition und Wiederstand in der SED-Diktatur*. Propyläen, Berlin.

Weber, H. (1999). *Geschichte der DDR*. dtv, München.

Yin, L. and Shaw, S.-L. (2015). Exploring space–time paths in physical and social closeness spaces: a space–time gis approach. *International Journal of Geographical Information Science*, 29(5):742–761.