The Journal of
HISTORICAL
NETWORK
RESEARCH

9 | 2023
special issue

Networks of Manuscripts, Networks of Texts

EVINA STEIN | GUSTAVO FERNÁNDEZ RIVA

C²DH
LUXEMBOURG CENTRE FOR
CONTEMPORARY AND DIGITAL HISTORY

uni.lu
UNIVERSITY OF
LUXEMBOURG

Imprint

Université du Luxembourg 2023

Luxembourg Centre for Contemporary and Digital History (C²DH)
Université du Luxembourg
Belval Campus
Maison des Sciences Humaines
11, Porte des Sciences
L-4366 Esch-sur-Alzette

Editors
Asst. Prof. Dr. Marten Düring (Luxembourg Centre for Contemporary and Digital History | C²DH)
PD Dr. Christian Rollinger (Universität Trier)
Dr. Cinderella Petz (IEG – Leibniz-Institut für Europäische Geschichte)
Dr. Ingeborg van Vugt (Königlich Niederländische Akademie der Wissenschaften)
Clemens Beck, M. A. (Friedrich-Schiller-Universität Jena)

ISSN 2535-8863

Contact
Principal Contact
JHNR-editors@historicalnetworkresearch.org
Support Contact
Dr. Marten Düring (Université du Luxembourg)
JHNR-support@historicalnetworkresearch.org

Cover Design and Typesetting
text plus form, Dresden, Germany
Cover image
Evina Stein
Copyediting
Andy Redwood, Barcelona, Spain

Published online at
https://doi.org/10.25517/jhnr.v9i1

The Journal of
HISTORICAL
NETWORK
RESEARCH

TABLE OF CONTENTS

C²DH
LUXEMBOURG CENTRE FOR
CONTEMPORARY AND DIGITAL HISTORY

uni.lu
UNIVERSITY OF
LUXEMBOURG

The Journal of HISTORICAL NETWORK RESEARCH

EVINA STEIN/GUSTAVO FERNÁNDEZ RIVA

Introduction: Fitting Manuscript Studies into the Historical Network Research

Journal of Historical Network Research 9 (2023) iii–xvi

Keywords Manuscripts, Network Analysis, Medieval Studies, Middle Ages

C²DH
LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

UNIVERSITY OF LUXEMBOURG

Introduction: Fitting Manuscript Studies into the Historical Network Research

In September 2023, when this special issue will be published, the application of network analysis and network visualization techniques to traditional historical topics can still be considered an emerging trend within the methodological landscape of humanities research.[1] While studies that make use of social network analysis have made sporadic appearances in historical journals since the 1970s,[2] the awareness of the network methodology among scholars engaged in the study of the past has only reached a critical mass within the last two decades with the advent of new computational tools. A testament to this development is the establishment of the *Historical Network Research Community* in 2009[3] and the *Social Network Researchers of the Middle Ages* (SNARMA) in 2018,[4] the organization of (bi)annual conferences dedicated to historical network research since 2013, and the appearance of this journal, the *Journal of Historical Network Research* (JHNR), in 2017.

Most historical research featured by the aforementioned venues and in collected volumes dedicated to historical network research belongs under the banner of social network analysis (SNA).[5] Indeed, social historians should be credited as the first among history practitioners to have taken an interest in network theory, using the methods devised by sociologists and social anthropologists for the examination of contemporary human relations and interactions and

Acknowledgements: We would like to use this introduction to thank all parties who contributed to making this special issue possible. First and foremost, the authors of the papers and the editors of the *Journal of Historical Network Research,* who agreed to publish it and assisted us throughout the whole process. We would also like to thank all the participants of the *Networks of Manuscripts, Networks of Texts* conference for their insightful presentations and interesting discussions. A special mention is due to Matteo Valleriani, the keynote speaker at the conference. We would also like to thank the anonymous peer reviewers of the papers in this issue, as it is their work that guarantees scientific standards and massively improves the contributions. Finally, we would like to acknowledge the infrastructural support provided for the *Networks of Manuscripts, Networks of Texts* conference by the Huygens Institute.
Corresponding author: Gustavo Fernández Riva, University of Heidelberg, gustavo.fernandez.riva@ub.uni-heidelberg.de

1 A general overview of developments in the field of historical network research in recent decades can be found in Rehbein, "Historical Network Research, Digital History, and Digital Humanities"; Fazioli, "Modeling the Middle Ages."

2 See for example Reinhard, *Freunde und Kreaturen;* Smith, "Kin and Neighbors in a Thirteenth-Century Suffolk Community."

3 At: https://historicalnetworkresearch.org/.

4 At: https://medievalsna.com/.

5 See for example Gamper, Reschke, and Düring, *Knoten und Kanten III;* Brughmans, Collar, and Coward, The Connected Past; Kerschbaumer et al., *The Power of Networks.*

transplanting them to historical subjects.[6] However, SNA is not the only network-based methodology relevant for historical research. Scholars other than historians have also been independently developing, and at times unwittingly experimenting, with network visualization and analysis that does not concern direct human interaction and therefore cannot be properly called social. A notable example is provided by archaeologists applying network analysis to man-made artifacts.[7] Another group of history practitioners who have made an incursion into the methodological terrain of network theory is the scholars of texts and their written media who are interested in their production, transmission, and reception.

These non-social types of historical network research have not yet seen the same degree of codification as historical SNA. Until now, we have lacked both publications specifically dedicated to this type of historical network research and the degree of theorization that would put it on a similarly firm footing within the field of historical network research as SNA.[8] This special issue of the JHNR aims to alleviate this situation. It brings together, for the first time, network-based research in manuscript studies. While one or two of the contributions in this issue have a social aspect, most represent examples of historical network research that do not qualify as SNA. Our hope as editors of this issue is twofold. First, we would like it to stimulate interest in (non-social) network-based approaches within manuscript studies by showcasing what has already been undertaken. Second, we would like it to serve as a basis for the discussion of the utility and viability of network-based research as a methodological framework applied to texts and text-bearing objects, especially beyond the confines of SNA. Perhaps the most pressing methodological and theoretical question underpinning the latter objective is to what extent can we, scholars of the text and text-bearing objects, borrow the concepts and models developed in the context of SNA, as our research subjects are not humans and therefore cannot be considered to have friends, families, or business associates. Can we say that texts form cliques, such as when they appear in a historical book in large numbers, as one researcher has?[9] And if we cannot, is it a matter of terminology, or is it because the concepts and models of SNA are in some way fundamentally unsuitable to the needs of manuscript and text specialists? This issue cannot and will not provide answers to these essential methodological and theoretical questions, but we as editors hope to at least open the

6 On the adoption of SNA by historians, see, among others, Wetherell, "Historical Social Network Analysis."

7 See the overviews in Knappett, "Networks in Archaeology: Between Scientific Method and Humanistic Metaphor;" Brughmans and Peeples, "Trends in Archaeological Network Research: A Bibliometric Analysis."

8 To our knowledge, the only attempt at theorization of non-SNA network model in the context of the historical network research has been undertaken in de Valeriola, "Can Historians Trust Centrality?"

9 de Valeriola, "Can Historians Trust Centrality?", 95.

door to further discussion with the scholarly community interested in historical network research.

The impetus to produce this special issue stemmed from the *Networks of Manuscripts, Networks of Texts* international conference taking place virtually at the height of the COVID pandemic on 21–23 October 2020. The conference, organized in the context of Evina Stein's *Innovating Knowledge* project,[10] was conceived as an opportunity to bring together manuscript and early print specialists who had either already tried their hand at network analysis or were planning to incorporate it into their ongoing research. It was hoped that the conference would serve as a space where researchers from different disciplinary backgrounds could benefit from each other's research successes and failures, and where a productive cross-fertilization could be achieved by uniting disparate strands of research that had been carried out without previous mutual awareness. The significant interest in the conference, which featured sixteen presentations that covered topics as varied as Old Norse sagas, early modern Kabbalistic texts, public inscriptions from the Tang Dynasty China, and medieval Latin sermon collections, made it clear that there is a community interested in deploying network-based methods on written artifacts and a need to cultivate such methods within the historical network research framework.[11]

Historical network research within manuscript and textual studies carried out prior to the publication of this special issue belongs to several distinct strands. The earliest attempts to deploy network-based methods on texts and text-bearing objects have stemmed from scholars' interest in understanding the transmission of pre-modern texts. The interest in multitext books and their role in textual transmission is not new and the issue has been studied from different perspectives, but the implementation of network analysis presents a brand-new perspective developed in the last decade, as several scholars, unbeknownst to each other, experimented with visualizing and modelling the relationship between manuscripts transmitting the same texts as networks.

---

10 See the website of the project at: https://innovatingknowledge.nl/. The funding for the conference was provided by the Dutch Research Organization (NWO) as a part of the VENI grant 275-50-016.

11 The conference program and presentation abstracts are available online at: https://homomodernus.net/2020/08/26/conference-programme-networks-of-manuscripts-networks-of-texts/ (accessed July 14, 2023). A report with short summaries of all presentations has already been published in Fernández Riva, "Conference Report: Networks of Manuscripts, Networks of Texts. Amsterdam (Online), 21–23 October 2020." Some of the conference presentations were recorded and are available via the *Digital Medieval Webinar Repository* (DMWR) at: https://zenodo.org/communities/dmwr (accessed July 14, 2023), and the YouTube channel of the *Networks of Manuscripts, Networks of Texts* conference, at: https://www.youtube.com/watch?v=IL9wTG_v5y8&list=PLK3oMFX57ubweIToV8O5S3vZWTpDjnHy (accessed July 14, 2023).

Prominent among the early pioneers were scholars of Nordic literature. In 2013, Alaric Hall and Katelin Parsons included a network representing the co-occurrence of Old Norse romance sagas in multitext manuscripts in their article about the transmission of *Konráðs saga keisarasonar*.[12] Two years later, Matias Blobel used network analysis in his MA thesis to explore the genre classification of Old Icelandic texts co-occurring in multitext manuscripts based on the metadata available in the manuscript database *Handrit*.[13] More recently, Katarzyna A. Kapitan and Tarrin Wills have expanded their research, deploying network analysis on a larger corpus of Old Norse manuscripts that transmit sagas in order to test the genre boundaries proposed for them by traditional scholarship.[14]

Independently of scholars of Nordic literature, networks of shared transmission in multitext manuscripts and books were explored by scholars of Latin and non-Nordic European vernacular literature. To provide just a few examples, Octave Julien employed network analysis to explore the co-occurrence of vernacular French and English texts in French and English multitext manuscripts from the late Middle Ages in 2016.[15] In 2018, Zdenko Vozár used the manuscript metadata available via the *Manuscriptorium* digital library[16] to examine the co-occurrence of Latin and Czech texts by the theologians of the Bohemian reformation in fourteenth-century manuscripts from Bohemia.[17] In the following years, Gustavo Fernández Riva articulated some of the general methodological precepts for analyzing shared manuscript transmission of medieval texts, demonstrating them on medieval texts written in German based on the metadata available in the online database *Handschriftencensus*.[18] Around the same time, N. Kıvılcım Yavuz experimented with visualizations of the co-occurrence network of late an-

12 Hall and Parsons, "Making Stemmas with Small Samples, and Digital Approaches to Publishing Them."

13 Blobel , "'Web'Scraping Parchment." The *Handrit* manuscript database is available at: https://handrit.is/ (accessed July 14, 2023).

14 They presented their preliminary results at the *Networks of Manuscripts, Networks of Texts* conference. The recording of their presentation is available at: https://youtu.be/bShYu_Ag-WA (accessed July 14, 2023). They were later published as Kapitan and Wills, "Sagas and Genre." Kapitan also made methodological observations about the limits of manuscript metadata currently available online for undertaking network analysis; see Kapitan, "Perspectives on Digital Catalogs and Textual Networks of Old Norse Literature."

15 Julien, "Délier, lire et relier."

16 *Manuscriptorium* is available at: https://www.manuscriptorium.com/en (accessed July 14, 2023).

17 Vozár, "Metadata for the Middle Ages."

18 Fernández Riva, "Network Analysis of Medieval Manuscript Transmission;" Fernández Riva and Millet, "Überlieferungsgemeinschaft in deutschsprachigen Handschriften." The *Handschriftencensus* database is available at: https://handschriftencensus.de/ (accessed July 14, 2023). Fernández Riva later expanded his research to also include metadata about medieval texts composed in French and Occitan (from the *Jonas* database) and the Ibe-

tique Latin narratives about the Trojan war in medieval manuscripts as a part of her project *Transtextual Networks in the European Middle Ages.*[19] Most recently, Elizabeth Archibald used network visualizations to study the distribution of grammatical texts in Carolingian grammatical handbooks and their mutual relationships,[20] and Sébastien de Valeriola and Bastien Dubuisson carried out a network-based examination of Latin hagiographies based on the *Bibliotheca Hagiographica Latina Manuscripta* database[21] to better understand which saints' lives tend to be clustered together.[22]

Another group of scholars who have engaged with network graphs is philologists, who look for alternatives to traditional stemmata for the transmission of complex textual collections and corpora.[23] To provide a recent example, networks proved useful for mapping the complexities of medieval glosses, which are often too 'unruly' to yield easily to a traditional genealogical approach. Two scholars, Bernhard Bauer and Evina Stein, attempted to disentangle the complex relationship between annotated manuscripts displaying gloss parallelism using network graphs as a supplement to, or a substitute for, traditional stemmata.[24]

Networks have also been used as models for understanding the complex relationships between ancient and medieval texts and their models and sources. Examples of this strand of historical network research include the 2014 study of the sources used by the medieval creators of the Würzburg commentary on Matthew by Malte Rehbein,[25] and the 2016 study of Plutarch's sources by Charlotte

---

rian languages (from the *Philobiblon* database), presenting his preliminary results at the *Networks of Manuscripts, Networks of Texts* conference, see the recorded presentation at: https://youtu.be/_MrS7FHGjQA (accessed July 14, 2023).

19 Yavuz, "Transtextual Networks in the European Middle Ages."

20 Archibald, "Carolingian Schoolbooks and Intellectual Networks: A New Approach." Archibald also presented her ongoing research examining connections between medieval monasteries using medieval library catalogs at the *Networks of Manuscripts, Networks of Texts* conference. The recording of this talk is available at: https://youtu.be/tJ_icEqd1h4 (accessed July 14, 2023).

21 Société des Bollandistes, "Bibliotheca Hagiographica Latina Manuscripta," accessed July 14, 2023, http://bhlms.fltr.ucl.ac.be/.

22 de Valeriola, "Can Historians Trust Centrality?"; de Valeriola and Dubuisson, "L'hagiographie à l'aune du numérique."

23 See Hoenen, "The Stemma as a Computational Model."

24 Bauer, "The interconnections of St Gall, Stiftsbibliothek, MS 251 with the Celtic Bede manuscripts;" Bauer, "Venezia, Biblioteca Marciana, Zanetti Lat. 349. An Isolated Manuscript?"; Steinová and Boot, "The Glosses to the First Book of the *Etymologiae* of Isidore of Seville: A Digital Scholarly Edition;" Steinová and Boot, "Editing Glosses as Networks." Stein also presented her preliminary results at the *Networks of Manuscripts, Networks of Texts* conference. The recording of this talk is available at: https://youtu.be/jL_409zSwL4 (accessed July 14, 2023).

25 Rehbein, "From the Scholarly Edition to Visualization."

Schubert.[26] Within this category we should perhaps also mention the potential utility of co-citation networks devised by modern scientometrists for adaptation in historical co-citation studies. A solitary example of such an adaptation is a 2021 study carried out by Richard Pollard and Anne-Gaëlle Weber, who examined the co-citation of ancient and medieval Latin authors in the *Patrologia Latina* database in order to assess which of them may have been counted among the medieval Church Fathers.[27]

Networks have likewise served to model the diffusion of ideas using texts and text-bearing objects as a proxy. Perhaps the most ambitious project of this type was carried out by Matteo Valleriani, whose research team studied the diffusion and evolution of astronomical knowledge in Europe based on early modern university textbooks in the context of *The Sphere* project.[28] Several other projects presented at the *Networks of Manuscripts, Networks of Texts* conference also fall into this category.[29]

Finally, this overview should not omit mentioning the existence of social network analysis that makes use of texts and text-bearing objects as proxies for human relations or historical social networks. There are far too many projects carried out within this strand of historical network research for us to be able to mention them all here, not to mention that many are not strictly concerned with manuscripts or texts, nor do they contribute to our knowledge about them. Nevertheless, two types of such research do belong to manuscript and textual studies. First is the promising (albeit rarely undertaken) use of social network analysis in paleographic research to map the relationships between copyists, il-

26 Schubert, "Die Visualisierung von Quellennetzwerken am Beispiel Plutarchs."

27 Pollard and Weber, "Le canon des Pères à l'époque carolingienne et la place de Flavius Josèphe."

28 Valleriani et al., "The Emergence of Epistemic Communities in the Sphaera Corpus"; Zamani et al., "Evolution and Transformation of Early Modern Cosmological Knowledge"; Valleriani et al., "The Network of Early Modern Printers and Its Impact on the Evolution of Scientific Knowledge." See also the project website; Valleriani et al., "The Sphere. Knowledge System Evolution and the Shared Scientific Identity of Europe."

29 These include *The Irish Foundation of Carolingian Europe – the case of calendrical science (computus)* led by Immo Warntjes at the Trinity College Dublin (https://computus.tchpc.tcd.ie/) making use of early medieval computistic manuscripts to trace the diffusion of scientific ideas in Carolingian Europe, and Sara Steffen's PhD project at the University of Basel *"Von der Eydgnoschafft will ichs heben an […]". Liedflugschriften als (vokale) Medien eidgenössischer Bündnisbeziehungen im 16. Jahrhundert* focusing on the diffusion of ideas via the practice of contrafactum in the printed ballads of the sixteenth-century Switzerland. The recording of Immo Warntjes's presentation is available at: https://youtu.be/AJPru72uUTA (accessed July 15, 2023). The recording of Sara Steffen's presentation is available at: https://youtu.be/CAnzH-AKNog (accessed July 15, 2023). Another project with a similar ambition of tracing the diffusion of ideas using medieval manuscripts is Bisagni et al., "Ireland and Carolingian Brittany."

luminators and other historical individuals that collaborated together within a manuscript-producing institution. To our knowledge, the only completed study of this type is the contribution by Katharina Kaska within this issue. Second, scholars have likewise been interested in using network visualizations and analysis to investigate the historical networks of book producers and owners in order to better understand the reception of specific works or authors. Recent examples of research projects in this area include the *ReVISION* project led by Laura Saetveit Miles, which examines the circulation of works by Saint Birgitta of Sweden in England.[30]

This special issue compiles six articles that grew out of and expanded upon some of the presentations at the aforementioned conference, *Networks of Manuscripts, Networks of Texts*. The papers not only offer concrete test cases of the use of network analysis in the field of manuscript studies, but also methodological reflections and diverse innovations. Thematically, they reflect a variety of perspectives and deal with many different kinds of texts from the European Middle Ages. The special issue follows an approximately chronological order, spanning from early medieval sermons and glosses to late medieval chronicles.

The article by Shari Boodts and Iris Denis showcases how network analysis tools can be incorporated and applied in the context of a wider research project and a scholarly database. The authors deal with patristic sermons and their textual transmission, and are the scholars behind a digital research tool for their study: PASSIM. These sermons, as with many other textual genres, display a high degree of variation and it is not uncommon to find extensive rewriting of one or more sources to create a testimony that can in many ways be considered a new text. Network visualizations are included in PASSIM to allow researchers to explore the database and the complex relationship between the sermons. The article explains the main features of these visualizations as well as their limitations and uses one sermon as an example of the kind of research enabled by their tool.

Evina Stein's article deals with mapping and analyzing organic corpora of medieval glosses. Unlike gloss commentaries or scholia, organic glosses did not come into being in a systematic fashion, but rather arose from the uncoordinated activity of many small, anonymous annotators with modest objectives. They were likewise circulated in an unsystematic fashion and as a result, annotated manuscripts of certain medieval texts display gloss parallelism that is highly indicative of some form of transmission; however, it does not allow scholars to speak of the existence of commentary traditions. Stein proposes a network-based method

30 Miles, Zieman and King, "ReVISION: Re-assessing St. Birgitta and her Revelations in Medieval England."

to understand how organic glosses may have been transmitted, demonstrating this on the corpus of early medieval glosses to the first book of the *Etymologiae* of Isidore of Seville. Based on a network analysis of a co-occurrence network of glosses shared by multiple manuscripts, she identifies several regional patterns of gloss transmission and chronological layers of the corpus studied. She also discusses some of the limits of the methods, avenues for its future refinement, and applications for the study of medieval glossing.

Katharina Kaska's article analyses scribal collaboration in the scriptoria of three Austrian Cistercian houses – Heiligenkreuz, Zwettl, and Baumgartenberg – in the twelfth century. This article combines newly produced data with previously available palaeographic observations, proposing a network-based model to represent and study participation of scribes in the copying of multi-scribe manuscript both within the three Cistercian houses and in collaboration among them. This method offers many advantages over the tabular presentation found in previous publications on the subject. The challenges of data gathering and analysis posed by the material are carefully considered. An important methodological aspect of this paper is the differentiation between manuscripts, texts and codicological units, which is not always considered in databases and analyses of manuscript transmission. Kaska then showcases the utility of her network model on the case of the hagiographic collection *Magnum legendarium Austriacum*.

The article by Dominique Stutzmann and Louis Chevalier deals with books of hours, a profusely transmitted kind of late medieval devotional manuscript book. As "compilations of compilations", these manuscripts consist of many discrete textual parts, which are often themselves constituted by smaller units. The order and composition of the individual units is key to identifying how different manuscripts of books of hours relate to each other, as well as the origins and circulation of each collection. In this paper, Stutzmann and Chevalier review past attempts to organize this vast material using diverse philological methods, and offer network analysis as a complementary tool that can help to determine connections and identify patterns of similarity. This method enables them, for example, to better explore hybridization – cases where certain books showcase a mix of two or more common and stable types. The authors understand that certain relevant aspects of the books cannot be easily modelled with networks, such as the order of the textual units; they thus combine network visualization with other methods to better explore these aspects.

Ina Serif's article explores the combination of network analysis and topic modelling. This combination should address two difficulties when performing a network analysis of shared manuscript transmission. It can create matrices of similarity between documents based solely on a full text transcription that can be created with handwritten text recognition tools, such as *Transkribus*. For that reason, the method can be applied to manuscripts that have not been previously edited or catalogued with enough metadata in a digital format. It also avoids an

often-subjective human determination of textual units. By replacing shared textual units with automatically detected shared topics, it is possible to compare the similarities between two documents and the text they transmit in a more abstract way, without the need to identify textual units. Using a late medieval German chronicle, Serif proposes and explores these methods, identifying their benefits but also their many as yet unsolved challenges and difficulties.

Catherine Emerson's article combines the more traditional social networks with networks of manuscripts in her analysis of manuscripts of fifteenth-century French chronicles and their owners, particularly the works of Nicole Gilles. Several different kinds of networks are represented in this article: networks of manuscripts and people linked by documentary evidence; networks of individuals connected by their ownership of manuscripts; and networks of manuscripts based on their material features. This last type of network provides an interesting way of comparing manuscripts, which is very different from the more usual shared textual transmission, and which is potentially worth exploring in a much wider scope.

As editors of this special issue, we hope this collective publication – the first to deal with historical network research in manuscript and textual studies – will not be the last. We are well aware that many methodological and theoretical obstacles still need to be overcome before the network-based approach appeals to the broader community of traditionally trained manuscript specialists and textual scholars. We realized this in part by observing that a common thread that runs through all the studies published in this issue, is that the authors faced similar challenges. These include problems arising from the reuse of pre-existing data, the extensive need for data preparation and cleaning, the definition of the units of analysis (manuscript, codicological unit, text), dealing with uncertainty within and fragmentation of historical evidence, as well as representing historically evolving networks. Even as this issue is published, it is fair to admit that the utility of the networks for the research of historical manuscripts and texts is yet to be proven. It is our wish that this special issue will soon be followed by other dedicated publications that lead to a maturation of these network-based methodologies within manuscript and textual studies. For now, let it be the first swallow heralding the coming of a new season.

References

Archibald, Elizabeth P. "Carolingian Schoolbooks and Intellectual Networks: A New Approach." In *Social and Intellectual Networking in the Early Middle Ages,* edited by Michael J. Kelly and K. Patrick Fazioli, 69–92. Binghamton, NY: Punctum Books, 2023. https://www.jstor.org/stable/jj.3452822.6.

Bauer, Bernhard. "The interconnections of St Gall, Stiftsbibliothek, MS 251 with the Celtic Bede manuscripts." *Keltische Forschungen* 8 (2019): 31–48. doi:10.25365/kf-8-2019-31-48.

— "Venezia, Biblioteca Marciana, Zanetti Lat. 349. An Isolated Manuscript? A (Network) Analysis of Parallel Glosses on Orosius' Historiae Adversus Paganos." *Études celtiques* 45 (2019): 91–106.

Bisagni, Jacopo et al. "IrCaBriTT – Ireland and Carolingian Brittany: Texts and Transmission." Accessed July 15, 2023. https://ircabritt.nuigalway.ie/.

Blobel, Mathias. "'Web'Scraping Parchment: Investigating Genre through Network Analysis of the Electronic Manuscript Catalogue Handrit.is." MA Thesis, Reykjavík: University of Iceland, 2015. https://skemman.is/handle/1946/22669.

Brughmans, Tom, Anna Collar, and Fiona Coward, eds. *The Connected Past: Challenges to Network Studies in Archaeology and History.* Oxford: Oxford University Press, 2016. doi:10.1093/oso/9780198748519.001.0001.

Brughmans, Tom, and Matt Peeples. "Trends in Archaeological Network Research: A Bibliometric Analysis." *Journal of Historical Network Research* 1 (2017): 1–24. doi.org/10.25517/jhnr.v1i1.10.

Fazioli, K. Patrick. "Modeling the Middle Ages: A Review of Historical Network Research on Medieval Europe and the Mediterranean World." In *Social and Intellectual Networking in the Early Middle Ages,* edited by Michael J. Kelly and K. Patrick Fazioli, 37–68. Binghamton, NY: Punctum Books, 2023. https://www.jstor.org/stable/jj.3452822.5.

Fernández Riva, Gustavo. "Conference Report: Networks of Manuscripts, Networks of Texts. Amsterdam (Online), 21–23 October 2020." *Comparative Oriental Manuscript Studies Bulletin* 7 (2021): 247–51. doi:10.25592/uhhfdm.9885.

— "Network Analysis of Medieval Manuscript Transmission: Some Basic Principles and Methods." *Journal of Historical Network Research* 3 (2019): 30–49. doi:10.25517/jhnr.v3i1.61.

Fernández Riva, Gustavo, and Victor Millet. "Überlieferungsgemeinschaft in deutschsprachigen Handschriften." *Zeitschrift für deutsches Altertum und deutsche Literatur* 151, no. 2 (2022): 179–200. doi:10.3813/zfda-2022-0006.

Gamper, Markus, Linda Reschke, and Marten Düring, eds. *Knoten und Kanten III: Soziale Netzwerkanalyse in Geschichts- und Politikforschung.* Sozialtheorie. Bielefeld: transcript Verlag, 2015. doi:10.1515/9783839427422.

Hall, Alaric, and Katelin Parsons. "Making Stemmas with Small Samples, and Digital Approaches to Publishing Them: Testing the Stemma of Konráðs Saga Keisarasonar." *Digital Medievalist* 9 (2013). doi:10.16995/dm.51.

Hoenen, Armin. "The Stemma as a Computational Model." In *Handbook of Stemmatology: History, Methodology, Digital Approaches,* edited by Philipp Roelli, 226–41. Berlin: De Gruyter, 2020.

Julien, Octave. "Délier, lire et relier. L'utilisation de l'analyse réseau pour construire une typologie de recueils manuscrits de la fin du Moyen Âge." *Hypothèses* 19, no. 1 (2016): 211–24. doi:10.3917/hyp.151.0211.

Kapitan, Katarzyna Anna. "Perspectives on Digital Catalogs and Textual Networks of Old Norse Literature." *Manuscript Studies* 6 (2021): 74–97. doi:10.1353/mns.2021.0002.

Kapitan, Katarzyna Anna, and Tarrin Wills. "Sagas and Genre: A Case for Application of Network Analysis to Manuscripts Preserving Old Norse-Icelandic Saga Literature." *Digital Scholarship in the Humanities* (2023), fqad013. doi:10.1093/llc/fqad013.

Kerschbaumer, Florian, Linda von Keyserlingk-Rehbein, Martin Stark, and Marten Düring, eds. *The Power of Networks: Prospects of Historical Network Research.* Digital Research in the Arts and Humanities. London: Routledge, 2020. doi:10.4324/9781315189062.

Knappett, Carl. "Networks in Archaeology: Between Scientific Method and Humanistic Metaphor." In *The Connected Past: Challenges to Network Studies in Archaeology and History,* edited by Tom Brughmans, Anna Collar, and Fiona Coward, 21–33. Oxford: Oxford University Press, 2016. doi:10.1093/9780198748519.003.0007.

Miles, Laura Saetveit, Katherine Zieman, and Julia King. "ReVISION: Re-assessing St. Birgitta and her Revelations in Medieval England: Circulation and Influence, 1380–1530." Accessed July 15, 2023. https://birgitta.uib.no/s/birgitta/page/home.

Pollard, Richard Matthew, and Anne-Gaëlle Weber. "Le canon des Pères à l'époque carolingienne et la place de Flavius Josèphe." *Revue d'Etudes Augustiniennes et Patristiques* 67, no. 2 (2021): 275–318. doi:10.1484/J.REA.5.131239.

Rehbein, Malte. "From the Scholarly Edition to Visualization: Re-Using Encoded Data for Historical Research." *International Journal of Humanities and Arts Computing* 8, no. 1 (2014): 81–105. doi:10.3366/ijhac.2014.0121.

— "Historical Network Research, Digital History, and Digital Humanities." In *The Power of Networks,* edited by Florian Kerschbaumer, Linda von Keyserlingk-Rehbein, Martin Stark, and Marten Düring, 253–79. London: Routledge, 2020.

Reinhard, Wolfgang. *Freunde und Kreaturen: "Verflechtung" als Konzept zur Erforschung historischer Führungsgruppen: römische Oligarchie um 1600.* Schriften der philosophischen Fachbereiche der Universität Augsburg 14. Munich: Vögel, 1979.

Schubert, Charlotte. "Die Visualisierung von Quellennetzwerken am Beispiel Plutarchs." *Digital Classics Online* 2, no. 1 (2016): 68–87. doi:10.11588/dco.2016.1.23825.

Smith, R. M. "Kin and Neighbors in a Thirteenth-Century Suffolk Community." *Journal of Family History* 4, no. 3 (1979): 219–56. doi:10.1177/036319907900400301.

Steinová, Evina, and Peter Boot. "The Glosses to the First Book of the Etymologiae of Isidore of Seville: A Digital Scholarly Edition." Amsterdam: Huygens Institute, 2021. https://db.innovatingknowledge.nl/edition/#left-home.

— "Editing Glosses as Networks: Exploring the Explorative Edition." *A Companion to Digital Editing Methods* (forthcoming).

de Valeriola, Sébastien. "Can Historians Trust Centrality? Historical Network Analysis and Centrality Metrics Robustness." *Journal of Historical Network Research* 6, no. 1 (2021): 85–125. doi:10.25517/jhnr.v6i1.105.

de Valeriola, Sébastien, and Bastien Dubuisson. "L'hagiographie à l'aune du numérique. Nouvelles perspectives d'étude des légendiers latins." *In Monte Artium* (forthcoming). Brepols. https://orbilu.uni.lu/handle/10993/52408.

Valleriani, Matteo, Florian Kräutli, Maryam Zamani, Alejandro Tejedor, Christoph Sander, Malte Vogl, Sabine Bertram, Gesa Funke, and Holger Kantz. "The Emergence of Epistemic Communities in the Sphaera Corpus: Mechanisms of Knowledge Evolution." *Journal of Historical Network Research* 3 (2019): 50–91. doi:10.25517/jhnr.v3i1.63.

Valleriani, Matteo, Hasan el-Hajj, Anika Merklein, et al. "The Sphere. Knowledge System Evolution and the Shared Scientific Identity of Europe." Accessed July 15, 2023. https://sphaera.mpiwg-berlin.mpg.de/.

Valleriani, Matteo, Malte Vogl, Hassan el-Hajj, and Kim Pham. "The Network of Early Modern Printers and Its Impact on the Evolution of Scientific Knowledge: Automatic Detection of Awareness Relationships." *Histories* 2, no. 4 (2022): 466–503. doi:10.3390/histories2040033.

Vozár, Zdenko. "Metadata for the Middle Ages: A Network Analysis of Manuscriptorium.com." International Medieval Congress, Leeds, 2018. https://www.muni.cz/vyzkum/publikace/1422901.

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

Wetherell, Charles. "Historical Social Network Analysis." *International Review of Social History* 43 (1998): 125–44. doi:10.1017/S0020859000115123.

Yavuz, N. Kıvılcım. "Transtextual Networks in the European Middle Ages: A Digital Corpus of the Trojan Narrative in Latin Manuscripts." Accessed July 14, 2023. https://www.transtextual.net/.

Zamani, Maryam, Alejandro Tejedor, Malte Vogl, Florian Krautli, Matteo Valleriani, and Holger Kantz. "Evolution and Transformation of Early Modern Cosmological Knowledge." *Scientific Reports* 10, no. 1 (2020): 19822. https://doi.org/10.1038/s41598-020-76916-3.

The Journal of HISTORICAL NETWORK RESEARCH

IRIS DENIS/SHARI BOODTS

The pseudo-Augustinian S. App. 121 and its medieval textual connections

A testcase for the development of network visualizations in the PASSIM Research Tool

Journal of Historical Network Research 9 (2023) 1–35

Keywords Network visualizations, Medieval manuscripts, Reception Studies, Patristic and Medieval Preaching

Abstract This article investigates the benefits and challenges inherent in using networks to visualize and analyze the textual connections between Latin patristic sermons as transmitted in medieval manuscripts. Patristic sermons, which had a dynamic reception in the Middle Ages and were the subject of an extensive and complex scholarly tradition, are an ideal test case for an inquiry into the manipulations of texts as part of the process of textual transmission in the Middle Ages. Using the pseudo-Augustinian *sermo* 121 as a case study, we will first describe the textual history of the sermon. Subsequently, we will translate this narrative of the scholarly history of PS-AU s 121 into network visualizations of increasing complexity and reflect on the accuracy, usefulness, and challenges of this method for the study of the myriad textual connections between patristic sermons in medieval contexts. This case study on network visualizations is part of a larger project to develop a digital application and database, the PASSIM Research Tool, to chart the dissemination and manipulation of patristic preaching in the Middle Ages.

C²DH LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

UNIVERSITY OF LUXEMBOURG

1. Introduction

This article investigates the benefits and challenges inherent in using networks to visualize and analyze the textual connections between Latin patristic sermons as transmitted in medieval manuscripts. This inquiry deals very specifically with sermons, which have particular characteristics both with regard to their genre conventions, as well as in terms of their transmission and reception in the Middle Ages. However, the method we discuss can be applied to many different types of texts, and the conclusions extrapolated to achieve a better and broader understanding of medieval textual transmission and knowledge dissemination. This understanding has been a focal point of historical and philological research for a very long time, but the Digital Turn in the Humanities, and in Manuscript studies more specifically, is helping scholars to overcome many obstacles that have been in place for centuries.

First, we will provide some context to our inquiry, describing the specificities of patristic sermons and their medieval reception and outlining their aptness as a test case for our particular investigation. We will also introduce the PASSIM Research Tool, which provides the interface and dataset for the networks we will generate. Next, we will describe the textual history of the pseudo-Augustinian *sermo* 121, which we have selected as our case study. Subsequently, we will translate this narrative of the scholarly history of PS-AU s 121 into network visualizations of increasing complexity. Finally, we will reflect on the accuracy, usefulness and challenges of this method for the study of textual connections between patristic sermons in medieval contexts.

Let us begin with the medieval context of textual transmission and knowledge dissemination, which has the manuscript as its central element. Every manuscript, whether plain or lavish, old or young, meticulously copied or haphazardly

Acknowledgements: This article was written in the context of two research projects: 'Patristic Sermons in the Middle Ages. The Dissemination, Manipulation, and Interpretation of Late-Antique Sermons in the Medieval Latin West (PASSIM)', funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No. ERC-2018-stg 802210 (2019–2023), and 'On the trail of Alanus of Farfa. Tracing the formation of Augustine's authority in medieval sermon collections for the liturgy (ALANUS)', funded by the Netherlands Organization for Scientific Research (NWO) under the Innovational Research Incentives Scheme: VENI, under grant number 016.Veni.195.127 (2018–2022). See https://applejack.science.ru.nl/passimproject/ for more information on both projects. We would like to thank Riccardo Macchioro, Gleb Schmidt, Menna Rempt, Erwin Komen, and Thijs Hermsen – all members of the PASSIM research team – for their contributions to the development of the dataset and research tool presented in this article.
Corresponding author: Iris Denis, Radboud University Nijmegen, iris.denis@ru.nl; and Shari Boodts, Radboud University Nijmegen, shari.boodts@ru.nl

thrown together, represents an act of reception, a very small cog in a very large medieval machine that toiled tirelessly to disperse knowledge and that has unalterably changed our collective literary and intellectual heritage. The problem is, of course, that the reconstruction of the medieval 'machine' of textual transmission is extremely challenging, and gaining an accurate picture of each individual manuscript as a driver of reception is to a certain extent contingent upon an understanding of the entirety of the transmission.

The main method by which scholars have attempted to penetrate the depths of this machine of knowledge dissemination is through the discipline of stemmatology: the charting of the genealogical relations, or 'family tree', of extant manuscript copies of a text, in order to reconstruct the most original version currently accessible to us based on the copies that remain.[1] Almost as a by-product, stemmata also involve information on the historical context of the medieval copies, and from there the medieval manipulation and interpretation – the medieval reception – of the text in question. As anyone who has ever attempted it can attest, building a stemma is an arduous and time-consuming task, often laced with frustration. There is usually a lot of uncertainty, textual variations can be interpreted in multiple ways, contamination in the manuscripts is rife, etcetera. Because of the labor-intensive nature of stemmatology, it tends to focus on small individual case studies and traditions. A natural consequence of this situation is the fact that modern scholarship is inclined to focus on texts and manuscripts that have been identified as 'significant' – a frustratingly vague term – by previous generations of philologists, and that we are still leaving large parts of the corpus of medieval texts and manuscripts untouched.

Still, much progress has been made since the Digital Turn in Manuscript Studies.[2] One of the objectives of the Digital Turn is to facilitate and speed up the process of generating a stemma, to make it possible to easily compare and combine traditions, both of manuscripts and of texts.[3] The experiment we undertake in this article is part of this scholarly trend. Though our focus is primarily on con-

1 For further information as well as a critical reflection on stemmatology, see Tarrant, *Texts, Editors, and Readers*; Roelli, *Handbook of Stemmatology*. Another useful resource is Echard and Partridge, *The Book Unbound*.

2 To get a sense of developments in the field, see, for example, Hamidović, Clivaz and Bowen Savant, *Ancient Manuscripts in Digital Culture*; van Lit, *Among Digitized Manuscripts* as well as the activities and publications of the *Digital Medievalist* community (https://digitalmedievalist.wordpress.com/, last accessed 24 July 2023). A special issue of the *Journal of Data Mining and Digital Humanities* is devoted entirely to recent and future developments in the field of Digital Manuscript Studies (https://jdmdh.episciences.org/page/on-the-way-to-the-future-of-digital-manuscript-studies#, last accessed 24 July 2023).

3 See Roelli, *Handbook of Stemmatology*; Andrews and Macé, *Analysis of Ancient and Medieval Texts*; Pratt et al., *The Dynamics of the Medieval Manuscript*.

nections between (the versions of) texts, we will see that the medieval vehicles of these texts, the manuscripts, are never far away.

Our particular approach is to look at Latin patristic sermons, i.e., the sermons preached by the Fathers of the Early Church – Augustine, Gregory, Leo – and their contemporaries in the period between, roughly, 350 and 750 AD.[4] The importance of patristic preaching has now been widely acknowledged for several decades. These sermons are valuable sources for our understanding of the early stages of Christianity. They tell us about the process of Christianization, the development of rituals, the competing heresies and their levels of success, and much more. They also provide unique insights into the personal development and thinking of some of the most important authorities in history, and especially show us the nuances that are absent from the more formal theological treatises of the Church Fathers.

However, the patristic sermons' relevance for the history of early Christianity is only half of their story. These texts had a dynamic and complex medieval afterlife. They circulated throughout the Middle Ages, usually as part of larger collections of patristic preaching. Both the collections and the sermons themselves were heavily manipulated as part of their medieval reception. Additionally, they circulated alongside and became intertwined with an enormous corpus of pseudo-epigraphic sermons that were attributed to one of the authoritative Church Fathers but of which the origin, whether late antique or medieval, is often uncertain.[5] Patristic sermons in medieval contexts demonstrate variation on multiple levels: there is variation in the composition of collections of texts and in their paratexts, there is textual variation of such intensity that texts can be considered deliberately different versions while clearly based on the same source, and of course there are the inevitable small (or not so small) errors introduced by the scribe who had the intention of copying the text faithfully.

It is on the second of these three types of variation that our focus will be in this article. Patristic sermons showed themselves to be a very malleable genre, inviting myriad adaptations as they were reused in and refitted for new contexts in the Middle Ages.[6] Without aiming to be exhaustive, we mention here for the purpose

4 For a general overview of Latin patristic preaching as well as references to further reading, see Dupont et al., ed., *Preaching in the Patristic Era*. For the Greek perspective, see, in the same series though much older, Cunningham and Allen, *Preacher and Audience*.

5 For more on the medieval reception of patristic preaching and the corpus of pseudo-epigraphic patristic sermons, see, for example, Boodts, "Navigating the vast tradition"; Diesenberger et al., *Sermo Doctorum*; Dolbeau, "La Transmission de la Prédication Antique de Langue Latine"; Pignot, *Latin Anonymous Sermons*; Weidmann, "Discovering Augustine's words."

6 See for example Martin, "The Italian Homiliary," especially 286–92, who argues that the author of the Italian Homiliary has occasionally adapted his patristic sources to his Caro-

of illustration some types of manipulation that patristic sermons underwent as part of their *Nachleben*.[7] Sermons could be abbreviated or text could be added to them, either newly created or taken from another existing sermon. Caesarius of Arles (d. 542), who adapted numerous Augustinian sermons for use by the priests of his diocese, is a well-known early example of this practice.[8] There were cut-and-paste jobs to merge sermons together, such as PS-AU s Mai 66, a *cento* which adapts Augustinian sermons for the martyrs Perpetua and Felicitas into a sermon for a different saint, Victoria.[9] Other times the connection is reduced to the level of an echo, where a few sentences or popular patristic one-liners are spliced into an otherwise entirely different text. We will encounter an example of this below, with the cluster surrounding the sermon known by its incipit, *hodie uerus sol*. On the other end of the spectrum, sermons can simply receive some new 'window dressing' with an alternative incipit or explicit. Such is the case for AU s 168, where an incorrect split between title and incipit in the archetype of the tradition has led different copyists to adjust the incipit slightly to correct the error.[10]

This abundance of versions and textual overlap, while highly interesting from the point of view of reception studies, can also be intensely frustrating when trying to rigidly catalogue and identify the sermons as found in the manuscripts.[11] Manuscripts are not printed books. Each rendition of a text in the manuscript tradition has the potential for much greater individual variation. Every copyist is a potential editor or even author. While scholars have certainly not been blind to the variety possible in the medieval reception of patristic preaching, there is a certain forcedness in trying to fit this reality into the existing modes of analysis and description, with an ever-present concern for the reconstruction of the original at the forefront. This focus makes it difficult to appreciate the dynamic of adaptation as a powerfully creative part of the medieval literary landscape. Of course, this tension is not exclusive to patristic preaching in medieval manuscripts. Many

---

lingian lay audience, even though the collection at a later stage would "make its entry into the monastic world" (295). See also Diesenberger, "Introduction," 8: "In the process of copying, compiling, and disseminating the sermons different techniques were used, which reflect the cultural variety that accompanied the production of these new texts."

7 We do not address here the issue of forgeries, which is a difficult term to use in the context of medieval attitudes to sermon adaptation, where many activities that to our modern eyes would be considered forgery (or at least plagiarism) were not perceived as such. Machielsen, "Contribution à l'étude" gives a list of possible categories of pseud-epigraphy, which, however, is not without its problems.

8 De Maeyer and Partoens, "Preaching in Sixth-Century Arles."

9 Weidmann, "Der Augustinuscento *Sermo Mai 66*."

10 Boodts, "The manuscript transmission of the *Quinquaginta homiliae*."

11 This tension is clearly visible in the seminal reference works for (pseudo-epigraphic) patristic preaching, such as Machielsen, *Clavis patristica*; Gryson, *Répertoire général*. See also Weidmann, "Discovering Augustine's words," which touches on both the problem of identification and the opportunities for further research in the corpus of pseudo-epigraphic patristic preaching.

other genres display the same flexibility and malleability, on the level of the text, the level of the collection, or both. The potential to document and explore this variety is present for schoolbooks, encyclopedias,[12] catenae, commentaries,[13] glosses, and verse narratives,[14] to name but a few. Our experiment with network visualizations of textual overlap in Latin patristic sermons and their medieval adaptations is a small contribution to this enterprise.

With the context and relevance of the inquiry now in place, we will move on to a description of the PASSIM Research Tool, which houses the data and interface for the networks we will generate.

2. The PASSIM Research Tool

The PASSIM Research Tool[15] in its current form aims at charting and analyzing the interrelations between the medieval manuscripts that transmit Latin patristic sermons. It consists of a database of metadata on manuscripts that transmit patristic sermon collections and an interface designed to allow for sophisticated querying of the data. In due course, this digital tool intends to grant access to a greater number of relevant manuscripts than has heretofore been studied. It is also meant to create a semblance of order in the extensive and complicated scholarly tradition of identifying Latin patristic sermons and their medieval incarnations. In doing so, the PASSIM Research Tool strives to open up new avenues for the study of the reception of Latin patristic preaching. Simultaneously, it intends to contribute to the methodological framework and digital toolbox necessary to further embed manuscripts in the study of medieval history.

The possibilities for analysis and visualization of the PASSIM Research Tool are emphatically not intended to supplant the existing research instruments. In fact, they make extensive use of existing reference works. The data are sourced from heuristic tools such as Claves, manuscript catalogues both digital and in print, relevant databases, critical editions, studies of the manuscript transmission of the Church Fathers, and case studies in the field's major journals, though

12 See, for example, the recent work of Evina Steinova; Steinova, "Two Carolingian Redactions"; Steinova, "The Oldest Manuscript Tradition."

13 See, for example, Witt, Christensen and Ueli, "Re-Conceiving the Christian Scholastic Corpus"; or the Scholastic Commentaries and Texts Archive (https://scta.info/, last accessed 24 July 2023).

14 See Pratt et al., *The Dynamics of the Medieval Manuscript*.

15 A beta-version of the PASSIM Research Tool is currently available, though (free) registration is required (https://passim.rich.ru.nl/, last accessed 24 July 2023). Additionally, the source code of the research tool is available on GitHub (https://github.com/ErwinKomen/RU-passim, last accessed 24 July 2023).

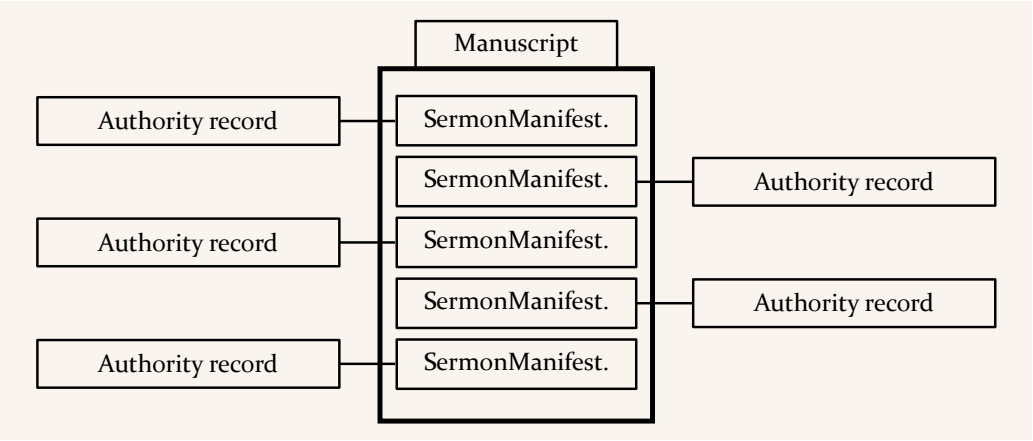

Fig. 1 Schematic representation of the PASSIM database structure.

they can, of course, be manually checked and enriched.[16] Simply put, the PASSIM Research Tool, through the integration and further enriching of the existing data, presents the information in such a way that more varied, wide-ranging, and nuanced inquiries are facilitated. Furthermore, whereas new discoveries or conclusions may currently remain isolated in journal articles for a long time, the digital context of PASSIM means that new findings can quickly become part of the shared knowledge of the scholarly community and the dataset at their disposal to do research in the field.

In the rest of this section we present the fundamental structural principles of the PASSIM database (Fig. 1). These structural features are important, as they will inform the network visualizations as we will present them later on.

At present, the interface allows entry into the database via three main queries: you can search for one or more manuscripts, you can look up the authority file of a sermon, which reflects the current scholarly *communis opinio* on a sermon's author and critical text,[17] and you can explore manifestations of a sermon, i.e., with all of the unique features the sermon has in a particular manuscript witness. Each manifestation of a sermon as found in a manuscript is linked to the corresponding authority file. In addition to these three central functionalities, searches for specific historical collections through keywords and other minor categories are also possible.

16 The entire list of sources used to populate the PASSIM database may be consulted through the Bibliography page (https://passim.rich.ru.nl/literature/list, last accessed 24 July 2023).

17 For the moment, however, the full text of the sermons is not accessible through the PASSIM Research Tool.

For every manuscript, the database contains a set of metadata including shelfmark, date, place of origin, provenance (if known), support, extent, format, as well as bibliography and external links to digital images or catalogue records. A composite manuscript is always separated into the codicological units it consists of. For each codicological unit, a detailed analysis of the content, including structural features such as the presence of *capitula,* is provided.

For every sermon manifestation (a sermon as it appears in a manuscript, with all the deliberate and accidental variants and changes made to the text and paratext by the scribe), the database contains the locus of the item in the manuscript, the attributed author, title (including section titles or *lectiones* that might precede the sermon in the manuscript), incipit, explicit, and, if applicable, postscriptum, feast or other liturgical occasions, Bible reference(s), and keywords. The manifestation is also linked to the corresponding authority file and the necessary identifying information from that authority file (reference codes or numbers, editions and literature) is provided.

An authority file for a sermon consists of the (real) author (according to the current academic consensus), incipit and explicit, a unique PASSIM code as well as other codes or reference systems currently in use, existing critical editions, and bibliography. Also listed are keywords, any known historical collections the sermon belongs to, and manuscripts in the database that contain this sermon. Furthermore, the authority file provides an archive of sorts of scholarly conclusions on the sermon – mostly author attributions – that are no longer upheld. Finally, the page also contains a list of all sermons (authority files) that – according to existing reference works and scholarship – share text with the sermon described. If known, the degree and direction of the overlap is indicated using a system of link types.

Links between authority files consist of two components, a quantitative and a qualitative/directional part (Fig. 2). The quantitative component roughly indicates the amount of textual overlap. The research tool in its current state offers three choices: 'nearly equals', for extensive textual overlap in the whole text, with differences only on the level of individual sentences; 'partially equals', when text overlap is concentrated on one or more paragraphs within the sermon; and 'echoes', for overlap limited to just a few sentences. Appointing a sermon to the 'partially equals' category is usually possible on the basis of information provided in secondary literature, as reference works tend to indicate which paragraphs are shared with another sermon. For the subtler 'nearly equals' and 'echoes' labels, recourse to a comparison of the critical texts is usually necessary. The label 'partially equals' is both the most common and most neutral, which is why it was chosen as the default setting. This simplest link type essentially indicates that there is some textual overlap between two authority files, but we cannot say anything more about it. Depending on the amount of information offered by the reference works we consult, as well as our own findings through a direct comparison

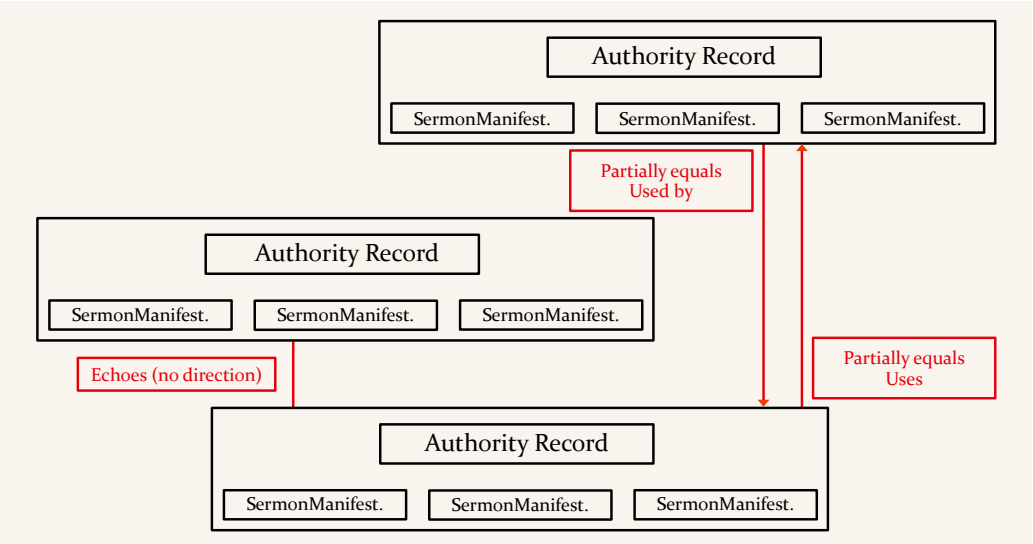

Fig. 2 Schematic representation of linked authority files in the PASSIM database.

of the texts, the link type can be further specified through the use of direction: 'is part of/has as its part', 'uses/used by', and 'uses/used by (indirect)'. The link type 'uses/used by (indirect)' indicates that we know that the connection is definitely not direct, and that there were known intermediaries. Another link type, 'common source' is used for texts that show textual overlap and that we know have independently used the same source, and 'unspecified' serves as a default directional specification, for those instances of textual overlap of which the precise circumstances are (still) unknown. Furthermore, it is possible to add a note, for example to indicate that a text is completely included in a larger text, to specify which parts of the texts overlap, or to refer to specific reference works or a bibliography. Of course, these specifications represent the current state of our knowledge on the texts, and may thus be updated as our knowledge evolves.

Quantitative	**Directional/explanatory (optional)**
Nearly equals	Is part of/Has as its part
Partially equals (default)	Uses/Used by
Echoes	Uses (indirect)/Used by (indirect)
	Common source
	Unspecified (default)

Tab. 1 Main link types used in the PASSIM database.

By connecting the authority files through this system of link types, we are effectively creating an annotated network of patristic and pseudo-epigraphic patristic sermons that reveals how the sermons changed over the course of their medieval transmission. The next step is to make this network visually accessible within the PASSIM Research Tool. To explore the benefits and challenges inherent in creating a serviceable network of a sermon's textual relations, we will use as a testcase the pseudo-Augustinian sermon 'App. 121,' a Christmas sermon on the topic of the Virgin birth.

3. The pseudo-Augustinian sermon App. 121

The pseudo-Augustinian sermon printed under the name 'App. 121' (hereafter PS-AU s 121)[18] is an apt example of the complex interrelations that can exist between texts labelled as 'patristic sermons.' The most recent edition[19] dates to 1845, but is in fact a near-exact reprint of a late seventeenth-century edition of Augustine's *Opera Omnia*.[20] It is the editors of this edition – the Benedictines of St. Maur – who are responsible for giving the sermon its number, and for placing it in the appendix to their edition of Augustine's authentic sermons, containing sermons they considered either inauthentic or of dubious authenticity. Hence the name 'Pseudo-Augustine, sermo Appendix 121', by which it has been identified ever since. The sermon has been the subject of several inquiries, which makes it very well-studied compared to many other pseudo-epigraphic texts.[21]

We also find an entry for PS-AU s 121 in a seminal resource for the study of pseudo-epigraphic preaching, Machielsen's *Clavis patristica pseudoepigraphorum medii aevi*. 1: *Opera homiletica* (CPPM).[22] This heuristic tool describes and cross-references a significant number of Latin sermons wrongly attributed to patristic preachers. Under the name Augustine, the most popular authority to which medieval scribes, readers and authors attached sermons, there are several thousand different items listed, a number that, while already very impressive, of

18 We present here a summary of the intertextual tradition of PS-AU s 121. The full dataset reflecting the scholarly tradition and underlying the network visualizations to follow, with reference to consulted literature and editions, can be accessed via "Authority file PASSIM 051.0103" in *PASSIM Research Tool, beta-version*. Radboud Institute for Culture and History, 2021 (https://passim.rich.ru.nl/ssg/details/1176, last accessed 24 July 2023).

19 Jean-Paul Migne, ed. *Patrologiae cursus completus. Series Latina*. Vol. 39, col. 1987–89. Paris: s.n., 1865.

20 *Sancti Aurelii Augustini Hipponensis episcopi operum tomus quintus, continens sermones ad populum* [...] Tomus V. Opera et studio monachorum ordinis S. Benedicti, è Congregatione S. Mauri, Parisiis, 1683–1684, col. 155–56.

21 Barré, "Le sermon pseudo-augustinien App. 121"; Bouhot, "L'homéliaire de Saint-Pierre du Vatican," 105–106; Aschoff, "Studien," 41–46. An overview of earlier scholarship on this sermon can be found in Barré, "Le sermon pseudo-augustinien App. 121," 111–12.

22 CPPM I 906, CPPM I 906a, CPPM I 906b.

course does not cover the full extent of sermons in medieval manuscripts wrongfully attributed to Augustine.

This lack of exhaustivity is an important realization, because it immediately highlights the inevitable incompleteness of any network generated by PASSIM. Until a full survey of every medieval manuscript that contains a sermon by or attributed to a patristic preacher has been made, it is a virtual certainty that there are still currently unknown versions or adaptations that we are missing. This is an inevitability to which we will return later on. For now, we want to make it clear that the narrative we will develop in this section and the visualizations that will follow, are based on a careful study of the existing reference instruments, studies and editions, supplemented here and there by the study of available manuscript witnesses. We have not, however, made a full transcription of every sermon referred to here in every manuscript witness. In other words, we have followed the traces and leads that are clearly defined in the existing scholarly literature, but have not gone far beyond that. This is a deliberate choice, because it is exactly the balance the PASSIM dataset as a whole is geared toward. We cannot make full analyses of all sermon manifestations, which run into the thousands, during the term of the project, but we do try to integrate as much as possible of what has been studied in the past within our database. Through the work of integrating, checking, and enriching these data we are, of course, able to add discoveries of new excerpts, versions, and manuscript witnesses, but our primary target is to amass as much data as we can.

Previous[23] scholarship has shown that **PS-AU s 121** as printed in the Maurist edition is, in fact, a compilation of two separate texts.[24] These two texts – **PS-AU s 121a** and **PS-AU s 121b** – were originally transmitted as part of decidedly different manuscript traditions.

PS-AU s 121a encompasses the first half of paragraph 1, plus paragraphs 4–5 from the text printed in *Patrologia Latina* as **PS-AU s 121**. Among its many manuscript witnesses we find various homiliaries (collections arranged according to the liturgical cycle) associated with the Abbey of Fleury.[25] It is further preserved in two well-known homiliaries currently housed in Wolfenbüttel[26] and Monte

23 To make it easier to compare the narrative we develop in this section with the visualizations in the next, the references to the sermons are printed in bold.

24 The Maurist editors already noted the interpolation that would later be labelled PS-AU s 121b (*Patrologia Latina* 39, col. 1987, n. b).

25 Orléans, Médiathèque municipale, 154 + Paris, BnF, NAL 1598 + Paris, BnF, NAL 1599, p. 14–17 (s. viii); Orléans, Médiathèque municipale, 155, p. 29–33 (s. x–xi).

26 Wolfenbüttel, Herzog August Bibliothek, Weiss. 12, f. 2r–2v (s. ix–x). This manuscript has lost its first leaves, which makes it impossible to say with certainty whether it transmits PS-AU s 121 or PS-AU s 121a. However, the Christmas sermons surrounding our text make it very likely that it was in fact PS-AU s 121a.

Cassino,[27] as well as the famous homiliary of Rochester Cathedral.[28] **PS-AU s 121a** often appears alongside several other Augustinian and pseudo-Augustinian Christmas sermons.[29] A shorter version of the sermon with a different incipit is transmitted in a codex associated with the Abbey of Lorsch.[30]

Textual overlap with **PS-AU s 121a** can be discerned in a few texts associated with Ambrose. Most notable among these is the fragment *In natale Domini* (**CPPM I 137**), which is considered by Barré to be an indirect source for **PS-AU s 121a**.[31] Our sermon also has textual passages in common with a number of other sermons and *centones*. These include two sermons found in the homiliary of Alanus of Farfa (**PS-ILD s 7**[32] and **PS-AU s 194**[33]).[34] Another variant connected to **PS-AU s 121a** is the Pseudo-Jerome's *cento* **CPPM I 5036**.[35] There is also an indirect connection between our sermon and the Late Antique compilation *Contra Iudaeos* (**AN Jud**).[36]

---

27 Monte Cassino, Biblioteca Statale del Monumento Nazionale, 12, p. 4–5 (s. x–xi).

28 Vatican, BAV, Vat. Lat. 4951, f. 15r–16r. A group of English manuscripts from the twelfth century (including Durham, Cathedral Library, Cod. B. IV. 12, f. 81v–82r, and London, British Library, Harley 3027, f. 157r–158r) transmits our sermon alongside a cluster of patristic texts that seem to overlap with the homiliary of Worcester; see Lambot, "La tradition manuscrite," 233 (especially n. 1); Sharpe, *English Benedictine Manuscripts,* 60–61.

29 See Wilmart, "Easter Sermons," 339–40 (especially n. 1); Barré, "Le sermon pseudo-augustinien App. 121," 112; Lambot "La tradition manuscrite," 227–39.

30 Vatican, BAV, Pal. Lat. 220, f. 69r–70v (s. ix$^{in.}$). Other early manuscripts containing PS-AU s 121a include Cambridge, University Library, Add. 3479, f. 178r–179r (s. ix$^{med.}$); Montpellier, Faculté de Médecine, H 59, f. 108v–109v (s. ix–x); Reims, Bibliothèque municipale, 296, f. 13v–14v (s. ix–x); Rome, Archivio di Stato, Rome, Osp. San Sal. 996, f. 45v–46v (s. xi).

31 The fragment is cited and ascribed to Ambrose by Cassian in his *De incarnatione contra Nestorium* VII, 25 (CSEL 17, 383–84, see also CPL 183). The fragment is seen as authentic and as an (indirect) source for PS-AU s 121a in Barré "Le sermon pseudo-augustinien App. 121," 115–21. However, Bouhot argues that the fragment is inauthentic and was probably taken from PS-AU s Cai I, 10a, which may have circulated under Ambrose's name; see Bouhot "L'homéliaire de Saint-Pierre du Vatican," 106, n. 15. See CPPM I 137 for an overview and further literature.

32 See CPPM I 5263 (= CPPM I 156).

33 It is the third paragraph of this sermon that shares text with PS-AU s 121a. See also CPPM I 979.

34 In Alanus's collection, these sermons are located in the Summer part, items II, 64–65 according to the (problematic) reconstruction in Grégoire, *Homiliaires,* 127–88. Bouhot mentions that these two sermons may have been created for the primitive Roman homiliary, of which Alanus of Farfa's collection is a close descendant; see Bouhot "L'homéliaire de Saint-Pierre du Vatican," 106, n. 14; and also below, section 4.4.

35 Folliet, "Deux nouveaux témoins," 181–82.

36 Aschoff, "Studien," 41–46; *Anonymi contra Ivdaeos,* especially viii, n. 16.

The second component, encompassing the second half of paragraph 1 and paragraphs 2–3 of **PS-AU s 121** as printed by the Maurist editors, is **PS-AU s 121b**. This sermon is thought to derive from a lost African sermon dating to the fifth century.[37] Our text is found in the ninth-century or tenth-century manuscript Montpellier, Fac. de Médecine H 59 (f. 81r–81v), which also transmits **PS-AU s 121a**. Moreover, the PASSIM description of the recently digitized Homiliary of Ottobeuren has made it unambiguously clear that this manuscript is another witness to **PS-AU s 121b**.[38] Our collation of these witnesses shows that the latter deviates in many small instances from the edited text, while the Montpellier manuscript is slightly closer to the text incorporated in **PS-AU s 121**, though both manuscripts also share a number of variants in contrast to the edition.[39] In addition, an abbreviated and interpolated variation of **PS-AU s 121b**, which incorporates various sources and has a different explicit, was apparently disseminated more enthusiastically. This reworking, associated in the reference instruments with Pseudo-

37 Bouhot, "L'homéliaire de Saint-Pierre du Vatican," 106. Two other derivations from this same lost sermon are PS-AU s Cai I, 13 and PS-AU s Cas II, 168. See also CPPM I 906b, CPPM I 1243, and CPPM I 1542.

38 A description of the Homiliary of Ottobeuren (Rome, Bibl. Naz. Centr. Vit. Em. II, Vitt. Em. 1190, s. ix$^{1/2}$) can be found via https://passim.rich.ru.nl/manuscript/details/1870/ (Menna Rempt 2022, last accessed 24 July 2023). Previous (otherwise excellent) descriptions of the manuscript are not unanimously clear on which version is actually present in the manuscript (f. 80r–81v). The digital images confirm that Barré rightly identified the Homiliary of Ottobeueren to contain PS-AU s 121b (expl. *Christus uobis hodie redemptor apparuit*); Barré, "Le sermon pseudoaugustinien App. 121," 121. More recent descriptions made this identification look doubtful. Gregoire refers to both PS-PET s Liv 5 and PS-AU s 121 but prints a different explicit than the one found in the manuscript; Bouhot writes that the text in the Homiliary of Ottobeueren is the version from 'Alanus of Farfa' (PS-PET s Liv 5); the entry for the manuscript in MANUS online refers to the text printed by Liverani and includes a different explicit; the detailed description found via Omeliari in scrittura beneventana refers to both PS-PET s Liv 5 and PS-AU s 121b; see Gregoire, *Homiliaires,* 325; Bouhot, "L'homéliaire de Saint-Pierre du Vatican," 106; MANUS online ((https://manus.iccu.sbn.it/risultati-ricerca-manoscritti/-/manus-search/cnmd/69660, last accessed 24 July 2023); and Omeliari in scrittura beneventana (http://omeliari.unicas.it/, last accessed 24 July 2023).

39 We limit ourselves here to only a few significant variants: in praesepio positum *ed. Ott*] positum in praesepio *Mont* fratres karissimi *add. Mont Ott*; nulli uirginitas seruituti succumbit *ed. Mont*] *om. Ott*; portabat *ed.*] gestabat *Mont Ott*; sancta credidit, sancta concepit *ed.*] sancta credidit *om. Mont Ott*; nasciturus ex uirgine *ed.*] ex uirgine nasciturus *Mont Ott*; mater *ed.*] antequam mulier *add. Mont Ott*; impraegnabatur *ed.*] implebatur *Mont Ott*; uita nobis hodie de coelo est data. Hodie super terram canunt angeli *ed.*] uita nobis hodie uenit super terram. Canunt angeli *Mont* uitam hodie super terram canunt angeli *Ott*; quia restauratur genus humanum per interitum *ed.*] quia restaurator hominum pro/per interitu *Mont Ott*; redemptor apparuit *ed.*] qui cum patre et spiritu sancto uiuit dominator (dominatur *Mont*) et regnat (regnat deus *Mont*) in saecula saeculorum amen *add. Mont Ott*.

Peter Chrysologus (**PS-PET s Liv 5**[40]), is found in the already mentioned Roman homiliary and homiliary of Alanus of Farfa, as well as several other prominent liturgical collections.[41]

To our knowledge, the compilation printed by the Maurists as **PS-AU s 121** is not extant in any known manuscript.[42] However, an intersection of **PS-AU s 121b** and **PS-AU s 121a** is not completely without precedent in the manuscript transmission. Both **PS-PET s Liv 5** (the version closely connected to **PS-AU s 121b**) and **PS-AU s 121a** are traced back to a shared principle source: **PS-AU s Cai I, 10a**.[43]

This is a brief but representative overview of the current consensus on this sermon and an apt illustration of the complex tangle of versions and manuscripts that characterizes the medieval reception of late antique sermons. We will now explore whether networks can have an added value in visualizing this tradition. Far from wanting to replace the meticulous textual comparisons between all these different versions executed by Barré, Bouhot and others, the network will not only be aimed at making the established textual links and connections instantly clear, in a single view, but also at allowing users to build on these scholarly observations through the accumulation and integration of the intertextual links of a greater number of texts in a single network. After all, the process of identifying and editing sermons, particularly anonymous or misattributed ones, is still very much ongoing, as testified by the newly edited texts and textual links discovered and published each year.[44]

---

40 See CPPM I 1965 (= CPPM I 6356). PS-PET s Liv 5 was edited in Liverani, *Spicilegium Liberianum* I 193–95; *Patrologiae Latinae Supplementum* III, col. 180–82. Its final paragraphs are also found in a sermon labelled PS-AU s Mai 176 (= CPPM I 1786), edited in Mai, *Nova Patrum Bibliotheca* I, 397–98; *Patrologiae Latinae Supplementum* II, col. 1274–75. The precise relations between this sermon and the compilation are not quite specified in scholarship.

41 The sermon is found, often combined with other texts, in at least the following collections and manuscripts: Homiliary of Alanus of Farfa I, 2e (cf. Grégoire, *Homiliaires,* 127–88; compilation); Homiliary of Egino of Verona, 2d (cf. Grégoire, *Homiliaires,* 189–221; compilation); Homiliary of St. Peter (Vatican, BAV, Arch.Cap.S.Peitro.C.105, 16d, f. 41v–43v, s. $x^{med.}$; cf. Grégoire, *Homiliaires,* 223–44; compilation); Venice, Biblioteca Marciana, ZL CLIII (1951), f. 72v (s. xii).

42 The assembly of a full heuristic overview is complicated by the fact that PS-AU s 121 and PS-AU s 121a (and sometimes even PS-AU s 121b) are not always clearly distinguished from each other in manuscript catalogues. See also Barré, "Le sermon pseudo-augustinien App. 121," 136.

43 See CPPM I 906a. Cf. Bouhot, "L'homéliaire de Saint-Pierre du Vatican," 106 (especially n. 15), who states that PS-AU s Cai I, 10a is an older, unedited version of the sermon printed in Caillau, *Collectio selecta,* 97–99 (PS-AU s Cai I, 10 = CPPM I 906).

44 A pertinent example is an Easter sermon recently edited and studied in Dolbeau, "Un sermon pseudo-augustien pour la fête de Pâques, confronté à ses sources," 111–23 (especially 116–17). Dolbeau's analysis reveals textual echoes between the opening lines of the newly edited sermon and PS-AU s 121 (and, we might add by extension, PS-AU s 121a), a

4. Network visualizations

The network visualizations shown and analyzed in this section are made using two programs. As part of the PASSIM Research Tool, the option to generate different types of visualizations is built into the application itself.[45] At the moment, we are still experimenting with customization options. Thus, the PASSIM visualization options are still under development. For the purpose of demonstrating the potential of such customization options, we have also used the program Node XL.[46] In the figures below, we have clearly distinguished which are at this time the product of the PASSIM application, and which were made using Node XL. It is our goal, however, to eventually integrate all options presented in this chapter into the PASSIM Research Tool.

The networks should be understood as follows. Every node represents an authority file, a 'standard' text (often the printed version). In our database, each of these authority files has its own PASSIM-code, and is linked to all manifestations of this text in the manuscripts currently in the database (Fig. 1 and Fig. 2 above). Every edge represents the recorded textual overlap – *any* textual overlap – between two sermons.

4.1 Simplest network

We start with the simplest version of the network, which shows all connections with PS-AU s 121 to the first level, i.e., mainly the texts mentioned in our summary of the tradition of PS-AU s 121a and PS-AU s 121b in the previous section (Fig. 4). These connections are quite easy to glean from the reference works and are also presented in a single list view on the page with the authority file for PS-AU s 121 in the PASSIM Research Tool (Fig. 3).[47]

While this network can, of course, be useful as an alternative way of looking at the data otherwise presented in a traditional list view, at this point, it does not

---

passage in Lactantius's *Divinae Institutiones* (IV, 18, 12) and several hagiographical texts. These discoveries have not yet been included in the network presented here. However, the flexibility of the PASSIM environment ensures that they can easily be added in future.

45 The PASSIM environment in principle offers the option to download and reuse both the images of the visualisation (in SVG and PNG format) and the underlying data (in JSON format). Note, however, that this functionality could be suspended for specific pre-existing datasets which have been imported in PASSIM with permission of their owners. Such datasets are recognisable in PASSIM through a project label.

46 Downloadable at https://nodexlgraphgallery.org/Pages/Registration.aspx (last accessed 24 July 2023). The visualizations presented in this article are generated through the Harel-Koren Fast Multiscale algorithm.

47 Cf. note 18.

echoes	Unspecified	(no Passim code)CPPM I 5036(by Undecided) hodie, dilectissimi nobis, uerus sol...-...pertingere merebimus A N
partially equals	Common source	(no Passim code)PS-AU s Mai 176 | CPPM I 1786(by Undecided) deus ab angelis proditur et agnus pastoribus demonstratur...-...coheredes christus dominus noster, qui… N
echoes	Uses (indirect)	PASSIM 012.0006AM vg | CPL 145(by Ambrosius Mediolanensis) si iuxta caelestis sententiam ueritatis uerbi...-...uicisset gladium quem quaerebat inuenit N
partially equals	Common source	PASSIM 051.0001AF I, 2e | PS-PET s Liv 5 | CPPM I 1965 | CPPM I 6356(by Anonymus) praedicamus hodie natum de uirgine...-...ut faceret cohaeredes christus dominus noster qui… N
partially equals	Uses (indirect)	PASSIM 051.0028PS-AU s Cai I, 10a | CPPM I 1240a(by Anonymus) diei huius aduentum...-...quia ipse est splendor gloriae aeternae domini nostri iesu christi N
partially equals	Common source	PASSIM 051.0186PS-AU s Cai I, 10 | CPPM I 1240(by Anonymus) diei huius aduentum si pleno possimus ore...-...bona sunt praestare dignetur per iesum christum dominum nostrum qui est benedictus in saecula saeculorum amen N
partially equals	Uses	PASSIM 051.0218PS-AU s 121a | CPPM I 906a(by Anonymus) quis tanta rerum uerborumque copia...-...gloria aeterna domino nostro iesu christo cui est…
partially equals	Common source	PASSIM 051.0243PS-AU s Cas II, 168 | CPPM I 1542(by Anonymus) dominus noster iesus christus fratres karissimi qui semper...-...christus nobis apparuit qui cum patre... N
partially equals	Common source	PASSIM 051.0244PS-AU s Cai I, 13 | CPPM I 1243(by Anonymus) si natiuitatem christi domini consideremus...-...manifestatus in carne qui cum patre... N
partially equals	Common source	PASSIM 051.0434AF II, 64 | PS-ILD s 7 | PS-MAX s 12 | CPPM I 156 | CPPM I 2257 | CPPM I 5062 | CPPM I 5263 | CPPM I 5578 | CPPM I 5945(by Anonymus) celebritas hodiernae diei nos admonet ut in laude...-...commendare dignetur in coelis apud dominum deum nostrum, qui… N
partially equals	Unspecified	PASSIM 051.0474AU s 194 add | CPPM I 581(by Anonymus) in hac enim die ad saluanda omnia...-...uirgineus uenter idoneus fuit
partially equals	Unspecified	PASSIM 051.1032AN Jud | CPL 360(by Anonymus) doctor gentium in fide et ueritate paulus exhortans nos...-...qui offerebat et quod offerebat explicit
partially equals	Uses	PASSIM 051.1290PS-AU s 121b | CPPM I 906b(by Anonymus) praedicamus hodie natum de uirgine...-...hodie redemptor apparuit N
partially equals	Unspecified	PASSIM 052.0005CPPM I 137(by Dubius) uidete miraculum matris dominicae...-...qui uenerat sanare corrupta
partially equals	Common source	PASSIM 052.0065AF II, 65 | PS-AU s 194 | CPL 368.194 | CPPM I 5041 | CPPM I 979(by Dubius) adest nobis dilectissimi optatus dies beatae...-...benedicta pretium ferre mundi N
partially equals	Common source	PASSIM 052.0186MAX s Mu 61B | [MAX] h 5 | CPL 219a.61b | CPL 220.5 | CPPM I 2072 | CPPM I 5759(by Dubius) proxima dominica dilectionem uestram admonuimus...-...ut possimus eum non timere cum iudicat

Fig. 3 List view of textual connections of PS-AU s 121 (PASSIM).

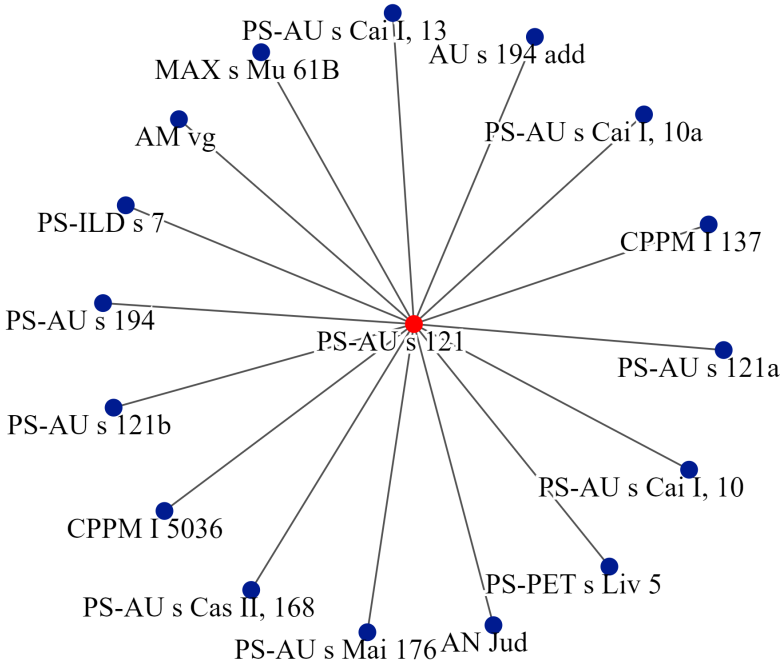

Fig. 4 Simple network of PS-AU s 121 (PASSIM).

offer any new information. Rather, the list view has the upper hand here, containing more details on the link types and on the sermons themselves.

4.2 Second-level network

The network does offer substantially new information once we visualize a higher degree of connections. For the first further development, we limit ourselves to the second level, which includes all of the sermons – or better yet, authority files – that demonstrate textual overlap with PS-AU s 121, and all of the authority files linked to this first set, i.e., those sermons that share texts with the sermons connected to PS-AU s 121, but not with PS-AU s 121 itself. In the network, the starting point, PS-AU s 121, is indicated in bright red, the first-level nodes are burgundy, and the second-level nodes are dark blue (Fig. 5).

This visualization is an interesting exercise in balance. It is not exhaustive, as there are still further connections that go beyond the current limits of the network. In fact, the visualization even distorts, or at least obscures, the reality of the texts. An example is the *cento* CPPM I 5036,[48] which is connected in the network to PS-AU s 121 and several of its sources and reworkings, and also to several nodes that represent versions of a sermon *Hodie uerus sol* (MAX s Mu 45 and PS-AU s Cai II, app. 22, but also the sermons that *use* either of these versions, PS-AU s Cai II, 26 and EUCH s 1). The network does not, however, show that these sermons also share text with each other, giving an impression of a large number of individual sermons overlapping separately with the compiled version CPPM I 5036, while in reality it is connected to a cluster.

However, this second-level network does offer an overview of all sermons that conceivably share a common source with the sermon we started with. While you can get to these data by clicking through the linked authority files on the page of PS-AU s 121, to find all further connections of the sermons listed there, it is clear the network has an advantage this time: it provides one comprehensive picture, but more importantly, the network also shows how those sermons listed as having textual overlap with PS-AU s 121 interconnect with each other.

Significantly, however, this added value is counterbalanced by several deficiencies. First, the network has no nuance: rather, every link type is represented by the same type of edge. Second, it does not contain any information about the manuscripts, which, if included, would provide context to the textual connections between the sermons. Third, the network is undirected, not offering any information on which sermons are the models, and which the derivations. The latter two deficiencies, we aim to correct in the PASSIM environment by including a number of custom options, which we will discuss further in section 4.4.

48 Folliet, "Deux nouveaux témoins," 181–82.

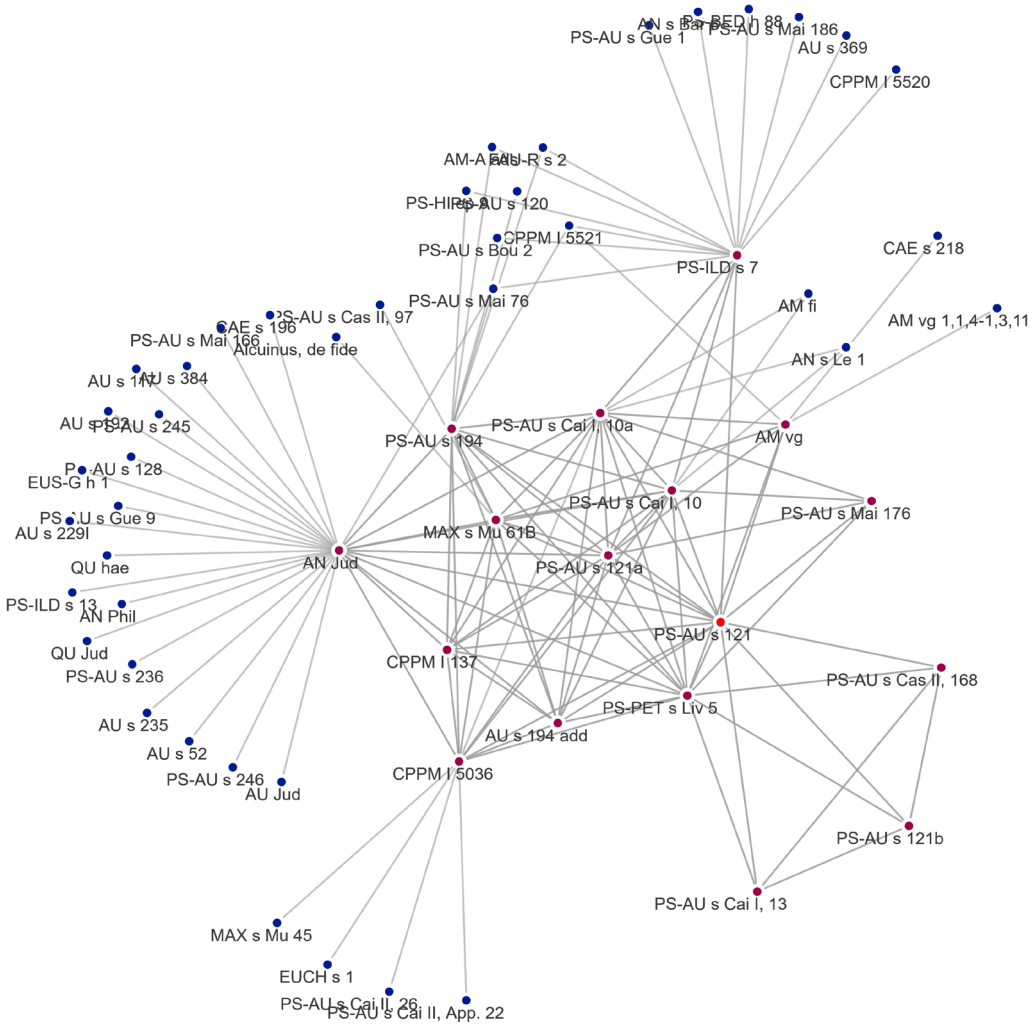

Fig. 5 Second-level network of textual connections of PS-AU s 121 (PASSIM).

The lack of nuance, however, remains a problem. For example, a passage from the aforementioned fragment CPPM I 137 (Pseudo-Ambrose's *In natale Domini*) is repeated in several sermons in the network. This short paragraph[49] – or part of

49 *Videte miraculum matris Dominicae: uirgo concepit, uirgo peperit, uirgo cum parturiit, uirgo grauida, virgo post partum, sicut in Ezechiele dicitur:* et porta erat clausa et non est aperta, quia Dominus transiuit per eam. *Gloriosa uirginitas et praeclara fecunditas. Dominus mundo nascitur, et nullus est gemitus parientis: uacuatur uterus, infans excipitur, nec tamen uirginitas uiolatur. Fas erat, ut Deo nascente meritum cresceret castitatis; nec per eius egressum uiolarentur integra qui uenerat sanare corrupta* (ed. Petschenig, *Cassianus,* 383–384).

it – is found in at least the following texts: CPPM I 137, PS-AU s Cai I, 10a, PS-AU s Cai I, 10, PS-AU s 121a, PS-AU s 121, MAX s Mu 61b, CPPM I 5036 (PS-HI), AU s 194 add, PS-PET s Liv 5, PS-AU s 194, and AN Jud. Some of these sermons, such as CPPM I 5036 and PS-AU s 194, share little more than a few lines with the *In natale Domini*-fragment, with some variation, while for others, such as PS-AU s 121a and PS-AU s Cai I, 10a, the overlap is far more substantial. We may include the option to customize the network by showing the link types through labels, or by making the edge darker or paler depending on the strength of the connection.

However, this does not address the true problem. The vast majority of the textual connections we have described for PS-AU s 121 fall in the default category 'partially equals.' Because our standard mode of data entry relies by necessity on reference works and catalogues, rather than full-text transcription and collation of all the relevant sermons, this will be the case for most sermons in the PASSIM database, especially considering the fact that most of them have not been as well-studied as the case study we are developing here.

The lack of nuance remains an issue, but it is, at least, an obvious one. Another, less conspicuous problem is the fact that the network shows us *all* textual connections expressed in the scholarly literature and reference works, i.e., also those that were hypothesized in scholarship, but may be mutually exclusive or proven incorrect at some stage. In the case of PS-AU s 121, different ideas by different scholars are at this point presented without preference for one over the other, because all instances of textual overlap are considered 'equal'. For example, even though PS-AU s Cai I, 10a is identified by Bouhot as a source for many of the sermons surrounding it (including PS-AU s 121a and PS-AU s Cai I, 10),[50] it appears in the network simply as one of the nodes connected in that cluster: it is the observation that there is textual overlap that determines the link between two nodes, not their *fons-usus* relationship. While the state of the data we use as source materials can only very gradually be improved upon, we must be aware of these weaknesses when using the visualization options.

Enriching the network with information on the types of links or their direction will always by necessity be an incomplete and imperfect process. It is highly unlikely that adding clear and unambiguous stemmatical information for all nodes in a large network is feasible. Also, adding stemmatical information highlights that many nodes are technically missing from the network. Two sermons which are connected because of textual overlap may be separated by several intermediary versions which are simply lost. The 'common source' link type refers to at least one other text which may not be extant. One could argue that adding specifications, directional or explanatory, to a network of textual overlap is muddying the waters. Given its imperfect state, it is certainly not recommended to use the

50 See note 43.

'nuanced' connections for grand proclamations on the entirety of the network. However, on a micro-level, when a single node or a small cluster is the point of interest, the nuances do give information that can be useful, if only to alert users to additional notes added to the authority files.

Researchers interested in PS-AU s 121 can, even before delving into the relevant secondary literature, gain a quick overview in a single glance of all of the texts (or better yet, authority files) they should investigate in order to get a full picture of the models, alternative versions and offspring of this sermon. They can spot, for example, that the two constituent parts of PS-AU s 121 – PS-AU s 121a and PS-AU s 121b – do not themselves connect to each other, that the former is part of an intricate cluster, but that the latter does overlap with a sermon – PS-PET s Liv 5 – that is also attached to this cluster. Also, if users were looking for their next case study, they might be intrigued by PS-ILD s 7, which appears – at least from this version of the network – to be either a *cento,* combining a great many sources, or a model that has been heavily used in other sermons.

This brings us to the next section, where we move away from what network visualizations can do for users who already know what they are looking for, and toward networks as gateways to serendipitous discoveries and motors for general, abstract, and overarching research questions.

4.3 Complete network

The visualization becomes vastly more complex and expansive still if we maximize the number of edges and follow the network to its natural limits (Fig. 6). This means that we add all the additional textual links to each of the nodes until there are no more textual links to add.[51] Eventually, we hope to generate a network that combines the entire set of authority files and their textual overlap in the PASSIM database. While this is at the moment still a distant point of the horizon, we can use a complete network of PS-AU s 121 as an illustration of the possibilities of such a comprehensive visualization.[52]

At this point, the connection with the sermon that formed our starting point is flimsy. As such, its usefulness for the study of PS-AU s 121 is limited. However, there are several other gains to be had. First, and most obviously, the network

51 Note, however, the methodological questions that arise when trying to establish textual links and visualizing information based primarily on previous scholarship (see below, 4.4 and 4.5). For the purposes of this experimental network, we have incorporated textual links (particularly between sermons) that are clearly and specifically defined in our reference works or the result of our own observations.

52 Because the functionality to calculate beyond six levels is not yet working properly in the PASSIM Research Tool, we show a visualization generated using Node XL, though based on the same dataset.

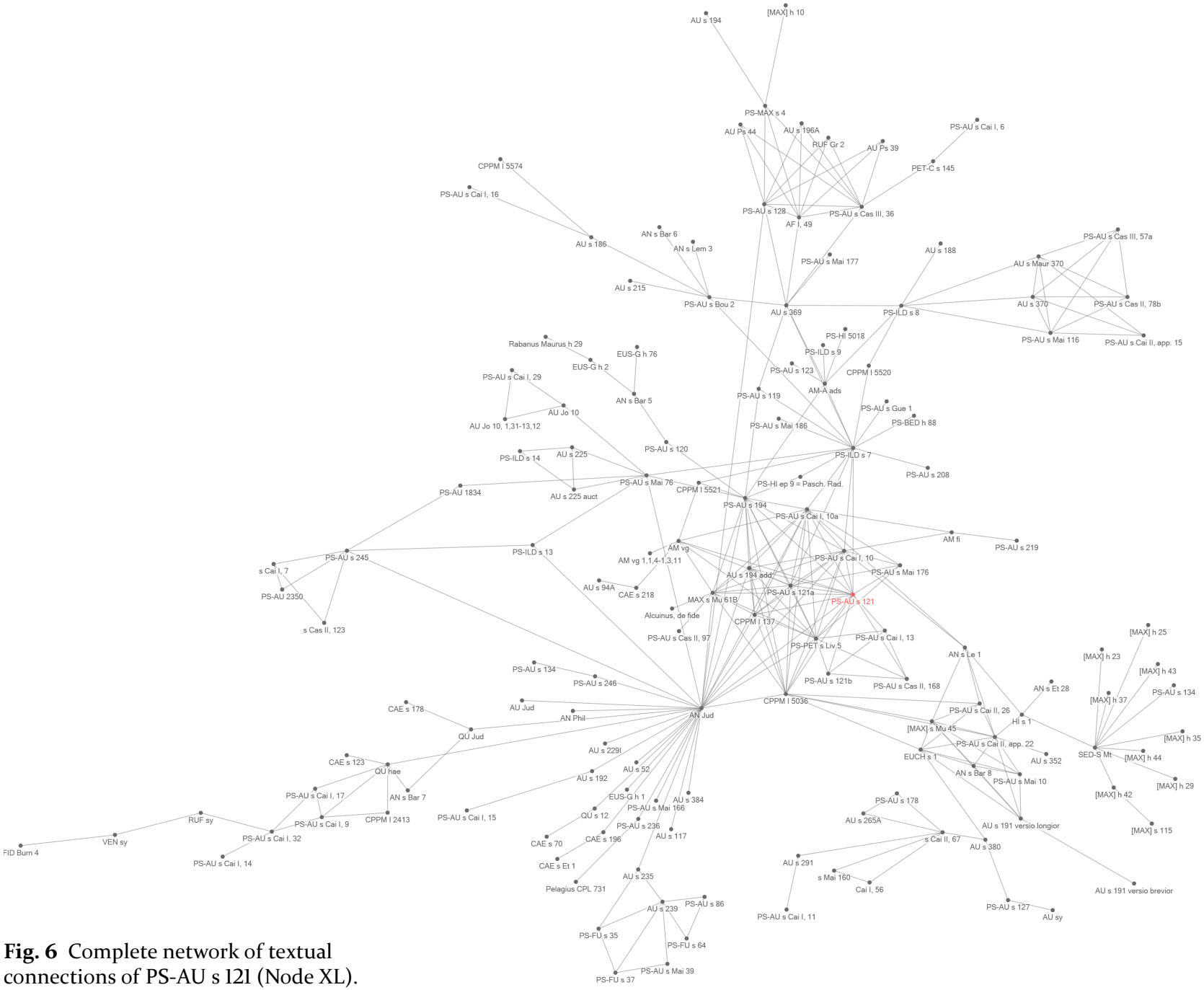

Fig. 6 Complete network of textual connections of PS-AU s 121 (Node XL).

presents in one view what would be a highly laborious task to puzzle together from reference works and secondary literature. Rather than giving any detailed information on the formation and influence of one particular pseudo-epigraphic sermon, the network gives an impression of the dynamic of a significant group of sermons, how they interacted, and how they are intertwined.

At this point, users would not necessarily have to start from a text they already know to find what they are looking for, but can rather explore the network for interesting case studies with a specific query or topic in mind. For example, for a scholar looking to investigate practices of rewriting and compiling patristic materials, the network can highlight the most promising texts to look into. The interesting clusters would be not only the set of overlapping texts surrounding PS-AU s 121a, but also the aforementioned cluster around sermons starting with *hodie uerus sol,* and – perhaps more significantly – the observation that at least two different texts (CPPM I 5036 and AN s Le 1)[53] form a 'bridge' between these two clusters without being connected to each other. In fact, both of these sermons are *centones* that incorporate and paraphrase snippets of texts that are shared by both of the clusters. This observation could, in turn, be an entryway for a researcher to also compare all versions, search for additional information on the manuscripts that transmit these versions, and – for instance – try to identify or reconstruct a manuscript or collection that may have been available to the compilers of both *centones*. Still, the network at this point of course does not guarantee that such potential gateways necessarily provide the expected results. To more efficiently highlight potentially interesting case studies, the network needs further specification and adaptability (see below, 4.4).

The function of each of these anonymous sermons as 'bridges' between clusters illustrates an additional point. When we move from specific queries to the methodological perspective, it is particularly interesting that this network partially reshapes the traditional perspective: the eye is not drawn primarily to the 'originals', namely the authentic sermons by the authoritative Church Fathers. Many of these are lingering at the outskirts of the network, while a more central position in the network is occupied by the texts that have the most edges, i.e., that connect the most to other texts. These can be (authentic or pseudo-epigraphic) sermons that have served as important source texts (e.g., AU s 369 or PS-AU s Cai I, 10a), but are more often still *centones* that combine several sources, such as PS-ILD s 7, PS-PET s Liv 5 or AN s Le 1. Many of the nodes at the center of the network represent the 'problematic' or 'unstable' texts that embody the medieval attitude to the patristic source material. So, as an exercise in shifting and broadening one's perspective and departing from traditional scholarly approaches, it is quite successful. Still, we must not overstate this feature. It is still possible that the nature

53 See (respectively) Folliet, "Deux nouveau témoins," 181–82; Leclercq, "Les inédits africains de l'homiliaire de Fleury," 55–56.

of the reference works distorts 'the true picture'. One question to be aware of, for example, is how often cataloguers described a sermon as 'authentic this-or-that' without noticing – or mentioning – that what is in the manuscript is in fact a variation on the authentic text? How much difference between two texts is necessary to no longer identify them as identical in the context of a catalogue?

It is perhaps through the warping of traditional research perspectives that this type of visualization provides a concrete opportunity to further delve into a question that emerges when examining pseudo-epigraphic patristic preaching materials: how do we explain the perceived textual instability of pseudo-epigraphic patristic sermons? At least for the set of sermons in our case study, the observation is that relatively few authenticated sermons intertwine extensively with other materials, while many of the pseudo-epigraphic patristic sermons we have examined exist in multiple versions and as part of complex constellations. This observation, however, requires careful examination. On the one hand, the transmission of sermons in multi-author collections, especially liturgical homiliaries, does seem to promote the malleability and the manipulation of texts. However, on the other hand, the way in which sermons are described in the scholarly reference works magnifies the perceived difference between authenticated and pseudo-epigraphic sermons, since altered versions of authenticated sermons are often not described and referenced as separate texts. We can and should search for contextual and historical factors to explain textual variation, but always in conjunction with an awareness of the biases inherent in the scholarly tradition.

While this version of the network triggers many fascinating questions, there is still something missing before we can truly approach the answers. What we really need is to add a layer on top of what we currently have, a layer that combines the textual data with information about the manuscript transmission of each sermon, its date and place of creation (wherever known), and its genealogical relations with the nodes it is connected to, i.e., whether it is a parent, child, or sibling of the sermons it shares text with. So, for the final part of this article, we want to spend some time exploring one type of further development of the network that would, in our view, significantly increase its usefulness.

4.4 Customized network

One of the most important challenges to overcome for the PASSIM Research Tool as we see it, is to bring together, to truly integrate, what we know about textual connections with what we know about manuscript traditions; in other words, to link the abstract, out-of-context intertextual relationships to information on their material, manuscript contexts. The key premise is that, at the point in time when a 'new' sermon was created, all of its sources must have been physically present in some form or other. The more information we have – about the direction of the *fons-usus* relationship between two sermons and about the manuscripts that transmit them both – the more vivid and detailed such a physical connection can

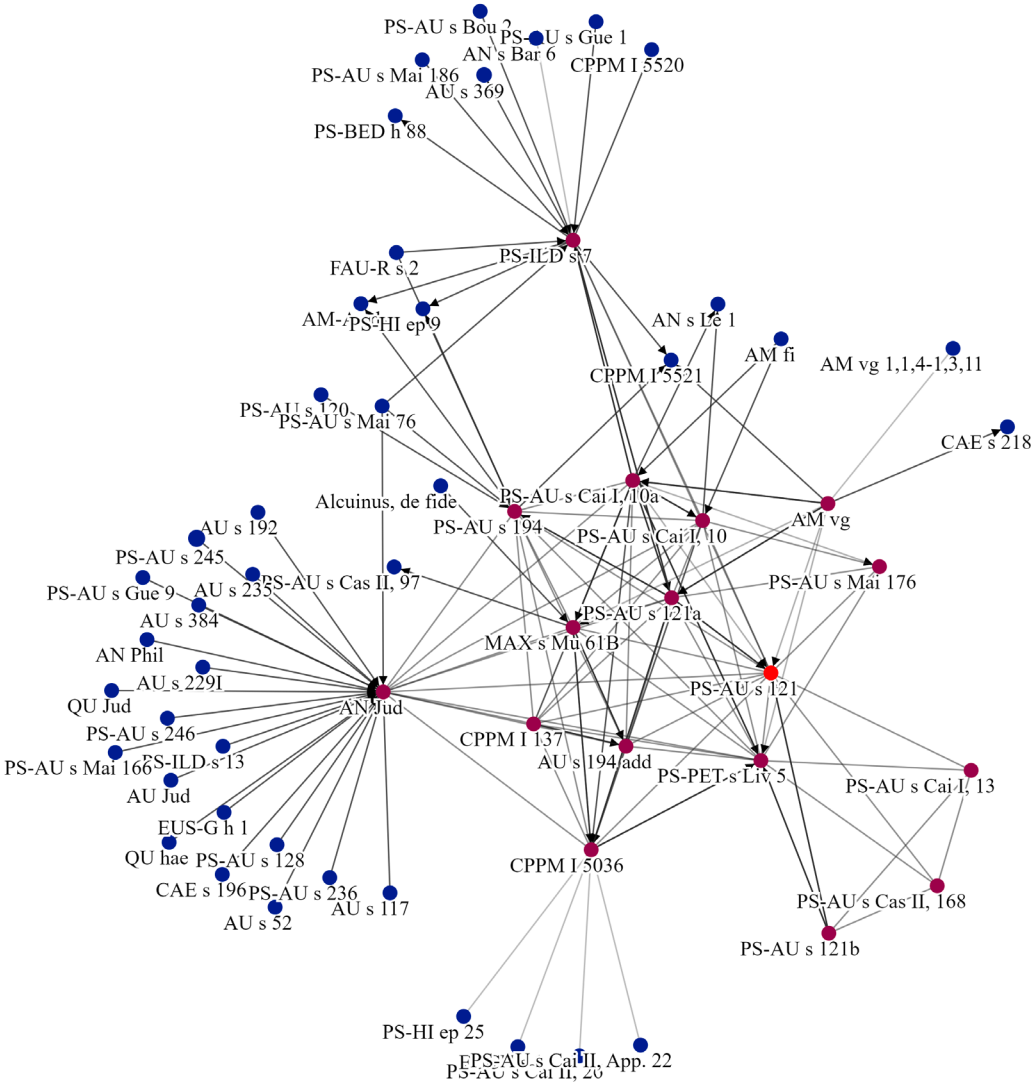

Fig. 7 Directed second-level network of textual connections of PS-AU s 121 (PASSIM).

become. This ostensibly simple principle can be a gateway to bigger topics such as patterns of transmission, the reconstruction of lost collections, or strategies of selection and dissemination. All of these are especially hard to determine for the early medieval period, when manuscript evidence is limited.

To pursue this integration of textual data and manuscript data in the PASSIM visualizations, we are working on several options to customize the networks. Firstly, there is the option to add a direction to the edges, when this is known (Fig. 7). This means that the optional directional link types (cf. Table 1) can be

visualized, either by simply adding arrowheads to the edges wherever possible, or a more sophisticated version, which differentiates between the different types of directional link types.

Second, there is the possibility to add information on the manuscripts that transmit the sermons that are part of the visualization (Fig. 8). In this version of the network, the size of the nodes corresponds to the number of manuscripts in which the sermon is found. The bigger the node, the more extant manuscripts. It is important to emphasize here that the size of the nodes in this particular visualization is, for many nodes, based on an estimation, with differing degrees of uncertainty. For certain sermons, we do have a full grasp of their manuscript transmission, while for others we can make a reliable reconstruction, but again for others we were only able to make an educated guess as the dataset is a work-in-progress and reliant on many different types of resources, some more comprehensive than others. Third, we are developing the option to visualize in which known historical collections a sermon is transmitted.[54] For this customization, the color of the node changes, according to a legend of historical collections represented in the network, which is organized by order of the number of sermons in the network that attaches to each collection. A user can choose to activate all custom options at once or move back and forth between them.

Without jumping to conclusions, this first, experimental network (Fig. 8) shows how a number of blue-coloured nodes in the network, representing texts such as PS-PET s Liv 5, PS-ILD s 7, PS-AU s 194, PS-ILD s 8 or PS-AU s Cas III, 36, reflect reworkings or *centones* preserved in the collection of Alanus of Farfa (and sometimes also in the Roman homiliary that formed its most important source). Even though some of these sermons and their sources have been studied in the past,[55] the network gives insight into the compiled texts in a collection and their potential sources on a larger scale.

As such, this development of our initial network has the potential to help the user in identifying hypotheses for further exploration in several interesting and important areas of research in the field. The network can show which texts were transmitted together in the same collection, and from there whether they potentially had a common source, thus putting the user on track to discover collections

54 The definition of a historical collection is not a straightforward one. Technically, every unique combination of sermons in a manuscript could be termed a collection. We reserve the term in the context of PASSIM for those groups of sermons that (1) were already put together as such in Late Antiquity or the early Middle Ages, or (2) were particularly widespread.

55 Bouhot offers examples of sermons that are thought to be the work of the compiler of the Roman homiliary; Bouhot, "L'homéliaire de Saint-Pierre du Vatican," 105–6, particularly 105, n. 12. See also Ibid., 106, n. 14 for the observation that PS-AU s Cai I, 10a or a text close to it may have been a source for PS-ILD s 7 (AF II, 64) and PS-AU s 194 (AF II, 65).

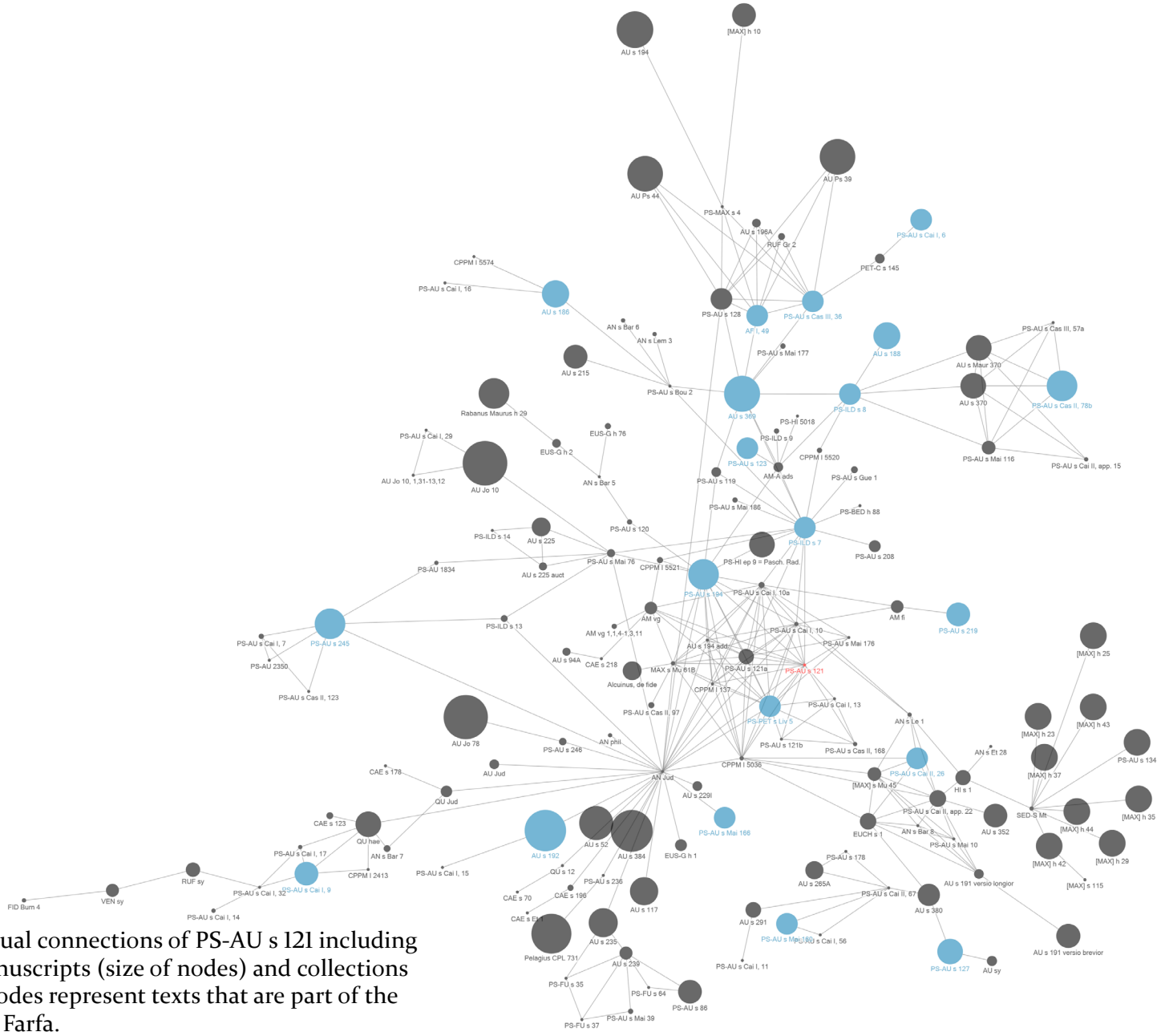

Fig. 8 Network of textual connections of PS-AU s 121 including data on number of manuscripts (size of nodes) and collections (Node XL). The blue nodes represent texts that are part of the homiliary of Alanus of Farfa.

eISSN: 2535-8863
DOI: 10.25517/jhnr.v9i1.133

that may have formed the basis of other collections. This is a line of research that has received considerable attention for the – often single-authored – collections of mostly authentic sermons, but not as much for pseudo-epigraphic or anonymous sermons. The network also draws attention to texts which were, seemingly, especially popular. While it is methodologically questionable to base the popularity of an individual sermon solely on the number of extant manuscripts, the network can also illustrate how dynamic a text was, by showing that it traveled in many different contexts (for instance, in lots of manuscripts that do not belong to the same collection). Furthermore, the network can identify those collections that served as key mediators or sources of inspiration for the rewriting of sermons. It can help differentiate between those collections that were relatively stable and were simply copied as such, and those collections where transmission involved a heavier manipulation of the texts. This can then be taken further into an investigation of the historical context and usage of the respective types of collections. Finally, while the network provides information on compilers we already know about, a next step would be to puzzle together traces of compilers or compilations we are not yet aware of.

However, it is important to keep in mind that these networks must be used not as providers of definitive answers, but as guides for further exploration. There remain several important caveats. The corpus of manuscripts in the PASSIM database is incomplete and thus, for the purposes of this demonstration, we have estimated the number of manuscripts based on our own research and consultation of the available heuristic tools and catalogues. However, this state of incompleteness will persist for a long time, and so the network will have to be used with caution. Aside from this straightforward note of caution, the network in this experimental phase reveals an important related problem. When looking at Fig. 8, it is immediately obvious that the nodes of the authentic texts are overall much bigger than those of the pseudo-epigraphic sermons. We do not think we can interpret this as an accurate reflection of the manuscript tradition. At least in part, this is due to the fact that the authentic sermons are much better studied: there are editions and reference lists, which make it much easier for cataloguers of manuscripts to identify authentic sermons. Thus, while it seems that authentic texts had a much larger dissemination in the Middle Ages, our intuition would be that this is often an effect of the scholarly tradition, and not necessarily the historical reality. This example shows that we cannot take this network at face value without an acute awareness of the imbalances within the material it is based on.

4.5 Methodological issues

Before moving on to the conclusion, we want to briefly mention two dilemmas that we are still trying to sort out.

The first can be summarized concisely as: 'Where to stop?'. Publications and critical editions, especially very recent ones, can go very far in listing potential or probable sources, derivatives, and alternative versions of texts.[56] An overlap of a few words may not even warrant an 'echoes'-link in the database. Then again, it depends where the words are located in the text. The *hodie uerus sol*-incipit, which (with some variation) is found in a number of sermons in our network, does appear to reflect a deliberate choice of the authors or compilers. It is debatable whether the single sentence (*deus factus est homo ut homo …*) that is shared between PS-AU s 128 and several of the sermons that have the *hodie uerus sol*-incipit is of sufficient weight to posit a connection. In these instances, the person working on the data-import has a particular responsibility. While individually, one of these conundrums does not endanger the validity of the dataset or the visualization based on it, an accumulation of such instances, or worse yet, a lack of consistency and uniformity in dealing with them, can eventually have an effect. There is a solution to this and several other issues, which is to work with full text-analysis and quantitatively define the percentage of textual overlap, but that is a challenge we simply cannot tackle within the confines of the current project, and which remains an ambition for the distant future.

The second issue concerns 'ghost' sermons, a term which can cover several different situations. PS-AU s 121 appears to be one of these, namely a sermon that is edited, but which, as far as we know now, may very well be a construction of the editors, since we do not know of any extant manuscript witnesses. It is a legitimate question whether or not there should be an authority file for this sermon in the PASSIM database. We believe it is better that there is, since we cannot be sure about the motives and resources of the seventeenth-century editors who printed it. Still, the scholarly tradition does not treat all such cases equally. In the case of PS-AU s 121, the constituent parts are identified, as we have seen, as PS-AU s 121a and PS-AU s 121b. Another example is AU s 194. In the Lovanienses edition of 1576, this sermon is printed with an addendum, identified with AU s 194 add (= CPPM I 581). However, there is no separate existing reference number for the

56 See for example the recent publications by Clemens Weidmann, "Sermo Mai 10"; Weidmann, "Zwei Weihnachtspredigten des Eucherius von Lyon." While the richness and detail of these studies is certainly a credit to their author, they present challenges when imported into the more rigid context of a database.

sum of both parts as it is printed in the edition.[57] Again, it makes a certain degree of sense to incorporate an authority file for the combination of both texts into the PASSIM database, but should they be considered in the same way as a sermon for which manifestations in manuscripts are known?

While these issues are not 'deal-breakers' and we can negotiate our way around them – eventually, hopefully even solve them – the case study we developed in this article brings them clearly to the fore. In this sense, the visualization of PS-AU s 121 and its textual network forces us to take responsibility for the data and the sources we are working with.

5. Conclusion

As we mentioned at the start of this article, our purpose was to explore the potential usefulness of network visualizations to get more (and different) insights into the textual connections between patristic sermons as recorded in the PASSIM database. In terms of potential added value, we feel the experiment was successful. We can conclude that looking at the data through a visualization indeed succeeds in pointing the user to promising research avenues and case studies beyond the traditional boundaries (and focus areas) of the existing scholarly tradition. We selected PS-AU s 121 because there was already some research indicating that it was part of a complicated tradition. This was certainly true. Assuming a researcher studied the network visualization before tackling the scholarly literature, he or she would have correctly identified the texts that overlap (significantly) with PS-AU s 121 and its parts, but would also have been alerted to the research potential of several other nodes, such as the *cento* CPPM I 5036, the cluster surrounding PS-AU s 128, or AU s 369's apparent use as a source for several other sermons. The complete network also contains several 'linear' connections, which might indicate a progression from one version to the next, depending on further research, namely the stemmatical relationship revealed by a full-text collation.

Furthermore – and this was a more unexpected result – the visualization encourages the user to adopt a different perspective that does away with the prioritization of the original that is prevalent in traditional visualizations of genealogical trees or stemmata. Instead, this visualization draws the eye to the medieval reception, to patterns of sermon circulation and reuse – which is exactly what the PASSIM project wants to promote. Moreover, our effort to encapsulate both textual data and manuscript data in a single visualization may provide interesting

57 According to Gryson, the fragment is "ein unechter Zusatz zu n. 1 in der Löwener Ausgabe von 1576, gedruckt in PL 38, 1015 adn. 4, aus [MAX] s Mu 61B, 1–2;" Gryson, *Répétoire Général,* 241.

insights into the combined and connected workings of physical transmission and textual adaptation.

Though it is an obvious advantage, we want to stress their adaptability as an important strength of the PASSIM dataset and the resulting networks. The case of PS-AU s 121 clearly illustrates how much more flexible a digital environment is to display these connections than, for instance, a printed stemma such as Barré's, or a cross-referenced listing in Machielsen's CPPM.[58] Although the network as displayed here is based on the work of these scholars and others – a debt we fully acknowledge – it may very well be possible that the direction, shape or number of links will change as scholarly research progresses.[59] Scholarship will continue to discover manuscripts or study them for the first time, contributing ever more nodes, edges, versions and links. However, these new discoveries have much less value in isolation than if we can integrate them with what we already know. The very flexibility of the network visualization will allow it to be adapted to new insights on particular sermons, to evolve with new research within the field, and to hopefully also prove useful to scholarship outside of the field of sermon studies.

Of course, it's not all good news. We have also been confronted with some challenging obstacles. It all comes down to a responsible use of the network, which will remain for a long time incomplete, which may have features that invite the user to attribute meaning to certain random effects, and which is subject to distortions produced by its dependence on previous traditions of scholarship. Also, we must not lose sight of the fact that the case study we selected for this experiment consists of the textual tradition of a sermon that has already been the subject of several scholarly inquiries. The manuscript tradition is littered with texts that have not (yet) enjoyed this privilege, but that may be part of networks just as complex as PS-AU s 121. It may be far more difficult to identify intriguing research avenues for the more obscure outliers. However, despite its limits, the existing tradition of heuristic research into the corpus of patristic preaching – both authentic and inauthentic – is truly remarkable, and we must realize the full potential of these tremendous achievements of previous scholarship. Additionally, we have already identified throughout the article several features that might be useful additions to mitigate some of the current defects or to expand the applicability of the network visualizations. In this conclusion, we want to take this 'Where to next?' further still, not by entering into details, but instead by singling out three further developments that could have a significant effect both on the use of networks in philological research and for the field of sermon studies.

58 Barré, "Le sermon pseudo-augustinien App. 121," 137; CPPM I 906.

59 All visualizations in the PASSIM database will therefore be exportable with the inclusion of all necessary metadata, including, for example, search criteria, time stamps, or modifications.

The point of integrating existing scholarly research in a digital context is to provide a new perspective, but also, of course, to save time. Researchers with access to a network based on a sufficiently clean and complete dataset will be able to more quickly and profoundly take their philological research to the next level. Their analysis can focus on how all these alterations and combinations affect the content, the literary qualities, and the theological stance of the texts in question and approach them from a historical or literary background, in addition to one that is purely philological. Although scholars such as Barré have rightly drawn attention to the way in which different versions of a text seem intended to be clearer or more accessible than the texts that may have been their source,[60] or to the theological or doctrinal emphases that can come to light through a detailed comparison of comparable passages,[61] these kinds of questions are rarely the focus of the philological investigations that look into individual sermons or collections. Studies such as Lisa Bailey's excellent work on the Eusebius Gallicanus collection[62] offer tantalizing glimpses of how we might dig deeper into the historical contexts of these texts and the motives and resources of their authors and compilers.

Of course, there remains an elephant in the room, and that elephant consists of the massive corpus of medieval sermons, both in Latin and the vernacular. At the moment, medieval authors are only very sporadically mentioned in the reference works we use as the basis for our dataset, and then usually only those that have famously made extensive use of patristic sources – Raban Maur, Sedulius Scotus and the like. This state of affairs immediately pinpoints a logical further step for the network visualizations and for the PASSIM Research Tool as a whole: to eventually include the connections between patristic and medieval preaching which are, at the moment, mostly overlooked. However, here, the data on which we could base our networks are far less complete still than was the case for late antique sermons as a separate category, so we must think very carefully on when and how to expand the dataset so as not to misdirect the user.

60 Barré analyses how the later MAX s Mu 61 b extr. appears to 'clarify' the similar text of AU s Cai I, 10, 1 (and/or AU s Cai I, 10a, 1); see Barré, "Le sermon pseudo-augustinien App. 121," 120. According to communication between Machielsen and Bouhot, however, this sermon likely also relies on PS-AU s Cai I, 10a; see CPPM I 2072. MAX s Mu 61 b extr. (CPPM I 5759, identical with [MAX] h 5, CPPM I 2072) is printed in Mutzenbecher, *Maximi Episcopi Taurinensis collectionem sermonum antiquam,* 253–55.

61 Several scholars have pointed out the inclusion of several 'Nestorian' elements in PS-AU s 121a that were, apparently, not a barrier for its medieval readers; see Barré, "Le sermon pseudo-augustinien App. 121," 113, n. 17. These sections are conspicuously absent from comparable passages in AU s Cai I, 10; see Barré, "Le sermon pseudo-augustinien App. 121," 117–18.

62 Bailey, *Christianity's Quiet Success.*

Finally, if we can get our corpus sorted, we must also not forget further developments in the field of network visualization and network analysis. The PASSIM project certainly does not claim to be a great innovator here. Rather, we operate with the intention to keep improving small features that will heighten the usefulness of networks wherever we are able. An important priority is to create networks that function as bridges, not end-points. This means enabling the user to click on each node, to see the details of that authority file, and to continue to explore the database from there. We hope to make the networks as dynamic and customizable as possible, so that users can visualize precisely those types of metadata (manuscript origins, attributed authors, historical collections) that are most useful to them. A particular challenge is to visualize not just historical collections, which remain after all, abstract, reconstructed entities, but also to include individual manuscripts and indicate, for instance with a second set of edges, which sermons co-occur within them. In all these pursuits, we remain very conscious of the limitations and ambiguities of network visualizations as tools for analysis.

Eventually, all of this together can help us work towards asking and answering bigger, more all-encompassing questions. We still know relatively little about early medieval traditions of preaching and the role that patristic heritage played in them. Our understanding of how collections traveled from one place to the next, and why certain collections or sermons were prioritized over others, is limited. Medieval copyists of sermons were apparently rather free with authorial attributions, but how and why the attributed author changed from one copy to the next, and what that meant for medieval ideas about authority, remains a mystery. Many such questions can be formulated, and each case study can contribute to the answer. For now, we continue to develop the visualizations in the PASSIM Research Tool, and to build up our dataset, sermon by sermon.

6. References

Andrews, Tara and Caroline Macé, ed. *Analysis of Ancient and Medieval Texts and Manuscripts: Digital Approaches.* Turnhout: Brepols, 2014.

Aschoff, Diethard, ed. *Anonymi contra Ivdaeos* (Corpus Christianorum. Series Latina 58B). Turnhout: Brepols, 2009.

— "Studien zu zwei anonymen Kompilationen der Spatantike. Anonymi contra Philosophos et Contra Judaeos." *Sacris Erudiri* 27 (1984): 37–127.

Bailey, Lisa. *Christianity's Quiet Success. The Eusebius Gallicanus Sermon Collection and the Power of the Church in Late Antique Gaul.* Notre Dame: University of Notre Dame Press, 2010.

Barré, Henri. "Le sermon pseudo-augustinien App. 121." *Revue d'Etudes Augustiniennes et Patristiques* 9, no. 1–2 (1963): 111–37.

Boodts, Shari. "Navigating the Vast Tradition of St. Augustine's Sermons. Old Instruments and New Approaches." *Augustiniana* 69, no. 1 (2019): 83–115. http://dx.doi.org/10.2143/AUG.69.1.3286703.

— "The manuscript transmission of the *Quinquaginta homiliae*. A stemmatical analysis with a new critical edition of Augustine's *sermo* 168." *Revue Bénédictine* 123, no. 2 (2013): 214–47.

Boodts, Shari, et al. *PASSIM Research Tool,* beta-version. Radboud Institute for Culture and History, 2021. Last accessed 24 July 2023. https://passim.rich.ru.nl/.

Bouhot, Jean-Paul. "L'homéliaire de Saint-Pierre du Vatican au milieu du VIIe siècle et sa postérité." *Recherches Augustiniennes et Patristiques* 20 (1985): 87–115.

Caillau, Armand Benjamin, ed. *Collectio selecta SS. ecclesiae Patrum complectens exquisitissima opera tum dogmatica et moralia, tum apologetica et oratoria,* Vol. 130. Paris: Méquignon-Havard, 1837.

Casavecchia, Roberta, et al. *Omeliari in scrittura beneventana.* http://omeliari.unicas.it/. Last accessed 24 July 2023.

Cunningham, Mary and Pauline Allen, ed. *Preacher and Audience. Studies in Early Christian and Byzantine Homiletics.* Leiden-Boston: Brill, 1998.

De Maeyer, Nicolas and Gert Partoens. "Preaching in Sixth-Century Arles. The Sermons of Bishop Caesarius," in *Preaching in the Patristic Era. Sermons, Preachers, and Audiences in the Latin West,* edited by Anthony Dupont, Shari Boodts, Gert Partoens and Johan Leemans, 198–231. Leiden-Boston: Brill, 2018.

Diesenberger, Maximilian, Yitzhak Hen, and Marianne Pollheimer, ed. *Sermo doctorum: Compilers, Preachers, and Their Audiences in the Early Medieval West.* Turnhout: Brepols, 2013.

Dolbeau, François. "La Transmission de la Prédication Antique de Langue Latine," in *Preaching in the Patristic Era. Sermons, Preachers, and Audiences in the Latin West,* edited by Anthony Dupont, Shari Boodts, Gert Partoens and Johan Leemans, 31–58. Leiden-Boston: Brill, 2018.

— "Un sermon pseudo-augustinien pour la fête de Pâcques, confronté à ses sources," in *Latin Anonymous Sermons from Late Antiquity and the Early Middle Ages (AD 300–800). Classification, Transmission, Dating,* edited by Matthieu Pignot, 111–24. Turnhout: Brepols, 2021.

Dupont, Anthony, Shari Boodts, Gert Partoens and Johan Leemans, *Preaching in the Patristic Era. Sermons, Preachers, and Audiences in the Latin West.* Leiden-Boston: Brill, 2018.

Echard, Siân and Stephen Partridge. *The Book Unbound: Editing and Reading Medieval Manuscripts and Texts.* Toronto: University of Toronto Press, 2004.

Folliet, Georges. "Deux nouveaux témoins du Sermonnaire carolingien récemment reconstitué." *Revue d'Etudes Augustiniennes et Patristiques* 23, no. 1–2 (1977): 155–98.

Grégoire, Réginald. *Homéliaires liturgiques médiévaux: Analyse de manuscrits.* Spoleto: Centro italiano di studi sull'alto medioevo, 1980.

Gryson, Roger. *Répertoire général des auteurs ecclésiastiques latins de l'antiquité et du haut Moyen Âge. 5e edition mise à jour du Verzeichnis der Sigel für Kirchenschriftsteller commence par B. Fischer, continué par H. J. Frede.* Freiburg: Herder, 2007.

Hamidović, David, Claire Clivaz, and Sarah Bowen Savant. *Ancient Manuscripts in Digital Culture. Visualisation, Data Mining, Communication.* Leiden-Boston: Brill, 2019.

Lambot, Cyril. "La tradition manuscrite des Sermons de saint Augustin pour la Noël et l'Épiphanie." *Revue Bénédictine* 77, no. 3–4 (1967): 217–45.

Leclercq, Jean. "Les inédits africains de l'homiliaire de Fleury." *Revue Bénédictine* 58, no. 1–4 (1948): 53–72.

van Lit, L. W. Cornelis. *Among Digitized Manuscripts. Philology, Codicology, Paleography in a Digital World.* Leiden-Boston: Brill, 2020.

Liverani, Francesco, ed. *Spicilegium Liberianum.* Vol. 1. Florence: Cambiagi, 1863.

Machielsen, Johannes. *Clavis patristica pseudoepigraphorum medii aevi.* Vol. 1, *Opera homiletica* (Corpus Christianorum. Series Latina). Turnhout: Brepols, 1990.

— "Contribution à l'étude de la pseudépigraphie médiévale en matière patristique. Problèmes d'attribution et de remaniement des textes homilétiques," in *Fälschungen im Mittelalter. Internationaler Kongreß der Monumenta Germaniae Historica. München, 16.–19. September 1986, Teil 5: Fingierte Briefe, Frömmigkeit und Fälschung, Realienfälschungen,* edited by Detlev Jasper, 345–59. Hannover: Hansche, 1988.

Mai, Angelo, ed. *Nova Patrum Bibliotheca.* Vol. 1. Rome: Typis Sacri Consilii Propagando Christiano Nomini, 1852.

Manus Online: Manoscritti delle biblioteche italiane. Last accessed 24 July 2023. https://manus.iccu.sbn.it/.

Martin, Michael. "The Italian Homiliary: An example *pro omnibus bonis operibus* produced according to the 'new' Carolingian homiletic genre and reform measures." *Sacris Erudiri* 48 (2009), 261–338.

Migne, Jean-Paul and Adalbert-Gautier Hamman, ed. *Patrologiae cursus completus. Supplementum.* Vol. II–III. Paris: Garnier Frères, 1960–1963.

Migne, Jean-Paul, ed. *Patrologiae cursus completus. Series Latina.* Vol. 39. Paris: s.n., 1865.

Mutzenbecher, Almut, ed. *Maximi Episcopi Taurinensis collectionem sermonum antiquam nonnulis sermonibus extravagantibus adiectis.* (Corpus Christianorum. Series Latina 23). Turnhout: Brepols, 1962.

Petschenig, Michael, ed. *Cassianus, De Institutis coenobiorum, De incarnatione contra Nestorium.* (Corpus Scriptorum Ecclesiasticorum Latinorum 17). Vienna: Österreichisen Akademie der Wissenschaften, 2004.

Pignot, Matthieu, ed. *Latin Anonymous Sermons from Late Antiquity and the Early Middle Ages (AD 300–800). Classification, Transmission, Dating.* Turnhout: Brepols, 2021.

Pratt, Karen, Bart Besamusca, Matthias Meyer, and Ad Putter, ed. *The Dynamics of the Medieval Manuscript. Text Collections from a European Perspective.* Göttingen: V&R unipress, 2017.

Roelli, Philipp, ed. *Handbook of Stemmatology. History, Methodology, Digital Approaches.* Berlin: De Gruyter, 2020.

Sharpe, Richard. *English Benedictine Libraries: The Shorter Catalogues.* London: British Library, 1996.

Steinova, Evina. "The Oldest Manuscript Tradition of the *Etymologiae* (Eighty Years After A. E. Anspach)." *Visigothic Symposia* 4 (2020): 100–43.

— "Two Carolingian Redactions of Isidore's *Etymologiae* from St. Gallen." *Mittellateinisches Jahrbuch* 56, no. 2 (2021): 298–376.

Tarrant, Richard. *Texts, Editors, and Readers. Methods and Problems in Latin Textual Criticism.* Cambridge: Cambridge University Press, 2016.

Weidmann, Clemens. "Der Augustinuscento *Sermo Mai 66:* Mit einem textkritischen Anhang zu Predigten auf Perpetua und Felicitas," in *Sermo doctorum: Compilers, Preachers, and Their Audiences in the Early Medieval West,* edited by Maximilian Diesenberger, Yitzhak Hen and Marianne Pollheimer, 59–79. Turnhout: Brepols, 2013.

— "Discovering Augustine's Words in Pseudo-Augustinian Sermons," in *Tractatio Scripturarum. Philological, exegetical, rhetorical, and theological studies on Augustine's sermons,* edited by Anthony Dupont, Gert Partoens and Mathijs Lamberigts, 41–58. Turnhout: Brepols, 2012.

— "Sermo Mai 10: Eine authentische Predigt des Augustinus?" *Augustiniana* 60, no. 3/4 (2010): 173–92.

— "Zwei Weihnachtspredigten des Eucherius von Lyon," in *Edition und Erforschung lateinischer patristischer Texte: 150 Jahre CSEL,* edited by Victoria Zimmerl-Panagl, Lukas J. Dorfbauer and Clemens Weidmann, 111–38. Berlin: De Gruyter, 2014.

Wilmart, André. "Easter Sermons of St. Augustine: Some New Texts." *The Journal of Theological Studies* 27 (1926): 337–56.

Witt, Jeffrey, Michael Stenskjær Christensen and Ueli Zahnd. "Re-Conceiving the Christian Scholastic Corpus with the Scholastic Commentaries and Texts Archive," in *Digital Humanities and Christianity. An Introduction,* edited by Timothy Hutchings and Claire Clivaz, 47–76. Berlin: De Gruyter, 2021.

The Journal of HISTORICAL NETWORK RESEARCH

EVINA STEIN

Parallel Glosses, Shared Glosses, and Gloss Clustering

Can Network-Based Approach Help Us to Understand Organic Corpora of Glosses?

Journal of Historical Network Research 9 (2023) 36–100

Keywords co-occurrence networks, medieval Latin glosses, manuscript studies, gloss parallelism, shared manuscript transmission

Abstract Glossing was an important element of medieval western manuscript culture. However, glosses are notoriously difficult to analyze because of their triviality, fluid nature, heterogeneity of origin, complex transmission histories, and anonymity. Traditional scholarly approaches such as close reading and the genealogical method often do not produce satisfactory results, especially in the case of gloss corpora that are highly organic, i.e., display the traits listed above to a significant degree. This article outlines a method for analyzing the organic corpora of glosses based on their treatment as networks. The theoretical model for the proposed method is the co-occurrence network, a network model in which relationships between entities (nodes) are established based on certain shared properties or constituent elements (edges). In the case of corpora of glosses, glossed manuscripts are assumed as nodes, and the glosses that specific manuscripts have in common constitute the edges between them. Since gloss parallelism can arise through different processes, including randomness, the article describes two strategies that reduce such noise so that the transmission of glosses can be effectively examined. The method is demonstrated on a representative corpus – the early medieval glosses to the first book of the *Etymologiae* of Isidore of Seville.

C²DH LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

UNIVERSITY OF LUXEMBOURG

I. Introduction

Glossing (a term used in this article interchangeably with annotation) represented an important aspect of many pre-modern written cultures, including in Europe before the advent of print. In the medieval Latin-writing world, handwritten texts, copied in the writing block of the manuscripts (the black space), were commonly equipped in the margins, between the lines and columns and in other spaces left blank on the page (the white space) with enriching information – commentary, explanatory vocabulary, grammatical and stylistic remarks, translation to other languages, diagrams, cross-references, and critical remarks about the text's quality and veracity (Fig. 1).[1]

Medieval western annotation has traditionally been of interest to various scholars: linguists, who have found it a valuable source of information on the development of various European languages; philologists, who edited and analyzed the most important commentaries, glossing traditions and glossaries in manners similar to how they treated other historical and literary sources; and historians of intellectual life, who studied specific commentaries, glossing traditions and glossaries in their historical and social contexts. Only in the last decades have we seen a growing interest in medieval western annotation as a phenomenon in its own right, its use as a source for the understanding of medieval western culture more broadly, and its consideration in the global context of annotation cultures of other regions and periods. As a result of this broadening of horizons, a development that owes much to the increasing permeation of Big Data approaches to Humanities, the advent of digitization and computer-assisted methods, and the re-envisaging of annotations as data rather than as a traditional historical, literary or linguistic source, it has been recognized that medieval western annotations could contribute to research questions for which they had not been traditionally exploited.

In this article, I explore such a novel direction in research, looking at how the study of medieval western annotations could benefit from the application of a network-based approach. I hope to demonstrate how this approach can open new avenues to answer long-standing questions about glossing and provide us

Acknowledgements: Data used in this article were gathered as part of the *Innovating Knowledge* project funded by the Dutch Research Organization (NWO) under the grant agreement VENI 275-50-016 in 2018–2021. The author of this study would like to thank Siamak Taati and Sara Najem for their valuable advice about network theory, Peter Boot for helping with various aspects of data processing, and Bernhard Bauer for his comments on an earlier draft of the article.

Corresponding author: Evina Stein, Huygens Institute, Dutch Academy of Arts and Sciences (KNAW), Amsterdam. evina.steinova@gmail.com.

1 For an introduction to medieval western annotation, see Holtz, "Glosse e commenti"; Tura, "Essai sur les *marginalia*"; Schiegg, *Frühmittelalterliche Glossen*.

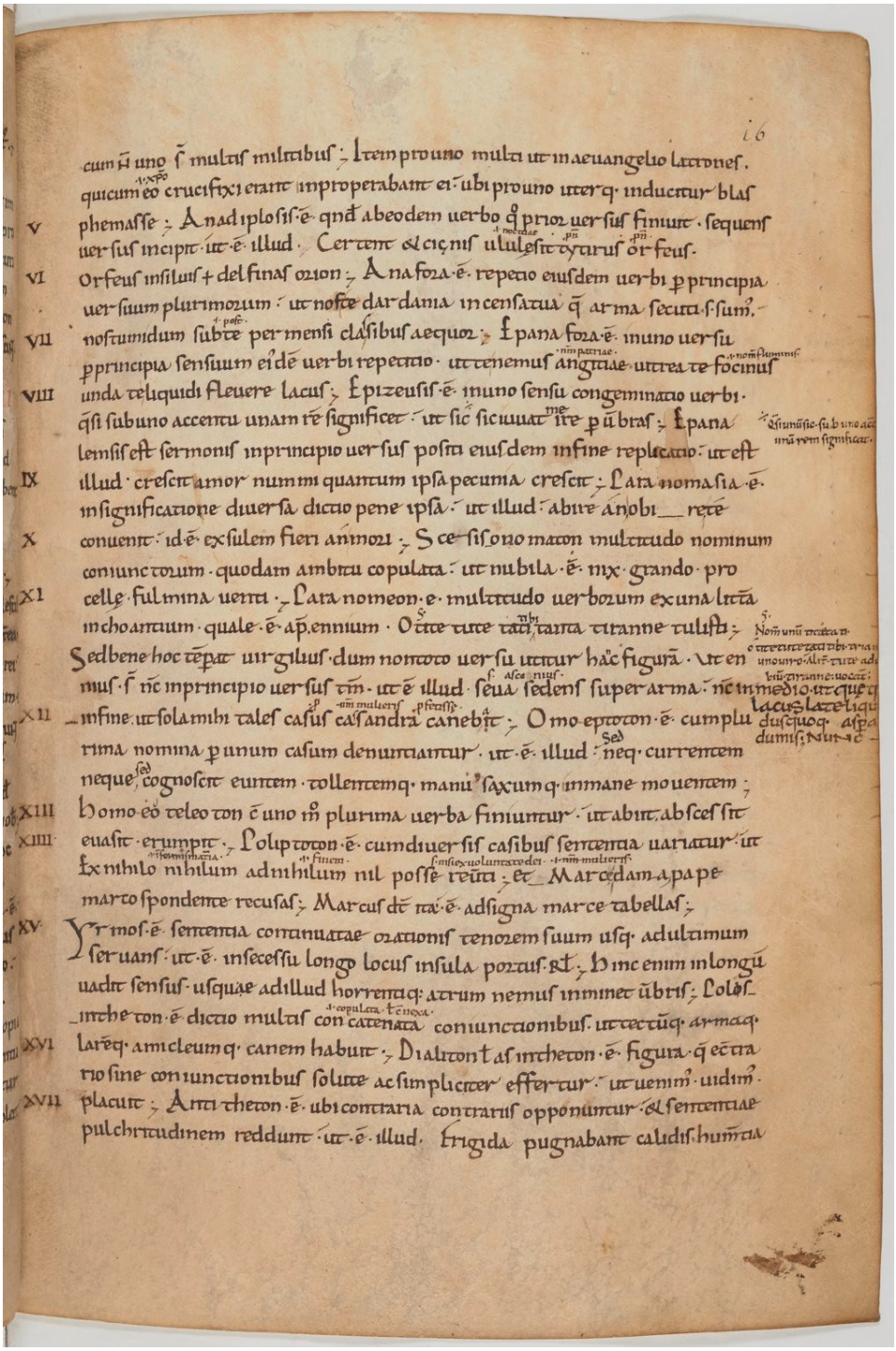

Fig. 1 An image of a glossed medieval western manuscript, Paris, Bibliothèque nationale de France, Latin 7585, fol. 16r. Source: Gallica, at https://gallica.bnf.fr/ark:/12148/btv1b10542288m/f35.item.r=%22Latin%207585%22.

with a means to overcome some of the well-known problems that scholars typically face. The new avenues relate to the how's, why's and what's of glossing: How were glosses produced and transmitted in the medieval Latin-writing world? Were they usually a result of spontaneous inspiration in response to momentary stimuli? Or were they rather handed down, exchanged, and collected?[2] In this regard, what is the significance of the same glosses recurring in different annotated manuscripts? How common is such gloss parallelism, and to what extent does it reflect transmission, as opposed to other historical processes, or chance? What does it suggest about the circulation of glosses in the medieval Latin-writing world? To tackle these questions, we can look at the patterns of gloss parallelism between manuscripts to establish their mutual relationships and examine the properties of networks constructed in this manner. The networks we can construct using this principle capture specific, intrinsic, dynamic aspects of glossing that are difficult to examine by other means, especially as extrinsic evidence about the production and circulation of glosses is often scarce or non-existent.[3]

As for the well-known problems in the study of medieval glossing, scholars need to tackle what may be termed their triviality. Glosses often amount to no more than a single phrase, word or even a syllable or a letter.[4] As a consequence, we cannot assume that multiple occurrences of a trivial gloss signal transmission, as would be the case with non-trivial glosses.[5] Moreover, the collections in which glosses typically survive are notoriously flexible and fluid, lacking the degree of coherence and sequentiality that define a typical text.[6] As a result, scholars may

2 Lapidge, "The Evidence of Latin Glosses"; Wieland, "The Glossed Manuscript"; Teeuwen, "Marginal Scholarship: Rethinking the Function of Latin Glosses in Early Medieval Manuscripts"; Teeuwen, "Writing in the Blank Space of Manuscripts."

3 This network-based approach is partially inspired by earlier attempts at employing network visualization to express the relationship between annotated manuscripts and explore gloss parallelism by Bernhard Bauer; see Bauer, "The interconnections of St Gall, Stiftsbibliothek, MS 251 with the Celtic Bede manuscripts"; Bauer, "The Celtic Parallel Glosses on Bede's 'De Natura Rerum'"; especially Bauer, "Venezia, Biblioteca Marciana, Zanetti Lat. 349. An Isolated Manuscript ?"

4 In the demonstrative corpus introduced in section 3, for example, the average length of a gloss is 2.6 words, and 45% of glosses are constituted by a single word. If opening formulas common to Latin glosses (see footnote 13) are discounted, the average length of a gloss in this corpus drops to 2.2 words, and 60% of glosses are constituted by a single word. See also Wieland, *The Latin Glosses on Arator and Prudentius in Cambridge University Library, MS Gg. 5.35,* 8; Nievergelt, "Glossen aus einem einzigen Buchstaben."

5 This observation is a variation on the well-known principle of indicative errors in genealogical textual criticism, which is explained in Chiesa, "The Genealogical Method: Principles and Practice," 79–80; Palumbo, "The Genealogical Method: Criticism and Controversy," 102–5.

6 See Teeuwen, "The Impossible Task of Editing a Ninth-Century Commentary," 197–200; Teeuwen, "Writing in the Blank Space of Manuscripts." They can be described as text colonies, using the terminology devised by the linguist Michael Hoey; Hoey, *Textual Interaction,* 74–76.

miss important connections within a corpus of glosses. Both traits of glossing pose a serious challenge in so far as we want to examine them with scholarly methods of traditional textual and historical scholarship (e.g., close reading).[7] As a result, certain types of glosses tend to be overlooked and understudied, while overconfident application of these methods may lead to inaccurate or misleading conclusions about other kinds of material. However, provided that we observe specific precautions, networks can be constructed even by relying on trivial glosses that do not form well-defined sequences. The network approach can, therefore, bypass some of the stumbling blocks of the research of medieval western glossing.

The questions articulated in this introduction are not fully resolved in this article. Instead, the present contribution aims to outline a particular methodology for analyzing medieval western glosses, demonstrate its utility on a representative corpus of glosses, and provide examples of network-driven analysis that can answer specific research questions. This article is, thus, primarily an invitation to: a) further develop a network-based approach to the study of glossing; b) apply it to different corpora of material; and c) test its usefulness. The proposed method is purposefully presented with a minimum of mathematical formalism and coding so that it is as accessible as possible to humanities scholars.

The article is divided into eight sections. The three sections following this introduction (section 1) provide the essential background for the network-based approach to annotated manuscripts. Section 2 defines concepts essential for the network-based approach to glossing and data pre-processing. Section 3 introduces a dataset used in this article to demonstrate this approach on a real-world corpus of glosses from medieval Europe. Section 4 outlines the general method used to construct a specific kind of network, namely the co-occurrence network, which can be used to harness gloss parallelism for research purposes. Sections 5 and 6 represent the analytical core of this article. In section 5, I describe and analyze several co-occurrence networks constructed from the data provided in section 3. In section 6, a selected network from section 5 is visualized and inspected in the light of extrinsic evidence. Finally, section 7 addresses the potential and limitations of the method, and section 8 presents the most important conclusions of this study.

2. Concepts and definitions

In this article, the term 'collection of annotations/glosses' is applied to manuscripts (e.g., a collection of annotations in Leiden VLF 48 or St. Gallen 904), while the term 'corpus of annotations/glosses' is used in connection to texts (e.g., a cor-

7 Other problems posed by glosses are described in O'Sullivan, "Problems in Editing Glosses: A Case Study of Carolingian Glosses on Martianus Capella."

pus of glosses to the Psalms, Virgil's *Aeneid,* or Priscian's *Institutiones*). A corpus of annotations represents all known glosses to a specific text. It is typically assembled from multiple collections of annotations found in manuscripts – its witnesses. Every gloss has two elements: a lemma (pl. lemmata), which corresponds to a specific word, phrase, or other textual unit in the black space, anchoring the gloss in a substrate (text or manuscript); and a body, in which enriching information is provided, usually in the white space. When the term gloss is used below it designates both elements, or the body of a gloss if it is explicitly distinguished from a lemma.

A gloss is the basic building block of a collection and a corpus of glosses, which may count tens, hundreds, or thousands of annotations. Although each gloss may be considered a micro-text as far as it is textually self-sufficient and can be added, removed, altered, and its position changed, glosses in a manuscript collection appear in a chain, i.e., we can state which gloss precedes or follows another and order them based on the sequence of folia and lines. However, without the support of a manuscript substrate, for example, if we compare glosses from different manuscripts or constitute a corpus, they disassemble into an unordered pool. The sequence of glosses in a collection may be relevant for certain types of research. However, the method described below is insensitive to it. The corpus and collections of glosses are, therefore, treated as pools, i.e., when a particular collection of annotations is discussed, the order in which glosses appear within it is taken into consideration only to a minimal degree.[8]

2.1 Systematic versus organic glossing

Based on the character of annotations in a collection or a corpus, it is useful to distinguish systematic from organic glossing. Systematic glossing can be defined as a programmatic annotation carried by one agent (a single individual or a group) with the intention of coherently engaging with a specific text, often meaning glossing it in its entirety, and therefore extensively.[9] Scholars frequently use

8 The 'minimal degree' applies here to glosses to identical lemmata appearing in different chapters of the annotated text, i.e., further apart than glosses to identical lemmata appearing within a single chapter. In theory, a researcher can encounter the same lemma-gloss pair in different chapters, as the text could contain the same words in multiple chapters, and these could attract the same glosses. However, these are not considered instances of gloss parallelism in this study.

9 Examples of medieval systematic glossing in the Latin-writing world include the ninth-century commentary on Martianus Capella's *De nuptiis Philologiae et Mercurii* by John the Scot, the twelfth-century commentaries of Anselm of Laon on the Gospel of John, and the coeval commentary on the Code of Justinian by the jurist Accursius. On these commentaries, see Jeauneau, "Le Commentaire érigénien sur Martianus Capella (De nuptiis, Book I)"; Rossi, *Atti del Convegno internazionale di studi Accursiani;* Andrée, "Anselm of Laon Unveiled."

terms such as commentaries, commentary traditions, and *scholia* to recognize the systematic nature of certain corpora of glosses and attribute them to specific individuals (authors) or groups (circles).[10] However, glosses were often inserted into manuscripts in a non-systematic manner, on an *ad hoc* basis, and perhaps even spontaneously, responding to the immediate needs and concerns of their makers rather than reflecting a program. As such, they do not form coherent collections nor provide a structured exposition, come from many different contexts of origin, and resulted from uncontrolled accumulation or growth not overseen by a specific agent. To distinguish these from the systematic collections of glosses and commentaries, I call them organic. A corpus or collection of annotations may possess a mixed set of traits, as it arose through both organic growth and systematic composition and compilation. While we can thus employ the designations systematic or organic for certain corpora of glosses that display very clear traits of one or the other type, it is more accurate to talk about the extent of organicity or systematicity in a corpus or a collection of glosses.

Traditionally, scholars have paid more attention to systematic than organic glosses due to the former's greater prominence among source material, their perceived higher aesthetic, literary and historical value, because these corpora better fitted the traditional notions of textuality and authorship, and due to the suitability of traditional approaches (e.g., genealogical editing and close reading) for their analysis. However, organic glossing may have been more prevalent in medieval Europe and thus more characteristic of medieval western annotation practices, particularly during certain periods. While the network-based approach may produce relevant results when applied to highly systematic corpora of glosses, these corpora tend to respond well to traditional methods, and thus the network-based approach may serve as a useful complement to these methods, although it is unlikely to be a scholar's primary option. This article is rather concerned with glossing that is organic to such a degree that its lack of coherence, heterogeneity of origin, purpose and language, multilayered character, and fluidity allow for a limited deployment of traditional methods.

2.2 Isolated, parallel and shared glosses

For the purposes of the network-based approach, a gloss corpus consists of two types of glosses. Certain glosses appear in it only once. I shall call these isolated glosses, and deal with them only marginally since they do not allow us to postulate a relationship between manuscripts. Other glosses feature in a corpus more than once since they appear in several of its manuscript witnesses. For example, the gloss *significat* is attached to the lemma *pingit* in two manuscripts of Bede's *De temporum ratione,* studied by Pierre-Yves Lambert and Bernhard Bauer

10 On this terminology, see, for example, Teeuwen, "Writing in the Blank Space of Manuscripts," 13.

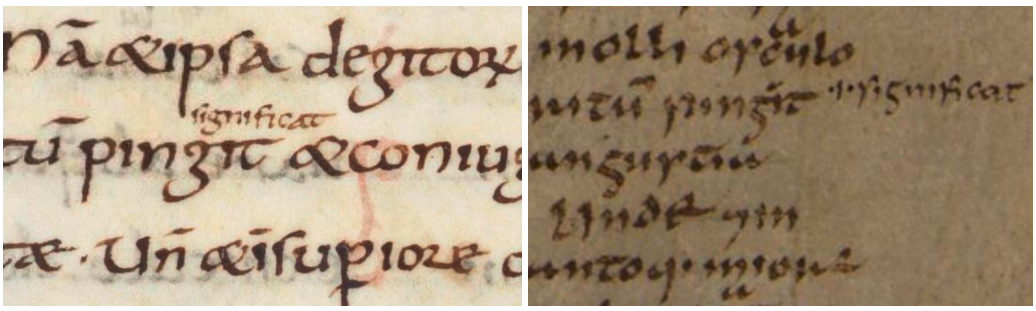

Fig. 2 A parallel gloss *pingit.significat* in two manuscripts of Bede's *De temporum ratione.* Left: Angers, Bibliothèque municipale, 477, fol. 45v (source: BVMM, at https://bvmm.irht.cnrs.fr/iiif/1097/canvas/canvas-375525/view). Right: Karlsruhe, Badische Landesbibliothek, Aug. Perg. 167, fol. 24r (source: Badische Landesbibliothek, at https://digital.blb-karlsruhe.de/id/20736).

(Fig. 2).[11] Such glosses establish a relationship between manuscripts that is at the heart of co-occurrence networks. I shall call these parallel glosses.[12]

Importantly, labelling a gloss as parallel does not provide an explanation as to why it recurs within a corpus. The term merely signals that due to the extent of their philological similarity, two or more glosses are judged as manifestations of the same philological entity.[13] Parallel glosses may be parallel due to transmission, that is, they reflect a relationship between manuscripts that implies contact between historical individuals, groups, and institutions. However, the parallelism illustrated by Fig. 2 could also result from processes other than transmission, especially if it concerns individual trivial glosses. Specifically, a trivial gloss that represents a logically derived explanation of a lemma (e.g., a synonym, a translation, or an etymology) could have been coined independently by multiple

11 Lambert, "Les commentaires celtiques a Bède le vénérable"; Lambert, "Les commentaires celtiques à Bède le vénérable"; Bauer, "The interconnections of St Gall, Stiftsbibliothek, MS 251 with the Celtic Bede manuscripts," 34.

12 I borrow this term from Bauer, "Venezia, Biblioteca Marciana, Zanetti Lat. 349. An Isolated Manuscript ?," 91.

13 The issue of philological similarity would deserve theoretical reflection, which is not possible in this article. Here, it can be noted that traditional textual scholarship also operates with a notion of similarity in assessing indicative errors, variant readings, and text versions as the same or similar. When considering glosses as parallel in this study, I ignore spelling variation, manner of abbreviation, morphological form (e.g., whether the gloss accepts the case, number, etc. of the lemma or not), word order, the presence of introductory phrases characteristic to Latin glosses that do not affect the meaning of a gloss (e.g., *id est, hoc est, scilicet, sicut, quasi, vel*), textual corruptions as a result of a mechanical error in a single witness, and omission of or variation in minor elements that do not alter the meaning of a gloss (e.g., prepositions and prefixes).

annotators, since the lexicon of a language provided them with limited options, they were glossing the same text and therefore the same lemmata, and they likely received similar training and had similar resources at their disposal. Furthermore, in a scenario described in this article, gloss parallelism can be expected to occur to an extent even randomly, not mirroring any historical process but rather arising from the method itself.

Since gloss parallelism is central to the method described in this article, it is essential to distinguish parallel glosses that reflect transmission from those that are the result of what I shall call spontaneous composition and random gloss parallelism. For this reason, I introduce a specifying category: the shared gloss. A shared gloss can be defined as a subtype of parallel glosses, the similarity of which can be explained as a consequence of transmission. Distinguishing parallel glosses from shared glosses is difficult in a real-life research context, especially in organic corpora of glosses, for the transmission of which we typically lack sufficient extrinsic evidence. A researcher can, nevertheless, establish philological criteria to assess parallel glosses as shared. In this study, I use a system of three ranks, following the principle that the more particular a gloss is, the less likely it is that it arose independently multiple times.[14] Beyond a certain degree of peculiarity, a scholar can consider a gloss monogenetic, i.e., having originated only once, and therefore assume that all its manifestations in a gloss corpus are instances of transmission. By contrast, the more generic information a gloss provides, the more likely that it is polygenetic, i.e., having originated independently multiple times, and therefore cannot be assumed to have been evidence for transmission.[15] In this particularity ranking, the lowest rank, 1, is accorded to generic parallel glosses that are possibly polygenetic, and are thus treated as instances of spontaneous composition in the following sections.[16] The intermediate rank, 2, is used

14 The following general criteria are used in this study to assess to what degree a parallel gloss is shared:
a) number of identical words appearing in the same or similar sequence in a gloss (here at least four);
b) the presence of the gloss in a significant number of witnesses (here at least five);
c) the gloss is a citation from a known source;
d) the presence of idiosyncratic, unusual, or erroneous information;
e) the presence of textual errors, corruptions, or paleographic features that are indicative of copying;
f) the gloss depends on an error in the substrate text but also appears in witnesses without this error;
g) if multiple parallel glosses form logically coherent sets within the text; and
h) in the case of lemmata that attracted many different isolated glosses if gloss parallelism is observed.

15 Monogenicity and polygenicity are discussed in Trovato, "Neo-Lachmannism: A New Synthesis?"; Conti, "A Typology of Variation and Error," 243–45.

16 A common example of glosses ranked 1 are glosses that expand an obvious ellipsis in the text. In the corpus introduced in section 3, for example, the lemma *Hebraeorum lit-*

for glosses that cannot be determined by philological assessment, i.e., they may have been transmitted, but it cannot be ruled out that they emerged as a result of spontaneous composition.[17] Finally, the highest rank, 3, is assigned to glosses so particular that they can be treated as transmitted.[18] For the most part, I shall use the term parallel gloss in the following sections of this study, but if the term shared gloss is used, it refers specifically to glosses that are assumed to have been transmitted, i.e., those with rank 3. Using this ranking method, rather than classifying parallel glosses binarily as shared or not, will allow us to account for the complexity of the real-world data introduced in the following section and consider different scenarios of gloss parallelism and transmission.

2.3 Gloss sets and gloss clusters

As glosses behave as self-sufficient micro-texts, we can expect to encounter some that circulated individually (as we shall see, this is the case with some shared glosses analyzed below). In practice, however, it is more common to encounter glosses preserved in particular groups of manuscript witnesses as sets, i.e., to observe that certain parallel glosses always travel together as a unit. As with the case of parallel glosses, dissecting a gloss corpus into sets does not explain why we encounter them today in this form. Some of the gloss clustering in a corpus is due to transmission, as the sets correspond to textual units circulating in the Middle Ages.[19] However, a degree of clustering is also a natural result of gloss parallelism due to spontaneous composition and even randomness.

For this reason, I distinguish gloss sets (any batches of glosses that appear together in two or more witnesses in the corpus), from gloss clusters (sets that can

---

teras a Lege coepisse per Moysen ("The Hebrew letters [are believed] to have begun from the Law through Moses") in the third chapter of the first book of the *Etymologiae* is glossed with *dicimus* ("we claim") in two manuscripts.

17 A common example of glosses ranked 2 are synonyms and glosses that provide non-specific clarifying information about the name of a person or place, or the grammatical category of the lemma. In the corpus introduced in section 3, for example, the lemma *repertus* ("found") is glossed as *inventus* ("discovered") and the name of the mythological king Cadmus is glossed as *rex* ("a king") in the third chapter of the first book of the *Etymologiae*.

18 An excellent example of a gloss from the corpus introduced in section 3 assigned rank 3 is a gloss to the term *sicilicus,* a special orthographic sign used to mark the duplication of letters in Latin, in chapter 27 of the first book of the *Etymologiae*. This gloss reads: *et sicilicus quia in Sicilia inveniebatur primo* ("and it is called a *sicilicus* because it was first invented in Sicily"). As the name of *sicilicus* is derived from *sicilis* ("a sickle") rather than related to the island of Sicily, this imaginative etymologization, found in three manuscripts, should be consider highly peculiar and therefore monogenetic.

19 We know from the extrinsic evidence that medieval scribes usually copied glosses from manuscript to manuscript in batches; see Dionisotti, "On the Nature and Transmission of Latin Glossaries"; Godden and Jayatilaka, "Counting the Heads of the Hydra," 365.

be assumed to reflect gloss transmission in the Middle Ages). In this study, I define a gloss cluster as a set constituted by at least ten glosses, or only by glosses with rank 3. Importantly, unless we possess historical evidence that would allow us to reconstruct a specific historical unit of transmission fully, which is rarely the case, it is a scholarly reconstruction. In the form in which we can reconstruct them, even large clusters that doubtlessly reflect transmission can be affected by spontaneous composition and random parallelism, and may therefore contain glosses attached to a genuine historical core by chance. We must therefore bear in mind that clusters inform us about the general contours of transmission, i.e., they attest to it and allow us to identify manuscripts containing transmitted material, but they do not provide us with an exact picture, i.e., we cannot be sure that all parallel glosses in a cluster were transmitted. In the case of clusters containing glosses with all ranks, only the core of these clusters, constituted by glosses with rank 3, can be considered as certainly transmitted, while we must remain in doubt about the glosses with the lower ranks, 1 and 2. For this reason, the smallest and most generically-looking sets, in particular those constituted by only one or two parallel glosses with ranks 1 and 2, may be phantoms created by scholarly reconstruction.

3. Data

To demonstrate the practical utility of the network-based method, I select a single representative corpus of medieval annotations – the glosses to the first book of the *Etymologiae* of Isidore of Seville.[20] This corpus displays characteristic traits of medieval western glossing, including those that cause the most problems to scholars applying traditional methods. It is therefore ideally suited for testing the network-based approach described below.

The *Etymologiae,* produced in the first decades of the seventh century in Visigothic Spain by the bishop Isidore of Seville (d. 636), was the most important encyclopedic work of the western Middle Ages. That it survives today fully or in parts in at least 1,400 manuscripts copied between the seventh and the sixteenth centuries is a lasting testament to the popularity of this work. Many of these manuscripts are annotated. The highest intensity of annotation took place in the early Middle Ages (c. 600–1000 CE), a period in which the *Etymologiae* was the only widely available encyclopedia and served as the ultimate go-to for

20 This corpus is published online at: https://db.innovatingknowledge.nl/edition/#right-network. The underlying data can be downloaded as an Excel file from Zenodo: 10.5281/zenodo.5359401. Those wishing to use this data will note that the published dataset uses a slightly different particularity ranking scale, with four ranks and a broader clustering scheme including sets larger than five glosses among clusters as small clusters (see footnotes 23 and 27).

the Latin-writing world. Today, 74 manuscripts of this work that were annotated in the period from the eighth to the beginning of the thirteenth centuries are known, preserving slightly more than 7,000 glosses. This corpus is highly organic, being the work of many anonymous annotators separated by time, space, linguistic context, interest, and skill.[21]

The corpus glosses are unevenly distributed, both across the identified witnesses (from one to more than a thousand glosses in a manuscript) and the twenty topic-based books into which the *Etymologiae* is structured (from 42 to more than 4,000 glosses per book). Because of this disparity, only parts of this larger corpus are suitable for network analysis. More specifically, one of the twenty books, the first book dedicated to the ancient and medieval discipline of grammar (*grammatica*), preserves most of the known glosses: 4,286 (i.e., ~ 62% of the entire corpus) and is annotated in the most manuscripts (54 of the 74 witnesses, i.e., 73%). These glosses to the first book of the *Etymologiae* form the main dataset for this study.

An overview of the 54 manuscripts that preserve the 4,286 glosses analyzed below, with the latter's full shelfmarks, shortened labels referenced in this study, assumed periods and places/regions of glossing, and the numbers of all glosses and parallel glosses of different ranks, are provided in Appendix I.[22]

3.1 Parallel and shared glosses in the dataset

Of the 4,286 glosses that constitute the corpus of glosses to the first book of the *Etymologiae,* 2,554 are isolated, and 1,732 are parallel. If described as a minimum corpus of unique glosses, i.e., parallel glosses encountered in multiple witnesses are considered manifestations of the same entity, then the corpus consists of 3,279 glosses, of which 2,554 are isolated (i.e., feature in the corpus exactly once) and 725 are parallel (i.e., feature in the corpus twice or more). Tab. 1 presents the distribution of parallel glosses with various ranks in the minimum corpus based on the number of manuscripts they appear in.[23] As can be gleaned from it, most parallel glosses in the corpus adopted for this study were assigned rank 2 (417, ~ 58% of parallel glosses) and appear in two manuscripts (547, ~ 75% of parallel glosses). Nevertheless, approximately 31% of parallel glosses from this data-

21 The identified annotated manuscripts and their historical context of origin are described in detail in Steinová, "Annotation of the Etymologiae of Isidore of Seville in Its Early Medieval Context."

22 As only 47 of these manuscripts contain any parallel glosses, seven manuscripts (highlighted in grey) are included in the Appendix only for the sake of completeness.

23 In this article, parallel glosses assigned ranks 3 and 4 in the original dataset have the rank of 3.

no. of mss. in which a gloss appears	all parallel glosses	rank 1	rank 2	rank 3
in two mss.	547	72	304	171
in three mss.	110	4	75	31
in four mss.	35	3	22	10
in five mss.	24	1	12	11
in six mss.	5	0	2	3
in eight mss.	2	1	1	0
in nine mss.	1	0	1	0
in ten mss.	1	0	0	1
Total	725	81	417	227

Tab. 1 Distribution of glosses in the minimum corpus based on the extent of their co-occurrence and particularity rank.

set were assigned rank 3, and up to ten manuscripts from the corpus feature the same parallel gloss.

The high number of isolated glosses and the limited extent of gloss parallelism, rarely extending beyond three manuscripts, are not the features of the corpus, nor do they inform us about the character of medieval western glossing. They are very likely a consequence of the loss of annotated manuscripts from the Middle Ages.[24] If we had access to the corpus of all glosses to the first book of the *Etymologiae* generated in the Middle Ages, as opposed to only those that are preserved by surviving manuscripts, we would likely see that many glosses that appear isolated were in fact parallel, and some of the parallel glosses had been shared by more manuscripts than is the case in the present-day corpus. For this reason, the reconstructions of relationships between the surviving annotated manuscripts of the *Etymologiae* must be understood as representing the best achievable minimalistic result, rather than faithfully corresponding to historical reality. Moreover, the links created between manuscripts by gloss parallelism should not be, as a rule, understood to reflect direct relationships between surviving witnesses

24 In this respect, it may be compared to the bifidity of stemmata in traditional textual scholarship; see Guidi and Trovato, "Sugli stemmi bipartiti. Decimazione, asimmetria e calcolo delle probabilità." On the extent of the loss of manuscripts from the Middle Ages, see Buringh, *Medieval Manuscript Production in the Latin West,* 179–252.

or the transfer of material from one witness to another, but rather, as in a stemma, indirect relationships facilitated by lost intermediaries.[25]

3.2 Gloss sets and gloss clusters in the dataset

The 725 parallel glosses can be split into 228 sets that are unique to anything between two and ten manuscripts. Of these 228 sets, 142 (62.3%) consist of a single gloss, twenty (8.7%) of two glosses, 26 (11.5%) of two to nine glosses, and 40 (17.5%) of ten or more glosses. The forty sets of ten or more glosses can be sorted into twelve clusters of 11 to 157 glosses, labelled by letters of the alphabet to distinguish them.[26] In addition, one set of seven glosses and nine sets of one to three glosses can be recognized as clusters following the principle that all their constituent glosses have a rank of 3. The latter sets, which illustrate that glosses could circulate independently, are labelled as C and subdivided into seven micro-clusters. These twenty clusters consist, on average, of eight or nine glosses and have an average rank of 2.87 (i.e., leaning strongly towards being particular rather than generic). An overview of the twenty clusters is provided in Tab. 2. The remaining 178 sets of one to nine glosses with lower ranks of 1 and 2 are assigned the generic label X, so that they can be filtered out from the following network analysis and visualization. These unassigned sets consist, on average, of one or two glosses with an average rank of 1.96 (i.e., leaning towards generic rather than particular).[27]

25 Compare with Roelli, "Definition of Stemma and Archetype," 213.

26 The discrepancy between the number of sets with ten or more parallel glosses and the number of established clusters is due to the consideration of extrinsic evidence. For example, Paris7490 today contains only chapters 5–17 of the first book of the *Etymologiae,* and Orleans296 chapters 21–44. Nevertheless, the analysis of glossing hands, layout and ruling pattern, and context of preservation suggest that the two manuscripts are closely related, and their collections of glosses may represent two parts of a single whole (e.g., two damaged codicological units from the same glossing circle); Steinová, "Annotation of the Etymologiae of Isidore of Seville in Its Early Medieval Context," 24. For this reason, Paris7490 and Orleans296 are assigned to the same clusters as Orleans296/Paris7490, even though they are treated as separate nodes in the analytical sections of this article.

27 The original dataset used for this study distinguishes small clusters of five to ten glosses (labelled as H, R, and T–Z), which are treated as unassigned sets (X) in this article and clusters distinguished by numerals (e.g., F1 and F2), which corresponds to the distinction between parallelism with Orleans296 and Paris7490 (see the previous footnote). The distinction of unassigned sets X1 and X2, introduced purely to distinguish different sets of parallel glosses to the same lemma, is also not maintained.

label	manuscripts that share the glosses from the cluster	no. of glosses	avg. rank	no. of mss.
Clusters (13)				
A	Hamilton689, Harley3941, MontpellierH53, Paris7585, Paris7670, Paris7671, Paris11278, Reims425, Reims426, Trier100, VLO41, VLF82	14	2.97	12
B	Reims426, VLO41	18	2.11	2
D	CesenaSXXI5, VeniceII46	11	3	2
E	IRHT342, CotCalAxv, GothaI147, Paris7585, Queen320	50	3	5
F	Harley3941, Orleans296/Paris7490, Reims426	157	2.29	4
G	IRHT342, CotCalAxv, Harley3941, Laon447, Paris7585, Queen320	30	2.4	6
I	Orleans296/Paris7490, VLO41	54	1.84	3
M	Paris7670, VLO41	17	2	2
N	Orleans296/Paris7490, Paris7670	29	2.06	3
O	Harley3941, Paris7670	13	2.31	2
P	RAH25, RAH76	17	2.12	2
Q	CotCalAxv, Harley3941, Paris7585, Paris11278	7	3	4
S	Orleans296/Paris7490, Reims426	21	1.93	3
micro-clusters (7)				
C1	Arundel129, Bern101, BrusselsII4856, Clm4541, Clm6250	1	3	5
C2	Hamilton689, MilanL99sup, VatLat5763	1	3	3
C3	IRHT342, Harley3941, Paris7490, VLF82, Wolfenbuttel64	1	3	5
C4	IRHT342, Bern101, Harley3941, Paris7559, Paris7671, Schaffhausen42, VLF82	3	3	7
C5	Clm4541, Clm6250, Laon447, Schaffhausen42	1	3	4
C6	Bologna797, Paris11278, Schaffhausen42	1	3	3
C7	IRHT342, GothaI147, Harley3941, Paris7559, Paris7585, Paris7670, Paris10293, Queen320, Schaffhausen42, Wolfenbuttel64	1	3	10

Tab. 2 Overview of gloss clusters in the corpus.

3.3 Historical context of the dataset

Clues put together based on manuscript evidence situate the glossing of the first book of the *Etymologiae* within the general contours of early medieval intellectual life. In the early Middle Ages, most of the intellectual production of the Latin-writing world, including glosses, originated in monastic and cathedral scriptoria, libraries, and schools. These religious centers, at least 650 of which have been documented and of which most if not all produced and used books, formed an interconnected network stretching across the Latin-writing world.[28] The circulation of glosses happened via this network through mechanisms that entailed written and oral transmission (e.g., the exchange of annotated manuscripts and personnel, and instruction).[29] The patterns of transmission of glosses to the first book of the *Etymologiae* does map onto the historical network of intellectual centers, although due to the loss of material from the Middle Ages, we can obtain only its faint echo.[30]

The manuscript evidence suggests that the glossing of the first book of the *Etymologiae* in the Carolingian environment was driven by the integration of this text into the grammatical curriculum starting from the end of the eighth century.[31] It thus appears to have been a response to the needs of school education, serving Carolingian schoolmasters and students. Many of the surviving annotated manuscripts of the first book of the *Etymologiae* reflect this purpose: they can be described as schoolbooks or instructional manuals and were produced and annotated during the ninth century in the modern region of northern France, the heart of the Carolingian empire. However, glosses are also found in manuscripts from Brittany, England, northern Italy, the German area, and Spain. Moreover, some of the annotated manuscripts, including those that preserve the richest collections of glosses, are books that were designed to sit on a lectern in a library, suggesting that while school study may have been an important stimulus for glossing, glosses nonetheless originated in and permeated other contexts of use. Overall, the extrinsic clues create the impression of a substantial circulation of the glosses to the first book of the *Etymologiae,* rather than the prevalence of spontaneous composition.

---

28 Ganz, "Book Production," 789; Contreni, "The Pursuit of Knowledge in Carolingian Europe," 127.

29 Teeuwen, "Marginal Scholarship: Rethinking the Function of Latin Glosses in Early Medieval Manuscripts," 30–32.

30 On these intellectual networks, see for example Moulin, "Paratextuelle Netzwerke."

31 Steinová, "Annotation of the Etymologiae of Isidore of Seville in Its Early Medieval Context," 19–29.

4. Method

4.1 A co-occurrence network as a model

The basic blueprint for networks constructed, described, and analyzed in the following sections may be called a co-occurrence network. A co-occurrence network is one in which similarity between entities, for example the sharing of properties or constituent elements, is used as a basis for establishing relationships between them.[32] Unlike social networks, the most common network model currently employed in historical network research, co-occurrence networks do not represent direct relationships facilitated by human interaction. Some of the concepts from social network analysis should, therefore, not be assumed to apply to them. Co-occurrence networks are instead suitable for exploring relationships and similarities between man-made objects, such as texts, written artifacts, or creative output (e.g., music and visual art).[33] In this regard, they resemble stemmata rather than social networks.[34]

This study is concerned with co-occurrence networks representing relationships between manuscript witnesses of a gloss corpus based on the patterns of gloss parallelism. In this model, manuscripts serve as nodes while parallel glosses supply the edges. Thus, if two manuscripts share a parallel gloss, their nodes are connected by an edge, if three manuscripts share it, all three are connected by edges, and if such a gloss appears in four manuscripts, all four are connected (Fig. 3). Because of this principle, a characteristic trait of co-occurrence networks is the presence of many locally complete sub-graphs.[35]

Since multiple glosses may be shared by two manuscripts, rather than plotting many parallel edges between two nodes, I provide each edge with a weight corresponding to the number of parallel glosses that form it. For example, the heaviest edge in the dataset chosen for this study, which connects Harley3941 and

32 A similar network model has been proposed in Valleriani et al., "The Emergence of Epistemic Communities in the Sphaera Corpus," 57–58. The co-occurrence network model can be considered a more generic version of the network of shared textual transmission developed in Fernández Riva, "Network Analysis of Medieval Manuscript Transmission. Basic Principles and Methods"; and explored in Kapitan, "Perspectives on Digital Catalogs and Textual Networks of Old Norse Literature." See also the networks of material culture explored in Peeples et al., "Analytical Challenges for the Application of Social Network Analysis in Archaeology," 65–67.

33 Compare with Brughmans, Collar, and Coward, "Network Perspectives on the Past," 11.

34 On stemmata as models and graphs, see Hoenen, "The Stemma as a Computational Model"; Roelli, "Definition of Stemma and Archetype."

35 The number of edges generated by a parallel gloss shared by N manuscripts can be calculated as $N(N-1)/2$. Thus, the parallel gloss shared by most manuscripts in this corpus, which is ten according to Tab. 1, generates 45 edges.

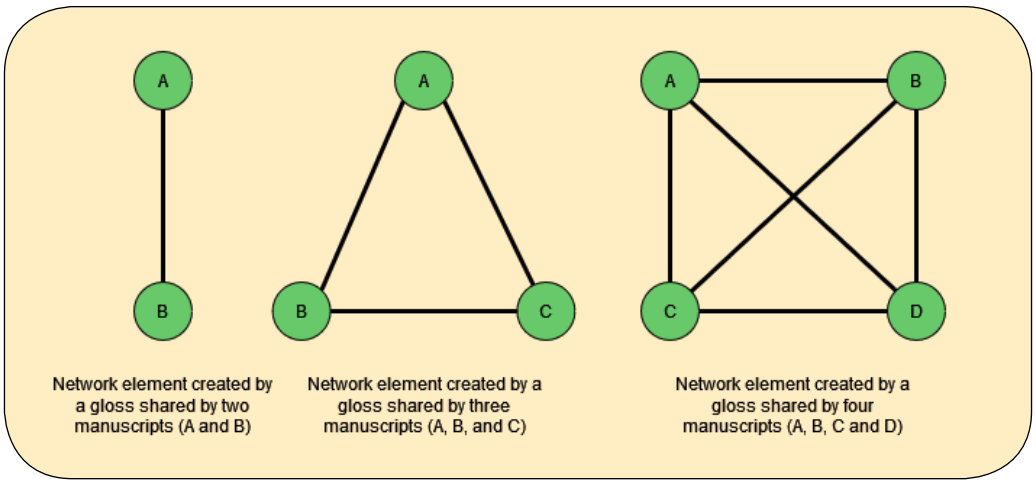

Fig. 3 The basic model for the network structure examined in this article. Illustrated here are the connections created by a parallel gloss shared by two, three, and four manuscripts. Produced with app.diagrams.net.

Orleans296, is constituted by 175 parallel glosses, and thus has a weight of 175. All network graphs described below are undirected.

4.2 Preparing the data

The data used for the construction of the co-occurrence networks was taken from a TEI-XML file containing a transcript of the glosses to the first book of the *Etymologiae,* produced in the context of preparing a digital scholarly edition of this gloss corpus.[36] As part of the encoding, each of the 4,000+ glosses in the XML file was equipped with attributes that indicate whether it was isolated or parallel, along with its particularity rank and cluster or set. This data was exported into an edge table suitable for network analysis using an XSL script.[37] The main edge table used in this study records the following information: a) labels of manuscript pairs sharing glosses (as source and target); b) cluster to which these edges belong (as cluster); c) the number of glosses of particular ranks constituting the edge (as rank 1, 2, and 3); and d) the total number of parallel glosses constituting the edge (as no. of glosses, see Tab. 3). The complete edge table has 417 rows, i.e., it corresponds to 417 edges between the 47 manuscripts containing at least one parallel gloss. It is provided in Appendix II.

36 This TEI-XML file is available at: https://github.com/HuygensING/isidore-glosses.
37 I would like to thank my colleague Peter Boot for writing this script.

Source	Target	Cluster	rank 1	rank 2	rank 3	no. of glosses
Harley3941	Orleans296	F	6	74	58	138
Orleans296	VLO41	I	3	35	0	38
Gotha1147	Paris7585	E	0	0	32	32
Harley3941	Paris7585	G	1	14	15	30
Harley3941	Orleans296	X	3	25	1	29
Harley3941	Reims426	X	2	19	3	24
Orleans296	Paris7670	N	5	18	0	23
Harley3941	Paris7490	F	0	10	9	19
Reims426	VLO41	B	1	14	3	18
RAH25	RAH76	P	0	15	2	17
Paris7670	VLO41	M	2	13	2	17
Harley3941	VLO41	X	1	12	3	16
Harley3941	Paris7670	X	1	14	1	16
Paris7585	Queen320	E	0	0	15	15
Orleans296	VLO41	X	1	13	1	15

Tab. 3 A segment of the complete edge table representing the co-occurrence network of parallel glosses in the studied corpus. Displayed are the top 15 rows ordered by the number of parallel glosses.

This edge table is complemented by a node table containing information about the 47 manuscripts containing parallel glosses to the first book of the *Etymologiae*. It was manually prepared by the author of this study and is also attached to this article in Appendix II. Its columns store the following information: a) manuscript label taken from Appendix I (as label); b) manuscript type (as type) based on whether the manuscript is a grammatical handbook containing only the first book of the *Etymologiae* (AI, 29%), a complete copy of the *Etymologiae* (BI, 67%), or a manuscript containing only excerpts from the first book of the *Etymologiae* (EXC, 4%); c) place of estimated glossing, represented as GPS coordinates (as latitude and longitude); and d) the number of parallel glosses found in the manuscript (as no. of parallel glosses, see Tab. 4). The GPS coordinates, manuscript type, and number of parallel glosses are used in the visualization plotted in section 6.

Fig. 4 represents a sample segment of a co-occurrence network of gloss parallelism between eight of the manuscripts studied here, constructed from the edge and node tables described above.

Label	type	latitude	longitude	no. of parallel glosses
Harley3941	BI	48.16667	–2.83333	305
Orleans296	AI	48.85341	2.3488	301
VLO41	AI	47.80281	2.31321	191
Paris7670	BI	48.85341	2.3488	130
Reims426	BI	49.25	4.03333	127
Paris7585	BI	51.27904	1.07992	116
Paris7490	AI	48.85341	2.3488	68
Paris7559	AI	48.85341	2.3488	42
Paris7671	AI	49.15964	5.3829	37
IRHT342	BI	48.16667	–2.83333	35
GothaI147	BI	48.16667	–2.83333	32
Paris11278	AI	43.71553	1.604	30
VLF82	BI	48.85395	2.33449	29
RAH25	BI	42.32962	–2.8722	28
RAH76	BI	40	–4	28

Tab. 4 A segment of the node table representing the nodes in the co-occurrence network of parallel glosses in the studied corpus. Displayed are the top 15 rows ordered by the number of parallel glosses.

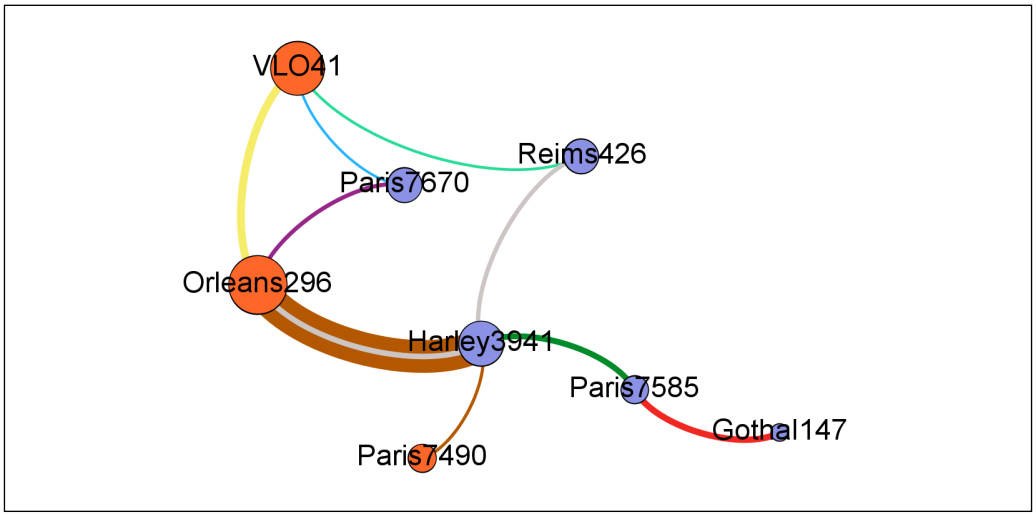

Fig. 4 A sample segment of a co-occurrence network of parallel glosses. The visualization displays manuscripts connected by edges with a weight larger than seventeen (i.e., it corresponds to the first nine lines of the edge table in Tab. 3). Edges are colored based on the cluster they represent (light green: B, dark red: E, brown: F, dark green: G, yellow: I, light blue: M, dark purple: N, and grey: X). The thickness of edges is proportional to their weight. The size of the nodes is proportional to the number of parallel glosses they contain. Parallel edges (here between Harley3941 and Orleans296) are overlaid. The visualization was created in Gephi with Yifan Hu layout.

4.3 Accounting for multiple scenarios

The combination of the particularity ranking and gloss clustering allows us to construct and examine several network scenarios based on different degrees of inclusivity of data (e.g., treating all parallel glosses as shared versus treating only glosses with rank 3 as shared), rather than having to represent the entire corpus with a single co-occurrence network. In this way, we can mitigate some of the issues stemming from the complexity of the real-world data and the limits of the scholarly reconstruction. The comparison of different scenarios yields insights into the stratigraphy of the corpus that would otherwise remain concealed. In the end, we can select one or more networks from the available scenarios that we find most suitable to answer our research questions.

The particularity ranking allows us to distinguish and isolate specific layers from the larger co-occurrence network of all parallel glosses, based on whether we consider them transmitted or not. The edge tables corresponding to individual layers can be derived from the main edge table by extracting and combining columns with glosses of specific ranks. We can work with up to six different scenarios: three for individual ranks (Rnk1, Rnk2, and Rnk3), two for a combination

of ranks (Rnk12 and Rnk23), and one representing all parallel glosses (Par). Gloss clustering helps us to distinguish those elements of the network that we can consider historical artifacts with certainty (clusters and micro-clusters) from those that may be mere noise (unassigned sets). We can use these to plot two types of networks: one in which all sets are included, irrespective of their weight and label (clustered), and one from which the sets labelled as X are removed (clustered-noX). We can also ignore the division into clusters and sets and construct co-occurrence networks, in which all glosses shared by two manuscripts establish an edge between them (unclustered). The two clustered network types differ from unclustered networks in that they are hypergraphs, i.e., two nodes can be connected by several parallel edges, since a pair of manuscripts can participate in multiple clusters.[38] The manuscript pair Harley3941-Paris7585, for example, features in clusters A, G and Q, as well as the micro-cluster C7 (see Tab. 2). Moreover, the values of the network properties of the two clustered network types can exceed the maximum values observable in unclustered networks. The edge tables for constructing clustered-noX networks can be derived from the main edge table by removing the rows assigned label X (180 rows). The edge tables for constructing the unclustered networks can be produced by contracting all rows with the same source and target (282 rows).

5. Analysis

By exploiting the particularity ranking and gloss clustering described in section 2, each of the three network types representing a different extent of gloss clustering (unclustered, clustered, and clustered-noX) can be paired with six network layers or layer combinations obtained through the particularity ranking (Rnk1, Rnk2, Rnk3, Rnk12, Rnk23, and Par). Thus, altogether we can construct eighteen co-occurrence networks from the data introduced in section 3. In this section, the main analytical part of this article, the network properties of these eighteen networks are examined to gain as complete a picture of the corpus as possible, probing its structure and dynamics by comparing various network types and layers and assessing their relative value. The following analysis does not require visualizing any of the eighteen networks.[39] Rather, the conclusions about the corpus are reached entirely from the network properties. At the end of this section, one of the eighteen networks is selected for visualization and detailed treatment in the following section.

The eighteen networks are labelled by a combination of a network layer and type in the following two sections, e.g., Rnk3-clustered-noX refers to a net-

38 Newman, *Networks,* 114–15.

39 The author of this paper, nevertheless, explored all eighteen network scenarios in Gephi to obtain some of their network properties.

Network		Rnk1	Rnk2	Rnk3	Rnk12	Rnk23	Par
Nodes	**Unclustered**	30	40	38	41	47	47
	Clustered	30	40	38	41	47	47
	clustered (no X)	7	12	35	12	35	35
Edges	**Unclustered**	87	176	160	201	262	282
	clustered	100	218	237	245	393	417
	clustered (no X)	13	23	172	24	180	180
Components	**unclustered**	2	1	2	1	1	1
	clustered	2	1	2	1	1	1
	clustered (no X)	1	2	3	2	3	3
Network diameter	**unclustered**	4	3	5	3	4	4
	clustered	4	3	5	3	4	4
	clustered (no X)	3	3	4	3	4	4
Density	**unclustered**	0.2	0.226	0.228	0.245	0.242	0.261
	clustered	0.23	0.279	0.339	0.279	0.364	0.386
	clustered (no X)	0.619	0.348	0.289	0.364	0.303	0.303
Avg. degree	**unclustered**	5.8	8.8	8.421	9.805	11.149	12
	clustered	6.667	10.9	12.53	11.951	16.42	17.745
	clustered (no X)	3.714	3.833	9.829	4	10.286	10.286
Median degree	**unclustered**	4.5	7.5	7	8	10	10
	clustered	5	8.5	6.5	9	13	13
	clustered (no X)	4	3.5	7	4	9	9
Max. degree	**unclustered**	16	27	26	30	33	35
	clustered	20	38	51	41	66	68
	clustered (no X)	6	12	34	12	36	36
Avg. edge weight	**unclustered**	1.61	4.97	3.31	5.05	5.35	5.47
	clustered	1.4	4.01	2.23	4.14	3.57	3.7
	clustered (no X)	2.31	11.56	2.52	12.33	3.88	4.05
Max. edge weight	**unclustered**	9	106	60	115	166	175
	clustered	6	74	58	80	132	138
	clustered (no X)	6	74	58	80	132	138

Tab. 5 Selected network properties of the eighteen possible networking scenarios described in section 4.

work that consists of only glosses with rank 3, belonging to one of the twenty clusters outlined in section 3.2. The network properties employed as descriptors of individual networks are: the number of nodes, edges[40] and connected components,[41] network diameter,[42] network density,[43] the average, median, and maximum degree,[44] and the average and maximum edge weight (Tab. 5).[45]

5.1 The number of nodes and edges

These two variables tell us how many annotated manuscripts contain parallel glosses of a specific rank or glosses belonging to sets and clusters (nodes) and how many instances of gloss parallelism there are within networks with certain properties (edges).[46] The number of nodes and edges gives us a glimpse of the similarity and robustness of individual networks. They reveal that Par-clustered-noX and Rnk23-clustered-noX both have 35 nodes and 180 edges and thus share network properties apart from their edge weight distribution. Moreover, Rnk3-clustered-noX (35 nodes and 172 edges) closely resembles the previous two, having an identical diameter (4) and number of components (3). As it contains more than 95% of the edges of the two less restrictive networks, it also has a similar average degree (9.829 compared to 10.286) and density (0.289 compared to 0.303). The identical properties of Par-clustered-noX and Rnk23-clustered-

40 The number of edges corresponds to the number of rows in an edge table after rows with specific ranks or cluster labels are removed or contracted (see section 4.3). The number of nodes corresponds to the number of unique manuscript labels that remain in the source and target columns of the same edge table. Both values can also be obtained via Gephi.

41 The number of connected components can be calculated using a method described in Barabási, *Network Science,* sec. 2.9. In this study, it was obtained via Gephi.

42 The network diameter can be calculated using a method described in Ibid., sec. 2.8. In this study, it was obtained via Gephi.

43 The network density was calculated from the number of edges and nodes following the method described in Newman, *Networks,* 128–30. It can also be obtained via Gephi.

44 The average degree was calculated from the number of edges and nodes following the method described in Ibid., 127–28. It can also be obtained via Gephi. The median and maximum degrees were established based on the degree distribution produced by Gephi. The degree distribution could also be obtained manually from an adjacency matrix constructed from the edge table; see Barabási, *Network Science,* sec. 2.3; Newman, *Networks,* 106–8.

45 The average edge weight corresponds to the average value of the column no. of glosses of an edge table. The maximum edge weight is equal to the highest value present in the same column in the same edge table.

46 The range of values we can expect for nodes is up to 47, the total number of annotated manuscripts containing parallel glosses. In unclustered networks, the number of edges can reach 1,081 (if all annotated manuscripts shared at least one parallel gloss with all other annotated manuscripts), while in clustered networks, the number could potentially be higher because two nodes can be connected by parallel edges. In practice, the number of edges is lower because few of the possible connections are present in real-life networks. See Barabási, *Network Science,* sec. 2.5.

noX indicate that glosses with rank 1 play a limited role within our co-occurrence networks. Indeed, Rnk1-clustered-noX has only seven nodes and thirteen edges, containing thus only 20% of the nodes and approximately 7% of edges of the largest network of the same type (Par-clustered-noX), and only 15% of nodes and 3% of the edges of the largest network that can be produced from the data (Par-clustered). This is partially because only 11% of glosses from the dataset have the lowest particularity rank. However, it also transpires from Tab. 5 that there is no manuscript connected to other manuscripts only by glosses with rank 1 (otherwise, the number of nodes of Par and Rnk23 networks would differ). Moreover, only 18 out of 282 edges of Par-unclustered and 24 out of 417 edges of Par-clustered (approximately 6% of the edges of both most inclusive networks) are constituted solely by glosses with rank 1. Rnk1 networks thus do not carry much weight on their own. Therefore, it seems safe to exclude glosses with rank 1 from further consideration in this analysis.

Rnk2-clustered-noX (12 nodes and 23 edges) and Rnk12-clustered-noX (12 nodes and 24 edges) also comprise a very small proportion of nodes and edges of the most inclusive network that can be constructed from the data available. They are thus not particularly robust when it comes to analyzing the connectivity within the corpus. However, Rnk2-clustered-noX has a very high average (11.56) and maximum (74) edge weights, an indication that glosses with rank 2 represent a significant portion of the volume of many clusters. Therefore, Rnk2-clustered-noX cannot be discarded altogether. As is shown below, it is a vital complement to Rnk3-clustered-noX, the most restrictive network that can be constructed from the data available and the network that is most relevant to understand the connections between annotated manuscripts of the first book of the *Etymologiae* that are certainly due to the transmission of glosses.

5.2 The number of connected components and network diameter

The number of connected components reveals whether all the nodes in a network are mutually interconnected (one component) or whether any parts of the network are isolated from each other (a higher number of components).[47] In this study, the latter means that certain annotated manuscripts mutually share parallel glosses, but otherwise differ from all other manuscripts in the corpus. As Tab. 5 shows, we find a single connected component in many of the network scenarios we can construct. However, once unassigned sets are removed, disconnected components appear. The extensive connectivity thus appears to be due to noise. If we furthermore limit our attention to glosses with rank 3, the network disintegrates into three components. This fragmentation points to the existence of several disconnected or weakly connected glossing communities.

47 Ibid., sec. 2.9; Newman, *Networks,* 133–37.

The network diameter corresponds to the longest direct path between nodes in a network component.[48] A network with a diameter of 1 is complete (as it is possible to reach every node from every other node), while in a network with the diameter of 2, all nodes in a component are connected through a single central node (called a hub) so that it is possible to reach any node from any other node through this central point. The network diameter thus provides us with a measure of connectivity within a network related to the presence of clusters and weakly connected segments. Translated into scholarly language, the larger the diameter, the more annotated manuscripts we can expect to display weak connections to other manuscripts. As in the case of connected components, identifying these weakly connected segments in a network is valuable for tracing parts of the corpus that are mutually distinct or relatively different and thus identifying glossing communities that are disconnected or poorly connected.

In our case, the smallest network diameter we observe in our co-occurrence networks is 3. This diameter, just one step beyond the network with a single central hub, appears in networks built solely or principally from glosses with rank 2 and one of the networks constituted by glosses with rank 1. It is an indication that glosses with these two lower ranks, particularly rank 2, play a role in shortening paths between nodes. By contrast, two networks constructed from glosses with rank 3, Rnk3-unclustered and Rnk3-clustered, have a large network diameter of 5, while Rnk3-clustered-noX has a smaller diameter of 4. The decrease in the diameter in Rnk3-clustered-noX compared to the other two Rnk3 networks signals that in the latter networks, particular far-flung nodes are only connected to the rest by glosses not assigned to any cluster. Once these weak, possibly phantom, connections are removed, the network becomes disconnected into several components.[49] For our purposes, Rnk3 networks show the most topographic detail and therefore merit further examination to identify weakly connected segments and potential bridges (i.e., manuscripts that connect otherwise disconnected parts of the network). Rnk2 networks, on the other hand, could be useful to inspect to determine to what extent glosses with rank 2 generate meaningful connections within our co-occurrence networks that do not feature in networks constructed from glosses with rank 3, and to what extent the increased connectivity is due to noise.

48 Barabási, *Network Science,* sec. 2.8; Newman, *Networks,* 133.

49 Thus, the number of components in Rnk3-clustered-noX increases as its network diameter decreases.

5.3 Network density

Network density informs us about how complete a network is, i.e., what proportion of the possible connections in the network, as given by the number of nodes, have been realized.[50] In this study, the network density provides us with a complementary perspective into the extent to which annotated manuscripts are connected to other annotated manuscripts containing glosses to the first book of the *Etymologiae*. A low density may suggest that manuscript annotators rarely acquired glosses from other intellectual centers (if we assume transmission, such as in the case of glosses with rank 3), or seldomly came up with glosses similar to those coined elsewhere (if we rather assume spontaneous composition), their glossing activity being one-of-a-kind. A high density, by contrast, is a significant indicator of gloss parallelism and therefore, potentially, of the extensive circulation of glosses. Typically, the density value can range from 0 (i.e., 0%, if no manuscript shares a gloss with another manuscript) to 1 (i.e., 100%, if every manuscript shares at least one parallel gloss with all other manuscripts).[51]

If we disregard networks with low robustness, the network density values in the corpus studied here range from 0.226 (Rnk2-unclustered) to 0.261 (Par-unclustered) in unclustered networks, from 0.279 (Rnk2-clustered and Rnk12-clustered) to 0.386 (Par-clustered) in clustered networks, and from 0.289 (Rnk3-clustered-noX) to 0.303 (Rnk23-clustered-noX and Par-clustered-noX) in clustered networks removing the unassigned sets. As can be seen, the gloss parallelism observable in the more robust co-occurrence networks is relatively stable, corresponding to between approximately 23% and 39% of the gloss parallelism we would see if all manuscripts shared glosses with all other manuscripts. These ratios make our co-occurrence networks relatively dense, in particular when compared to other real-world networks, such as those examined by Lászlo Barabási in his *Network Science*.[52] Some of this density is certainly due to the method described here, for example the decision to remove isolated nodes from the co-occurrence networks studied here (thus the number of nodes varies per network), and at least in part due to the presence of many locally complete sub-graphs (see section 4.1). It can also be partially attributed to noise due to spontaneous composition and random gloss parallelism. We can see this in Par-unclustered and Par-clustered, the two most inclusive co-occurrence networks, if we subject them to a small test, removing edges constituted by less than a certain number of glosses. The density of Par-unclustered (0.261) drops to 0.14 if we exclude edges

50 Barabási, *Network Science,* sec. 2.5; Newman, *Networks,* 128–30.

51 Due to the presence of parallel edges, the value could theoretically exceed 1 in a clustered network.

52 For example, the science collaboration and the citation networks provided as examples of real-world networks by Barabási have densities of 0.00035 and 0.000046, respectively. See Barabási, *Network Science,* sec. 2.2.

constituted by a single gloss, to 0.093 if we also exclude edges constituted by two glosses, and to 0.06 if we exclude edges constituted by less than six glosses. The density of Par-clustered (0.386) similarly drops to 0.168 if we remove edges constituted by a single gloss, to 0.127 if we also exclude edges constituted by two glosses, and to 0.069 if we exclude edges constituted by less than six glosses.

This rapid decrease in density could mean that the relatively high densities of our networks are due to noise rather than meaningful connections between annotated manuscripts. If this were the case, the gloss parallelism in an organic corpus of glosses could be assumed to occur mainly due to processes other than transmission. However, the most restrictive network, Rnk3-clustered-noX, which filters out potential noise in a stricter fashion than the test described above, is significantly denser than even the most lightly filtered network containing unassigned sets, and has a higher density (0.289) than even its unclustered counterpart, Rnk3-unclustered (0.228). The high densities of the clustered-noX networks can be interpreted as an indicator that transmission contributes to the high gloss parallelism observed in our co-occurrence networks to a rather significant degree. Moreover, it also tells us that by gloss clustering and removing unassigned sets, it is possible to remove some of the noise from the corpus without significantly imperiling, and even increasing, its informative value. Filtering out edges based solely on their weight, on the other hand, degrades the networks wholesale, eliminating not only noise but also valuable information. Gloss clustering and the removal of unassigned sets are thus essential data pre-processing strategies, if we want to obtain high-quality insights into organic gloss corpora, while ignoring the presence of gloss clusters and sets of low importance is likely to produce unreliable results, particularly if the study of gloss transmission is the main concern.

5.4 The average, median and maximum degree

The average degree (the average number of connections a node has with other nodes in a network) and the median degree (the number of connections that a node in the exact mid-point of the degree distribution has with other nodes in a network) inform us how well-connected individual nodes are to each other and what role different types of connections (e.g., clusters and unassigned sets) play in forging this connection.[53] Depending on the network type, the average degree indicates the average number of manuscripts that any given manuscript shares glosses with (unclustered), or the average number of sets and clusters (clustered) or clusters alone (clustered-noX) through which any manuscript is connected to

53 Ibid., sec. 2.3. In an unclustered network, the average and median degrees can range from 1 (in a network constituted by isolated manuscript pairs) to one less than the maximum number of nodes (in a complete network). In clustered networks, the average and median values can be higher because a node can be connected to another node by several parallel edges.

other manuscripts. The median degree tells us that at least half of the manuscripts in a network have the same or a higher number of connections with other manuscripts in the same network than the median value. In unclustered networks, these connections refer to manuscripts, while in clustered networks, they refer to sets and clusters (clustered) or clusters alone (clustered-noX).[54] The maximum degree reveals the largest number of manuscripts with which a manuscript from the corpus shares parallel glosses (unclustered), and the largest number of connections of any manuscripts facilitated by sets and clusters (clustered) or clusters alone (clustered-noX). Ideally, the following analysis of average, median and maximum degrees would complement the degree distribution plotted for the eighteen network scenarios described by Tab. 5. However, since this would require significant space, we shall rely on these three network properties to understand degree, the centrality measure chosen to characterize our co-occurrence networks.[55]

The values of average degree observed in unclustered networks range from 5.8 (Rnk1) to 12 (Par), meaning that, depending on the layer of the corpus examined, an annotated manuscript shares parallel glosses with, on average, between approximately six and twelve other manuscripts (20% to 26% of manuscripts in the respective networks).[56] The average degree ranges from 6.667 (Rnk1) to 17.745 (Par) for clustered networks and from 3.714 (Rnk1) to 10.286 (Rnk23 and Par) for clustered networks excluding unassigned sets, telling us that sets and clusters facilitate on average approximately seven to eighteen connections and clusters alone, approximately four to ten connections within the corpus.[57] The median degree values of unclustered networks occupy a range from 4.5 to ten manuscripts. This piece of information tells us that, depending on the layer of corpus examined, half of the manuscripts share glosses with at least four other manuscripts in the smallest network and with at least ten other manuscripts in the largest network.[58] The median degree ranges from five to thirteen connections in clustered networks and 3.5 to nine connections in clustered networks excluding unassigned sets.[59]

---

54 Thus, in Par-unclustered (median degree of 12), half of the manuscripts share glosses with twelve or more manuscripts; and in Rnk3-clustered-noX (median degree of 7), half of the manuscripts have seven or more connections to other manuscripts via clusters.

55 Alternative centrality measures used in network research are described in Newman, *Networks,* 159–77. On the relative utility of the four most common centrality measures, including degree, in historical network research, see Valeriola, "Can Historians Trust Centrality?"

56 The general corpus average is nine to ten manuscripts.

57 The general corpus average is twelve to thirteen connections if both clusters and sets are considered and seven connections if only clusters are considered.

58 The average median for the entire corpus is seven to eight manuscripts.

59 The average median for the entire corpus is nine connections between manuscripts facilitated by sets and clusters and six connections facilitated by clusters alone.

Even in the absence of an available comparison with co-occurrence networks constructed for other types of material, such as highly systematic corpora of glosses and regular texts, the observed average and median degree appear very high. In particular, it can be noted that the large majority of parallel glosses (~ 90%) are shared by only two or three manuscripts, i.e., they generate one or three edges, while only a small number of glosses (~ 4.5%) are shared by five or more manuscripts, i.e., they generate ten or more edges (see Tab. 1).[60] We could therefore expect many nodes in our co-occurrence networks to have relatively low degrees and very few nodes that have high degrees.[61] Yet, there are relatively few manuscripts with the lowest degrees of one to three in our co-occurrence networks and relatively many nodes with degrees of ten or higher. To provide examples: only seven nodes have a degree of three or lower (14.9%) in Par-unclustered (47 nodes), but 28 nodes have a degree of ten or higher (60%); meanwhile, only six nodes have a degree of three or lower (12.8%) in Par-clustered (47 nodes), but 29 nodes have a degree of ten or higher (61.7%). In the most restrictive network, Rnk3-clustered-noX (35 nodes), these ratios are slightly more balanced; nonetheless, only seven nodes have a degree of three or lower (20%), while sixteen nodes have a degree of ten or higher (45.7%).[62] The degree distribution of our co-occurrence networks does not follow the distribution of parallel glosses (or edge weights treated below), nor does it resemble the degree distribution common to many real-world networks, which follows the power law.[63]

The high average and median degree values warrant further investigation that cannot be fully carried out in this article. Intuitively, it could be assumed that this is an effect of noise, i.e., that the degree distribution is distorted by gloss parallelism due to spontaneous composition and randomness. It can be noted that the average and median degree values tend to be highest in clustered networks, whose densities suggest that they are not entirely reliable, and the lowest in clustered networks excluding unassigned sets, that is, in networks constructed with

60 Even if we considered the total number of edges a parallel gloss can generate based on the number of manuscripts it connects (rather than the number of parallel glosses), based on Tab. 1, 57% of the edges in the dataset are due to parallel glosses shared by two or three manuscripts (i.e., generating one or three edges), 29% of the edges are due to parallel glosses shared by four or five manuscripts (i.e., generating six or ten edges), and only 14% of the edges are due to parallel glosses shared by six or more manuscripts (i.e., generating 15 to 45 edges).

61 Compare with Barabási, *Network Science,* sec. 2.3, 3.5 and 4.2.

62 We encounter the highest ratios of nodes with degrees of one to three in Rnk12-clustered-noX (5 nodes, 42%) and Rnk2-clustered-noX (6 nodes, 50%). However, these networks are very small. These ratios are, therefore, less meaningful than in networks with more nodes and edges.

63 The power-law distribution means, in the most general terms, that nodes with the lowest degree should be most numerous and nodes with the highest degree should be least common in a network. Degree distribution following power law is a feature of the so-called scale-free networks treated at length in Barabási, *Network Science,* sec. 4.2.

a precaution taken against the distortive effect of noise. Unclustered networks tend to have a value lower than clustered ones and are similar to clustered-noX networks. Thus, the co-occurrence network with the most edges, Par-clustered, displays 47% more connections than Par-unclustered and 73% more connections than Par-clustered-noX. This inflation in the degree values in clustered networks is due to unassigned sets, many of which are presumably noise.

We may opt to derive our insights entirely from clustered-noX networks, considering their average and median degree values are minimally distorted by noise, or less distorted than in clustered or unclustered networks. Even so, the values of these two properties in the three most robust clustered-noX networks (Rnk3, Rnk23, and Par) are still surprisingly high. They indicate that an annotated manuscript of the first book of the *Etymologiae* contains, on average, parallel glosses from ten clusters, and at least half of such manuscripts contain parallel glosses from seven to nine clusters, even though the loss of manuscript evidence is potentially substantial. What is more, Rnk3 networks deviate from the trend described in the previous paragraph, as the average degree of Rnk3-clustered-noX (9.829) is higher than that of Rnk3-unclustered (8.79). As in the case of network density, this deviation suggests that meaningful transmission-related connections between manuscripts missing from unclustered networks are revealed in Rnk3-clustered-noX. Given the rich topography of Rnk3-clustered-noX, as suggested by the number of connected components and network diameter, it is tempting to relate the high average and median degree values to the multilayered character of the organic corpus studied in this article (a trait demonstrated in the following section).[64] The high average and median degree values of the clustered-noX networks, especially Rnk3-clustered-noX, may thus reveal the extent to which collections of annotation in manuscripts of the first book of the *Etymologiae* are amalgamating batches of glosses originating in distinct contexts (more on this in section 7).[65]

Unlike spontaneous composition and random gloss parallelism, the accumulation of glosses of heterogeneous origin in a manuscript is bound to generate hubs, i.e., manuscripts that stand out because they share glosses with an un-

64 We can engage in a thought experiment, imagining how the process of transmission of glosses together with the substrate text (e.g., copying from an annotated exemplar to its apograph) differs from the process of copying batches of glosses or individual glosses into a manuscript. While the former manner of transmission can increase the degree of a node in a co-occurrence network of parallel glosses only by one, the latter transmission process can increase the degree of a node by n, where n is the number of manuscripts that already contain the same batch. The collection of glosses thus has the potential to increase the degree of a node at a rate significantly higher than copying from an exemplar to an apograph and create hubs.

65 The corpus thus confirms the scholarly theories about the cumulative nature of early medieval glossing; O'Sullivan, "Text, Gloss, and Tradition in the Early Medieval West."

usually large number of other manuscripts, or because they are connected to other manuscripts by an unusually large number of clusters.[66] Indeed, the small difference between the average and median degrees, ranging from –0.29 to 2.8 in unclustered and clustered-noX networks, informs us that we should expect some, albeit not too many, hubs in our co-occurrence networks.[67] The maximum degree values in Tab. 5 reveal that in Par-unclustered, we encounter a manuscript that shares glosses with as many as 76% of the other manuscripts. The same manuscript shares glosses with 66% of other manuscripts in Rnk3-clustered-noX. The manuscript in question is Harley3941, which looms large among the annotated manuscripts of the first book of the *Etymologiae* due to its remarkable extent of gloss parallelism and gloss sharing – it is the most significant hub in our networks.[68] Several other nodes with high degrees compared to both average and median degree values also qualify as hubs. Rnk3-clustered-noX can be considered the most informative in this regard, given its difference between the average and median degree values (2.8). In this restrictive network, we encounter four manuscripts other than Harley3941 that are connected to a higher number of manuscripts than average in this network (29%): Paris7670 and Schaffhausen42 (50%), Paris7585 (44%), and IRHT342 (38%). These are examined against the background of extrinsic evidence in the following section.

5.5 The average and maximum edge weights

The average and maximum edge weight (the average and highest number of glosses shared between a pair of manuscripts), gives us an insight into the volume of gloss parallelism within the corpus. In our co-occurrence networks, the average edge weights range from 1.4 glosses (Rnk1-clustered) to 12.33 glosses (Rnk12-clustered-noX), while the maxima range from 6 glosses (Rnk1-clustered-noX) to 175 glosses (Par-unclustered). The edge weights are distributed in a more standard pattern than degrees, with the majority of edges consisting of very few glosses, and few edges being very heavy.[69] In the two most inclusive networks, Par-unclustered and Par-clustered, for example, edges constituted by one or two

66 Hubs are described in Barabási, *Network Science,* sec. 2.11; Newman, *Networks,* 178–80.

67 The difference between the average and the median degree values tells us to what extent the average is skewed by nodes with unusually high degrees, which are not entirely representative of the network. The more positive its value, the more outliers with high degrees (i.e., hubs) there are in a degree distribution of a given network. The more negative its value, the more outliers with low degrees there are in a degree distribution of a given network.

68 Harley3941 has the largest degree in most of the eighteen networks described in Tab. 5. The exceptions are Rnk1-clustered-noX (Reims426), Rnk2-unclustered (VLO41), Rnk2-clustered (VLO41, Orleans296 and Reims426), Rnk12-unclustered (VLO41), and Rnk12-clustered (VLO41).

69 Plotting the distribution of the edge weights on a logarithmic scale suggests that it follows the power law.

glosses comprise 65% (183 edges) and 67% (280 edges) of all network edges, respectively, while edges constituted by more than ten glosses comprise 12% (34 edges) and 8.4% (35 edges) of the network edges, respectively. This distribution matches what was observed in section 3.1, namely that most cases of gloss parallelism between manuscripts in the corpus (77.6%) are due to one or two glosses. Importantly, these lightweight edges correspond not only to unassigned sets, which could be interpreted as noise, but also to the seven micro-clusters C1–C7 and several clusters constituted by glosses with rank 3. This is why the average edge weight remains low in Rnk3-clustered-noX (2.52) but increases significantly in Rnk2-clustered-noX (11.56), which does not feature glosses that belong to unassigned clusters or those with rank 3. This network provides us with a particularly undiluted view of the volume of gloss parallelism in our co-occurrence networks, both because of the significant proportion of parallel glosses with rank 2 in the corpus (58%) and because of their role in adding weight to clusters whose contours are provided by glosses with rank 3. Indeed, if the edge tables of Rnk2-clustered-noX and Rnk3-clustered-noX are compared, it can be observed that the former contains only two edges that do not appear in the latter, i.e., most of the glosses with rank 2 appear in the same clusters as glosses with rank 3 and can be therefore considered to reflect transmission.

Just as the analysis of average and maximum degrees reveals some nodes to be hubs, the average and maximum edge weights reveal certain edges as outliers, constituted by an exceptionally high number of glosses. One edge that stands out in this regard, across all networks, is Harley3941-Orleans296, which corresponds to cluster F in clustered networks. In Par-unclustered, this edge consists of 175 glosses, while the next heaviest edge (Orleans296-VLO41) is constituted by 57 glosses (i.e., less than a third of the former's weight); in the most restrictive network, Rnk3-clustered-noX, Harley3941-Orleans296 amounts to 58 glosses, followed by Gotha1147-Paris7585 with 32 glosses (55% of the former); and in Par-clustered, the network with the most edges, this edge consists of 138 glosses, followed by Orleans296-VLO41 with 38 glosses (i.e., approximately a quarter of the former's weight). For the same reason that Harley3941 and other manuscripts with exceptionally high degrees can be considered hubs, we can identify this edge as a highway.

5.6 General trends in the co-occurrence networks of parallel glosses to the first book of the Etymologiae

After examining all co-occurrence networks constructed from the data introduced in section 3, we can identify several trends that characterize the corpus of glosses to the first book of the *Etymologiae*. First, we have seen that these co-occurrence networks are relatively dense and tend to have very high average and median degrees. This is somewhat surprising, given the historical context of the generation and circulation of organic glosses. Medieval scribes and masters may have been eager to acquire glosses, but they had, in practice, limited access to the totality of

glosses circulating in the Latin-writing world. Crucially, the degree of gloss parallelism among the annotated manuscripts is unusually high even in networks constructed by applying the most restrictive criteria that are intended to curb any potential inflation of connections due to spontaneous composition and other noise. It cannot thus be attributed to medieval annotators frequently generating glosses similar to those independently composed by others. It rather seems that a substantial gloss parallelism is an intrinsic quality of the corpus studied here. If we ask ourselves what real-world property the high density, average and median degree may correspond to, they may be a characteristic of the multilayered character of the corpus, telling us that manuscripts of the first book of the *Etymologiae* attracted glosses of heterogeneous origin. As such, these network properties may serve as an indicator of the high organicity of this corpus.

Second, while some of the gloss parallelism observable in the corpus studied here is due to spontaneous composition, generating glosses that appear identical but are not different manifestations of the same transmitted items or even due to random similarity, two properties of our co-occurrence networks suggest that it is mostly due to transmission. First, the fact that the most restrictive network, Rnk3-clustered-noX, does not look and behave like a network we could obtain by applying filters to unclustered networks, having significantly more edges. In addition, glosses with rank 2 do not form new edges in clustered networks excluding unassigned sets (i.e., eliminating noise), but rather add volume to edges that we can reconstruct as reflecting transmitted gloss clusters based on glosses with rank 3. The properties of the co-occurrence networks constructed in this section thus confirm that the circulation of the glosses to the first book of the *Etymologiae* in the early Middle Ages was extensive, as proposed based on the extrinsic evidence in section 3.3.

Third, Rnk3-clustered-noX, the network most geared towards investigating transmission patterns, is relatively topographically rich, featuring isolated components, weakly connected segments, and hubs. Based on the network analysis performed in this section, we cannot yet tell to what extent the observed disconnectedness is due to the extent of manuscript loss, and to what extent it reflects the relative mutual isolation of certain glossing communities. For this purpose, we need to conduct a qualitative examination in section 6. Another network element we could expect to see in the visualization in the following section are bridges, i.e., nodes that appear at the interconnection of otherwise unconnected or poorly connected segments.[70] These network elements are a reflection of human activity, which scholars may be particularly interested in identifying via network analysis in order to study them in greater detail with traditional methods.

70 Bridges are discussed in Barabási, *Network Science,* 2.9.

Finally, our analysis revealed that not all co-occurrence networks that can be constructed from the data described in section 3 have the same degree of robustness and quality for analytical purposes. Glosses with rank 1, for example, can be excluded from consideration without reducing the accuracy of the analysis of the corpus examined in this article, both because they are not particularly numerous within this corpus and because it can be shown that they do not generate any significant connections. We have also seen that clustered networks, that is, network scenarios in which sets and clusters are recognized but unassigned sets are not removed, suffer from an 'edge bloat' that can be interpreted mostly as noise and are therefore not particularly useful (apart, perhaps, from an analysis of this noise and its sources). For this reason, it also does not seem sufficient to treat all parallel glosses as shared. Overall, low-key gloss parallelism due to one or two glosses with the lower ranks 1 and 2 has a distortive effect on the quality of the co-occurrence networks we can construct from the dataset provided. It generates many weak connections between nodes that should not be trusted. However, since many connections facilitated by glosses with the highest particularity rank 3 are also due to one or two glosses (e.g., micro-clusters C1–C7), constructing unclustered networks and filtering out low-weight edges does not appear to be a good strategy to obtain high-quality data.

In seeking to answer the questions articulated in the introduction of this study, in particular the question of the extent and shape of the transmission of organic glosses, it seems most profitable to focus on co-occurrence networks that account for gloss clusters but exclude unassigned sets and glosses with the lowest rank, 1. In the following section, therefore, Rnk23-clustered-noX is visualized.[71] With this configuration, we benefit from a single network scenario that fleshes out both a) the general contours of connectivity within the gloss corpus studied here, namely how far and where the circulation of glosses extended to, and where it may have been weak or non-existent; and b) the intensity of this connectivity, i.e., where the exchange may have been most significant.

6. Visualization

In this section, Rnk23-clustered-noX is visualized and interpreted with the support of extrinsic evidence (i.e., paleographic, philological, linguistic, and historical information) in light of the network analysis performed in the previous section. This visualization was produced with Gephi by following these steps:

71 It could have been even more profitable to visualize Rnk3-clustered-noX and Rnk2-clustered-noX separately to obtain a subtler picture. However, as the space offered by this article is limited, the choice went to Rnk23-clustered-noX on the basis that only two edges in Rnk2-clustered-noX do not feature in Rnk3-clustered-noX (on these, see below) and thus there is a good overlap between the two layers of the corpus.

1) The node table and the appropriate edge table were loaded into Gephi. Nodes from the node table that do not feature in the edge table (i.e., have a degree of 0) were removed so as not to be displayed as isolated nodes.
2) Node color was adjusted to represent the manuscript type (orange: grammatical handbooks containing only the first book of the *Etymologiae;* purple: library books containing the entire *Etymologiae;* green: manuscripts containing excerpts from the first book of the *Etymologiae*). Node size was set to correspond to the number of parallel glosses.
3) The color of the edges was adjusted to correspond to clusters (A: dark blue; B and O: light green; C1–C7: orange; D and P: turquoise; E: dark red; F: brown; G: dark green; I: yellow; M: light blue; N: dark purple; Q: pink; and S: black).[72]
4) The Yifan Hu layout was applied to the network because of its suitability for small undirected weighted networks and good cluster detection. In addition, Geo Layout using the GPS coordinates in the node table was used to produce visualizations sensitive to the regional localization of annotated manuscripts.[73] The position of nodes was further adjusted with the Label Adjust algorithm to give the network graph a cleaner look.
5) The network graph was manually adjusted to accentuate particular connections and make the visualizations more compact, e.g., nodes were moved apart to prevent edge overlap and make segmentation more visible, and isolated components were moved closer to the main component to decrease the size of the visualization.

Fig. 5 contains two visualizations of Rnk23-clustered-noX: in Fig. 5a, the layout created with the Yifan Hu algorithm reveals the clustering within the network based on gloss parallelism between annotated manuscripts; in Fig. 5b, the geographical relationships between manuscripts featuring shared glosses is mapped with the Geo Layout algorithm.

72 In several cases, two unrelated clusters that concern unrelated manuscripts were assigned the same color. This was done to reduce the color palette used in visualizations and make them more readable.

73 As we typically cannot pinpoint the location of the glossing of a manuscript more precisely than to a region, the position of the nodes in the geo-located visualization should be considered approximate at best. In most cases, it does not inform us about the relation between specific places, such as early medieval monasteries. However, it is precise enough to allow us to consider the circulation of shared glosses in six regions of the early medieval Latin-writing world mentioned in section 3.3: France, the German area, Brittany, England, northern Italy, and Spain.

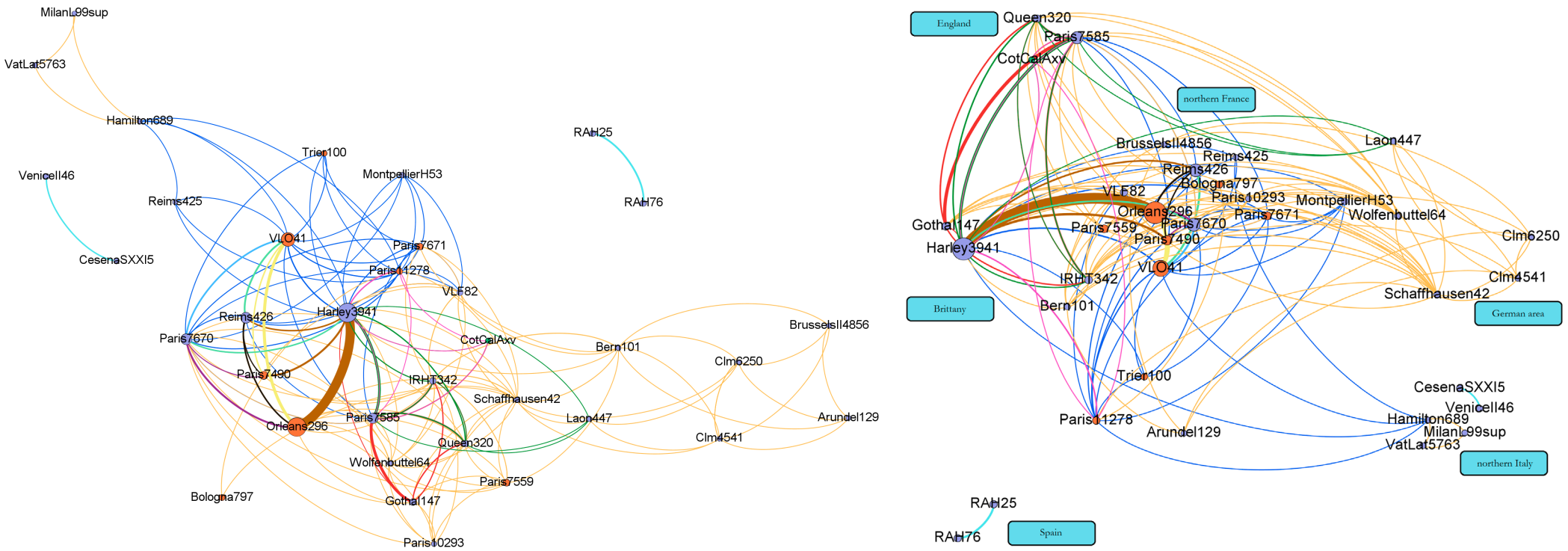

Fig. 5 Rnk23-clustered-noX projected with Yifan Hu (left, 5a) and Geo Layout (right, 5b) algorithms. Clusters displayed, in the order of the proportion of displayed edges include: C1–C7 (orange, 51.7%), A (blue, 23.3%), E (red, 6.1%), G (dark green, 6.1%), F (brown, 3.3%), Q (pink, 3.3%), I (yellow, 1.1%), N (dark purple, 1.1%), S (black, 1.1%), B (light green, 0.6%), D (turquoise, 0.6%), M (light blue, 0.6%), N (dark purple, 0.6%), O (light green, 0.6%), and P (turquoise, 0.6%). Node color: library book containing the entire encyclopedia (purple, 26 manuscripts), grammatical handbooks (orange, 6 manuscripts), and excerpts (green, 1 manuscript). Node size corresponds to the number of parallel glosses present in the manuscript.

6.1 Components and the regional glossing patterns

Rnk23-clustered-noX is dominated by a large component in which we find the majority of manuscripts (31 out of 35 nodes), flanked by two small, isolated components constituted by manuscript pairs CesenaSXXI5-VeniceII46 and RAH25-RAH76. Fig. 5b reveals that manuscripts in the large component were annotated in different parts of France, Brittany, England, and the German area. Its densest core corresponds to the area of northern France, where the most important Carolingian intellectual centers were situated.[74] The two isolated components reflect glossing in northern Italy (CesenaSXXI5-VeniceII46) and Spain (RAH25-RAH76). This figure also shows a trio of interconnected manuscripts (Hamilton689-MilanL99sup-VatLat5763) attached weakly to the large component, also glossed in northern Italy (they appear in the upper left corner of Fig. 5a, at the periphery of the main component). Were it not for a single parallel gloss from cluster A in Hamilton689 (blue), these three manuscripts would be separated from the large component.

An examination of the extrinsic evidence shows that the manuscripts in the two isolated components are closely related philologically: VeniceII46 is a direct copy of CesenaSXXI5, and RAH25 and RAH76 are either parent and offspring, or two siblings.[75] Edges connecting these manuscript pairs thus correspond to the copying of glosses from an exemplar to its apographs, a transmission pattern not observed elsewhere in the corpus. Given this connectivity pattern, it does not seem likely that the isolation of the two small components in Rnk23-clustered-noX is solely due to the loss of connection to the large component as a result of the disappearance of manuscripts. Rather, it seems indicative of distinct attitudes to glossing the *Etymologiae* and limited contact between the glossing communities in Spain, northern Italy, and other regions of the medieval Latin-writing world.

While this may not be particularly visible in Fig. 5b, the poor connectivity to the Carolingian glossing communities is characteristic not only of northern Italy and Spain but also of the German area represented in Rnk23-clustered-noX by Clm4541, Clm6250, Laon447, Schaffhausen42, and Wolfenbuttel64. These manuscripts appear close to each other in the base of the segment extending on the right from Fig. 5a, connected by several micro-clusters both mutually and with other manuscripts, mainly from France. Were it not for the micro-clusters, the German area would vanish from this network graph almost entirely.[76] Comparing the 35 nodes displayed in this visualization with the 54 annotated manuscripts

74 Contreni, "The Carolingian Renaissance," 721.
75 Bellettini, "Il codice del sec. IX di Cesena, Malatestiano S. XXI.5," 75–91; Steinová, "Annotation of the Etymologiae of Isidore of Seville in Its Early Medieval Context," 38.
76 The exception is Laon447, which contains glosses from cluster G (see below).

of the first book of the *Etymologiae* in Appendix I, we can note that many manuscripts from northern Italy and the German area are absent from Rnk23-clustered-noX. Even in light of the possibly substantial loss of annotated manuscripts, we can conclude that while the first book of the *Etymologiae* was annotated in all major regions of the early medieval Latin-writing world, the transmission of glosses to the first book of the *Etymologiae* was principally restricted to three regions – France, Brittany, and England – and was most intense in the Carolingian heartland in northern France.

6.2 Three layers of the main component

The general inspection of Fig. 5 revealed certain qualitative differences between specific regions in terms of the nature of the glosses circulating within them (e.g., transmission from an exemplar to an apograph in northern Italy and Spain, as opposed to the prevalence of the micro-clusters in the German area). To take a further step in this visually-supported analysis, we can dissect Rnk23-clustered-noX into layers corresponding to specific clusters and cluster groupings. By plotting them separately, we can better appreciate that these layers show limited overlap, have different network properties, feature glosses with distinct philological profiles, and correspond to different manuscript contexts. They thus appear to reflect distinct historical circumstances of transmission and regional trends.

We can recognize three layers in Rnk23-clustered-noX. The most prominent of these is the layer of glosses assigned to the micro-clusters C (Fig. 6a, 51.7% of edges of Rnk23-clustered-noX, three parallel edges with cluster A, seven parallel edges with other clusters). Manuscripts containing these glosses were annotated in all the regions mentioned above, apart from Spain. Most contain no other glosses to the *Etymologiae,* although some (Schaffhausen42) attracted glosses from multiple micro-clusters. Looking at the manuscript context of their transmission, we can understand why these glosses were transmitted in isolation rather than as parts of clusters and are attested in regions in which glosses to the *Etymologiae* do not otherwise seem to have circulated widely. Most appear fossilized or semi-fossilized in the main text, meaning that they passed or survived in the process of passing from the white to the black space of a manuscript.[77] Indeed, some of the connections visible in Fig. 6a are due to the fact that medieval scribes did not discern these fossilized and semi-fossilized glosses as glosses, but copied them as a part of the main text, as corrections, or as variant readings. The fossilization indicates their older age relative to the annotated manuscripts that preserve them, which mostly date from the ninth century. Since some of their witnesses are early (e.g., BrusselsII4856 copied and annotated at the end of the

77 On gloss fossilization, see the introduction of Steinová and Boot, "The Glosses to the First Book of the Etymologiae of Isidore of Seville"; Stagni, "Nell'officina di Paolo Diacono?"

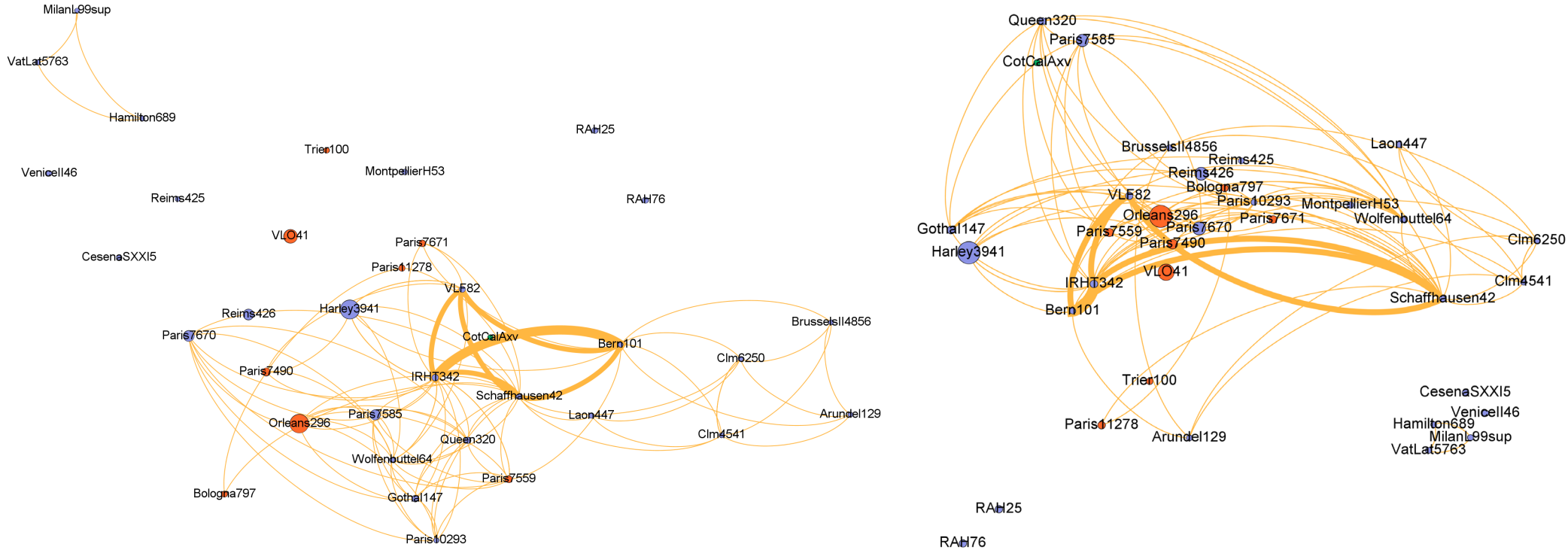

Fig. 6a The layer of Rnk-23-clustered-noX network corresponding to micro-clusters C1-C7 (orange). Layouts: Yifan Hu (left) and Geo Layout (right).

eISSN: 2535-8863
DOI: 10.25517/jhnr.v9i1.198

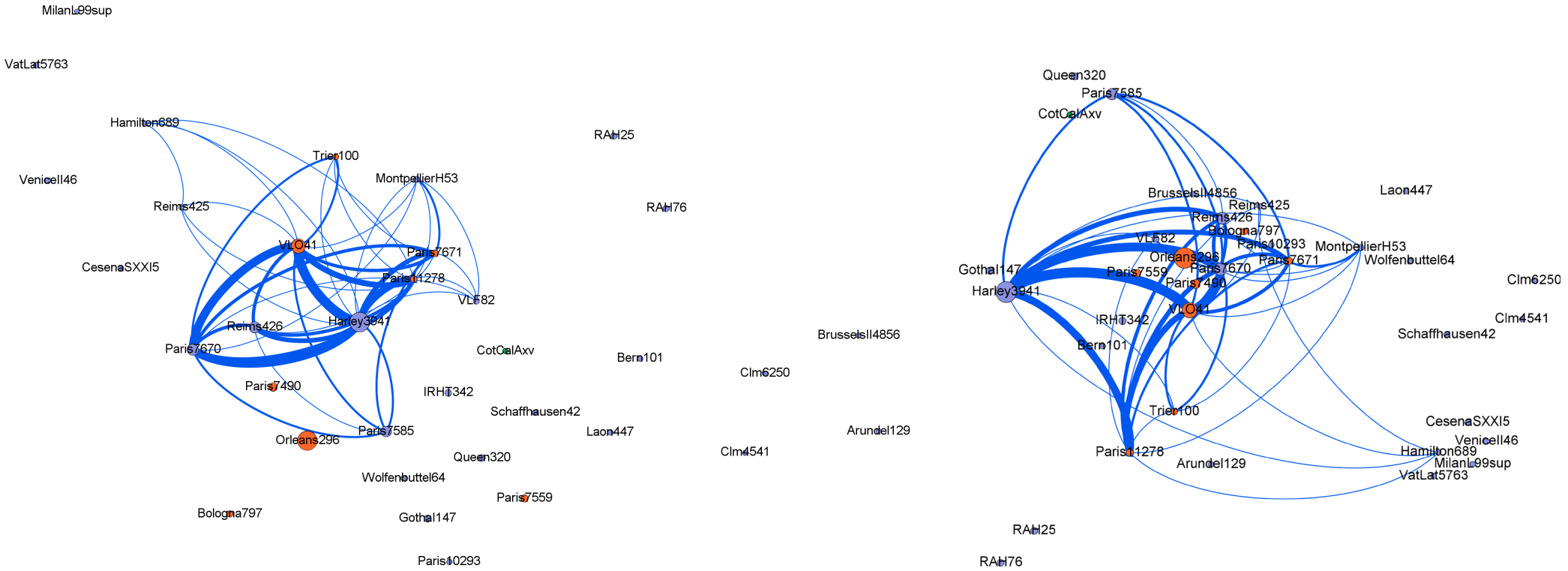

Fig. 6b The layer of Rnk-23-clustered-noX network corresponding to cluster A (blue). Layouts: Yifan Hu (left) and Geo Layout (right).

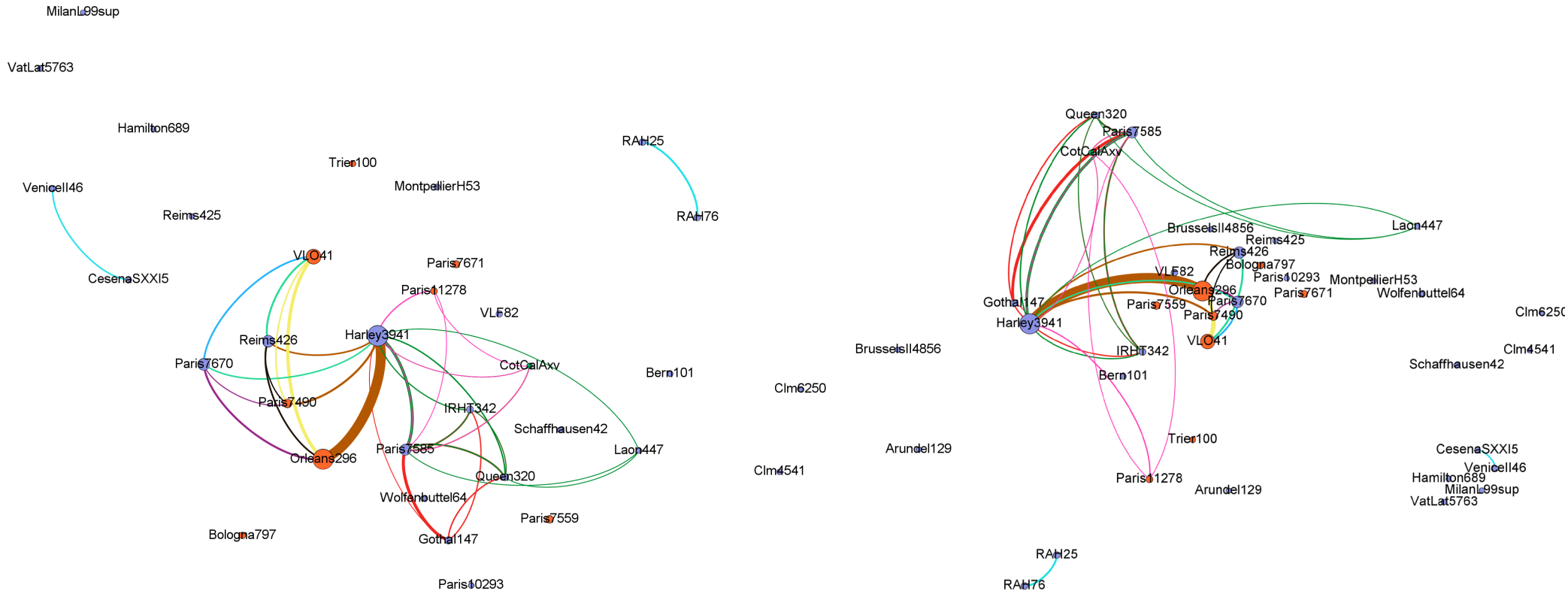

Fig. 6c The layer of Rnk-23-clustered-noX network corresponding to clusters other than A and C1-C7 (other colors). Layouts: Yifan Hu (left) and Geo Layout (right).

eighth century in Corbie, and MilanL99sup copied and annotated in the second half of the eighth century in Bobbio), and given their wide diffusion range, which also presupposes a long period of transmission prior to the earliest attestation, the micro-clusters look like a remnant of pre-Carolingian glossing. We know very little about glossing before the year 800, but we can presuppose it to have taken place given the popularity that the *Etymologiae* had already enjoyed in the Latin-writing world from the seventh century.[78] These glosses must have been more numerous and had a similar transmission dynamics as other glosses analyzed in this study (e.g., transmission in batches rather than in insolation and independently from the substrate text). However, only a small number of them survived due to fossilization. Once embedded into the black space, moreover, the patterns of co-occurrence based on the micro-clusters mirror the transmission patterns of the substrate text (the *Etymologiae*), rather than conforming to what can be expected of glosses. The extremely fragmentary state in which these witnesses of pre-Carolingian engagement with the *Etymologiae* reach us means we cannot reconstruct their context of origin or the direction of their diffusion.[79]

The second layer of glosses, corresponding to cluster A (Fig. 6b, 23.3% of the edges of Rnk23-clustered-noX, three parallel glosses with micro-clusters and six parallel edged with other clusters), display similarities with the layer constituted by micro-clusters C1–C7. Here, too, we are looking at many manuscripts that share a small number of glosses and sometimes do not transmit any other annotations. However, glosses belonging to cluster A appear consistently in the white space. Looking at their manuscript context, we can note that they represent a coherent set of annotations to the first three chapters of the first book of the *Etymologiae,* i.e., the very beginning of Isidore's encyclopedia.[80] They also inhabit a different geographical range, principally occurring in manuscripts from France, an indication that they may have originated in this region. Given their pattern of occurrence in Rnk23-clustered-noX, they may also be a remnant of an entity distinct in its age and character from other clusters, such as a larger body of glosses to the *Etymologiae,* of which only the opening sections remain due to the hazards of transmission. However, this entity is probably not as old as the glosses belonging to micro-clusters C1–C7, and perhaps not older than the early ninth century, since glosses from cluster A are not fossilized.

78 Bischoff, "Die europäische Verbreitung"; Ryan, "Isidore amongst the Islands."

79 Nevertheless, extrinsic clues indicate that some micro-clusters may have originated in the insular environment. Micro-cluster C4, for example, consists of citations from pseudo-Isidorean *De vitiis et virtutibus* that survive fully only in an Irish manuscript; see Schindel, *Die lateinischen Figurenlehren des 5. bis 7. Jahrhunderts und Donats Vergilkommentar;* Schindel, "Zur Datierung des Basler Figurentraktats (cod. lat. F III 15d)."

80 Steinová and Boot, "The Glosses to the First Book of the Etymologiae of Isidore of Seville." https://db.innovatingknowledge.nl/edition/#left-II.

Finally, we can establish a separate layer consisting of the other major clusters (Fig. 6c, 21.51% of the edges of Rnk23-clustered-noX, seven parallel glosses with micro-clusters and six parallel glosses with cluster A). While different clusters appear in this layer, they share certain commonalities. For example, manuscripts in this layer share glosses with a smaller number of manuscripts than in most other layers (thus, the average degree of this layer is 2.571, in contrast to 5.31 for layer C1–C7, and 10.286 for Rnk23-clustered-noX as a whole), although these connections are stronger (thus the average edge weight of this layer is 11.16, in contrast to 2.31 of layer A, 1.08 of C1–C7, and 3.88 of Rnk23-clustered-noX as a whole). Except for Laon447, which was annotated by a group of Irish and Carolingian scholars in Mainz,[81] all manuscripts belonging to this layer were annotated in France (particularly in the north), England, or Brittany.

We can further recognize two regional segments of this layer. First, we find manuscripts annotated in England and Brittany (CotCalAxv, GothaI147, Harley-3941, Paris7585, Queens320), as well as manuscripts annotated on the continent in insular-influenced milieus (IRHT342, Laon447) concentrated to the right of Harley3941 in Fig. 5a. These manuscripts are interconnected by three clusters, each common to at least four manuscripts: E (red, 50 glosses), G (dark green, 30 glosses), and Q (pink, 7 glosses). E and G are absent from manuscripts annotated in France, while Q appears in Paris11278, a manuscript annotated in southern France or northern Italy, which should be perhaps considered to reflect an insular influence on the grounds that it contains glosses from Q. All of the manuscripts mentioned above, apart from CotCalAxv and Paris11278, are codices of the complete *Etymologiae* into which glosses were copied. These insular clusters represent glosses that circulated in the early Middle Ages only or predominantly in areas under the insular influence, and are in all likelihood of insular origin.[82]

The parts of this layer corresponding to the main component positioned to the left of Harley3941 in Fig. 5a include, aside from this codex, manuscripts annotated in northern France, the area of most intense Carolingian intellectual activity. They display specific features when compared with the insular segment of this and other layers. First, the manuscripts belonging to this Carolingian Frankish segment feature clusters common to only one to three other manuscripts, but share glosses with many manuscripts in this manner.[83] Furthermore, except for the highway cluster F (brown, 157 glosses, 73 of which are rank 3), none of the edges visible in this part of the layer are constituted by more than five glosses with

81 Calloni, "Allegorizzare le 'Etymologiae': l'irlandese Probo e gli estratti esegetici del codice Laon BM 447."

82 Compare with Steinová, "Annotation of the Etymologiae of Isidore of Seville in Its Early Medieval Context," 29–37.

83 Orleans296/Paris7490 contains glosses from four clusters and Paris7670, Reims426, and VLO41 from three clusters.

rank 3, although glosses with rank 2 add significant weight to all of them. Clusters M (light blue, 17 glosses) and S (black, 21 glosses) contain only two glosses with rank 3, but M includes thirteen glosses and S fifteen glosses with rank 2, and cluster B (light green, 18 glosses) is built from three glosses with rank 3 but fourteen glosses with rank 2. By contrast, the insular clusters are constituted by more glosses with rank 3 than with rank 2 (G, with fifteen glosses with rank 3 and fourteen with rank 2), or only by glosses with rank 3 (E and Q). Finally, the only edges in Rnk23-clustered-noX that are constituted by no glosses with the highest particularity rank 3, corresponding to clusters I (yellow, 54 glosses) and N (dark purple, 29 glosses), appear in this segment.

Interpreting the last two clusters in their network context is difficult. On the one hand, we cannot dispel the possibility that they are phantoms, rather than historical entities since they are constituted only by glosses with rank 2; on the other hand, they are the second (I) and fifth (N) heaviest clusters in the dataset, constituted by more glosses than eight clusters whose genuineness as transmitted units is not in doubt. Cluster I, moreover, displays a peculiarly scattered pattern of gloss distribution in its two witnesses (Orleans296/Paris7490 and VLO41). In contrast to clusters constituted by many glosses with rank 3, E, F, and G, in which glosses appear concentrated in specific chapters of the first book of the *Etymologiae,*[84] the glosses in cluster I are spread across Orleans296/Paris7490 and VLO41 rather uniformly. As a result, they are interspersed by both isolated and other parallel glosses to such an extent that cluster I is invisible to the human eye, unlike clusters E, F, and G, which can be partially detected via close reading. To a lesser extent, the scattered pattern of gloss distribution also characterizes cluster N and other clusters from the Carolingian Frankish segment.

This distribution pattern, as opposed to the appearance of glosses in blocks, could be expected to arise as a result of either spontaneous composition or random parallelism. Whether we should assume that neither I nor N are genuine clusters, as is perhaps the case with parts of clusters B, M, O, and S, depends on how plausible we find the idea that spontaneous composition or random parallelism could generate phantom sets constituted by a large number of glosses, in particular more than ten glosses, the demarcation assumed for clusters in section 3.2 (see section 7). In the absence of relevant data, we can rely on traditional philological reasoning that tells us it is unlikely that cluster I, with its 54 glosses, and cluster N, with its 29 glosses, could be phantoms in their entirety. Instead, we can seek the explanation for the specific features of these two clusters, and others, in the historical processes that gave rise to them and yet that seem distinct

84 83% of glosses from cluster F appear in chapters 36–44 in Harley3941 and Orleans296, 64% of glosses from cluster E appear in chapters 37–40 in Gotha1147 and Paris7585, and 50% of glosses belonging to cluster G can be found in chapters 32–37 in Harley3941 and Paris7585.

from those described in this article so far. For example, it can be pointed out that manuscripts sharing these two clusters are grammatical handbooks rather than library books (nodes colored orange) and display paleographic features consistent with the use in or design for school use.[85] Since we have learned in section 3.3 that instruction seems to have been an important stimulus for the production of glosses to the first book of the *Etymologiae* in northern France, the peculiarities of clusters I and N, and of the Carolingian Frankish segment more broadly, may reflect the transmission of glosses in the context of instruction (e.g., involving oral transmission or selective collection with the aim of reuse for teaching).

6.3 Hubs in the network

The cluster-detection algorithm applied in Fig. 5a makes it evident that one manuscript sits at the intersection of all three layers described in section 6.2: Harley3941. As was established in the previous section, this codex stands out among manuscripts containing glosses to the first book of the *Etymologiae* due to its extensive gloss parallelism and should be considered a hub. Fig. 6c reveals that Harley3941 also acts as a bridge between otherwise disconnected insular and Carolingian Frankish segments of the large component in Rnk23-clustered-noX. Paleographic and philological evidence corroborate network analysis and visualization.[86] Harley3941 is a manuscript of the entire *Etymologiae* that was produced and annotated at the end of the ninth or the beginning of the tenth century in Brittany. Glosses from several different clusters can be shown to have been copied into it at the time of its production and during the following century, including a batch added by a single hand that corresponds to cluster G.[87] The latter act of copying suggests that this Breton codex was used to collect glosses of diverse origin, including material known only from the Carolingian environment or the insular world. The network properties of Harley3941 can be matched to its real-world status as what may be termed a depository manuscript, i.e., a codex dedicated to the accumulation of glosses for preservation and potential reuse.

While Harley3941 is the most evident hub in Rnk23-clustered-noX, four other manuscripts were flagged as potential hubs in section 5.4: IRHT342, Paris7585, Paris7670, and Schaffhausen42. Of these, Schaffhausen42, a codex produced in the second quarter of the ninth century in Mainz and annotated in the second half of the same century in St. Gallen, can be excluded from the list. It holds a prominent place in the layer of Rnk23-clustered-noX constituted by micro-clusters, as it contains the highest number of parallel glosses belonging to these micro-clusters (6) and is connected to the highest number of manuscripts trans-

85 Steinová, "Annotation of the Etymologiae of Isidore of Seville in Its Early Medieval Context," 13–15.
86 Ibid., 31–33.
87 Ibid., 17.

mitting them (18), but is otherwise not particularly central to the entire network. By contrast, Paris7585 and Paris 7670 feature in all three layers distinguished in Fig. 6, and IRHT342 appears in two of the three layers, not containing glosses from cluster A. Paleographic and philological evidence also identify IRHT342, copied in the tenth and annotated in the following century and a half in an unknown location, but showing a clear affinity to the insular world in its collection of glosses, and Paris7585, produced in France and annotated in Canterbury in the second half of the tenth century, as depository manuscripts.[88] Both assemble glosses from two insular clusters, E and G, which are transmitted separately in older manuscripts, most notably in the Breton GothaII47 (the most important witness of E) and Harley3941 (the most important witness of G). Furthermore, like Harley3941, these two manuscripts were annotated during or shortly after their production, and the glosses they contain can be shown to have circulated at least a century before the copying of the manuscripts; they are library books rather than schoolbooks, and they represent rare examples of manuscripts that attracted glosses not only to the first but to all books of the *Etymologiae*. The case of Paris7670 is more intriguing, as we lack paleographic and philological evidence that would classify it as a depository manuscript. Nonetheless, it has the properties of such a manuscript (e.g., it is a library book). Its network properties may be an indicator that it should also be considered a depository manuscript.

6.4 Gloss parallelism and geographic distribution

We can conclude the visual inspection of Rnk23-clustered-noX by visually comparing Figures 5a and 5b, noting which nodes from specific layers and segments are pulled together by the Yifan Hu algorithm even if they do not represent geographically close manuscripts, and which nodes are pushed apart even if they represent geographically close manuscripts. Such a discrepancy between geographic proximity and gloss parallelism is particularly notable in the case of Orleans296 and VLO41, the two manuscripts with the highest number of parallel glosses after Harley3941 (301 and 191, respectively) and the two most densely annotated surviving manuscripts of the first book of the *Etymologiae* (768 and 682 glosses, respectively). While these two manuscripts are connected by cluster I, the second heaviest edge in Rnk23-clustered-noX, the Yifan Hu algorithm places them on opposite sides of the Carolingian Frankish segment of the main component because of their otherwise distinct connectivity to other parts of the network. Nevertheless, both manuscripts have ties to the same location: Fleury, a monastery in central France. VLO41 was annotated there at the end of the ninth or during the early tenth century. Orleans296 was present in Fleury from the tenth century at the latest, and probably earlier.[89]

88 Ibid., 31–33 and 35.
89 Ibid., 50–54.

Another manuscript pair displaying a similar discrepancy is GothaI147 and Harley3941. These two codices, produced and annotated in Brittany, share only a single gloss with rank 3, although they contain many other glosses with this rank shared with manuscripts in England and France. In other cases, such as the Spanish RAH25 and RAH76, the northern Italian CesenaSXXI5 and VeniceII46, which we have established reflect the transfer of glosses from an exemplar to an apograph, we can observe a correlation between gloss parallelism and geographic proximity. This is also true for CotCalAxv, Paris7585, and Queens320, which seem to have been annotated in Canterbury[90] and share glosses from clusters E and G. A visual examination of Rnk23-clustered-noX is insufficient to reach a definitive conclusion about the relationship between geography and gloss parallelism. To explore this dimension of co-occurrence networks, we have to deploy a different strategy (see section 7).

6.5 The network visualization in perspective

In conclusion to this section, we can reflect on the utility of the network visualization for analyzing the corpus of glosses to the first book of the *Etymologiae*. While network visualization does not replace proper network analysis, this section has hopefully demonstrated how useful it is to complement the latter with the former, in particular as an exploratory technique that could direct scholars to avenues for further investigation[91], and how much can be gained from interrogating network visualization against the backdrop of the available extrinsic evidence. Firstly, the visualization exercise suggested that certain network properties can be a good match for extrinsic properties. To name but two examples, the detection of hubs could help us identify manuscripts used for the collection and preservation of glosses; and as specific network patterns seem to have been generated by different historical processes of gloss generation and transmission, their detection and analysis could provide us with crucial insights into how glosses were produced and circulated in medieval Latin-writing Europe. In addition, the visualization brought home how multilayered the corpus of glosses to the first book of the *Etymologiae* is, strengthening the observations made in section 5.4 based on the analysis of average and median degrees. Finally, it allowed for the inclusion of the chronological and geographical dimensions of the data, which were not directly tapped within the network analysis. Especially insofar as a more rigorous analysis of the geographical relationships between the witnesses of a gloss corpus is not feasible within a specific research project, the geo-sensitive visualizations can supply scholars with preliminary observations about regional trends and patterns.

90 Ibid., 31–32 and 35–36.

91 Compare with Fernández Riva, "Network Analysis of Medieval Manuscript Transmission. Basic Principles and Methods," 38; Lemercier and Zalc, *Quantitative Methods in the Humanities: An Introduction,* 129–36.

7. Avenues for further research and the limits of the method

The utility of the network-based approach outlined in this article is not restricted to the analysis of the network properties of co-occurrence networks (section 5), nor the visualization and interpretation of these networks in light of available scholarly evidence (section 6). It can be developed further to directly address specific research questions, as some of the network properties and models can serve as relevant proxies for historical processes or circumstances we wish to study. Unfortunately, the scope of this article does not allow us to develop such applications. Nevertheless, some potential uses of the network-based approach for the study of glosses and avenues for expanding and refining this method in the future can be sketched here.

- **Organicity of gloss corpora:** As we have seen in section 5.4, the degree of co-occurrence networks of gloss parallelism appears to be tied to the multilayered character of the corpus of glosses to the first book of the *Etymologiae,* and thus its organicity. It would be expedient to further test the utility of degree as a quantitative measure of the organicity of gloss corpora by comparing the degree distributions of co-occurrence networks constructed from different types of material (corpora of glosses with different expected levels of organicity/systematicity and standard textual traditions). If it turns out to be a good proxy for organicity, it could provide a basis for assessing gloss corpora quantitatively, as opposed to giving them a label based purely on qualitative assessment, and allow for a comparison of different gloss corpora.
- **The extent of spontaneous composition and random gloss parallelism:** It would be useful to model gloss parallelism due to spontaneous composition and randomness in order to better understand how co-occurrence networks due to transmission differ from those due to spontaneous composition and randomness, and what may be the extent and type of distortion that we should expect to observe in a co-occurrence network corresponding to real-world data. In both cases, one model that comes to mind is a type of random network model, a random intersection graph following hypergeometric distribution.[92] It is vital to further develop this or other random network models

92 On random network models in general, see Barabási, *Network Science,* chap. 3. Hypergeometric distribution corresponds to a situation in which objects (glosses) taken from a certain pool (e.g., Latin lexicon or its parts) are assigned to containers (manuscripts), either entirely randomly (random gloss parallelism) or according to specific criteria with an element of chance (spontaneous composition). Depending on the size of the pool and the number of objects assigned to a container, we should observe that the same objects are assigned to different containers and thus form a basis for the construction of co-occurrence networks at a certain rate with different probabilities, i.e., that co-occurrence networks constructed based on the co-occurrence of the same objects in different containers tend to have properties within certain ranges, and display properties in other ranges with

that could approximate random gloss parallelism and test their utility for modeling the historical process of spontaneous composition.

- **Identifying and distinguishing different transmission processes:** In the two analytical sections of this article, sections 5 and 6, it was noted that certain network patterns observed in our co-occurrence networks seem to reflect different types of transmission (e.g., copying from an exemplar to an apograph, transmission of fossilized glosses within the main text, and the collection of glosses in a depository manuscript), and that different transmission processes can be expected to generate different network properties and elements. For this reason, it could be productive to develop network models that simulate different types of transmission, in particular, genealogical transmission typically represented as a stemma and the collection of glosses in a depository manuscript, just as we can establish a network model to approximate spontaneous composition and random gloss parallelism.[93] In this way, network analysis could help to distinguish different transmission processes from one another in cases when extrinsic evidence is lacking.
- **Geographical distance as a factor in gloss parallelism:** The relationship between geographical distance and gloss parallelism could be explored more rigorously than through a visual comparison of network graphs. Taking the data from the corpus explored in this article, we could, for example, compare the distance between any two manuscripts containing parallel glosses in km (derivable from their GPS coordinates) with the extent of their gloss parallelism (represented by the number of parallel glosses or glosses of certain ranks they share) and plot them against each other.[94]
- **Gloss-hopping:** In this study, the network-based approach was used to examine the internal dynamics and structure of a single gloss corpus. In this respect, we did not stray from the traditional scholarly paradigm, which treats glosses to each text separately, i.e., as unique and distinct from those to other texts. It is still uncommon for scholars to acknowledge that such boundaries may be due to our modern scholarly perceptions and editorial needs rather than to medieval annotation practices.[95] As the network-based approach treats corpora and collections of glosses as pools, it can be used to trace gloss parallelism

---

a negligible probability. On hypergeometric distribution, see Pitman, *Probability,* 127. The random intersection graph model was developed in Singer, "Random Intersection Graphs."

93 See Hoenen, "The Stemma as a Computational Model," 229–30.

94 Since, as was explained in section 4.2, the GPS coordinates are only used to approximate the region of the glossing of manuscripts, we can hope to uncover only very general trends, for example that glosses in manuscripts from a certain region tend to be more similar than glosses in manuscripts coming from different regions. Alternatively, we could restrict ourselves to using manuscript pairs for which we know the precise location of origin (e.g., Canterbury for Paris7585) to make this experiment more precise.

95 For a rare example of the awareness of this issue, see Teeuwen, "The Impossible Task of Editing a Ninth-Century Commentary," 200–202.

> across the boundaries of text-defined corpora, transcending the compartmentalization of glosses by text to assess to what extent glossing was text-bound, and whether other boundaries may be more relevant for the understanding of medieval reality (e.g., because of the role of memorization of glosses as self-sufficient units). As long as we can formulate criteria for postulating gloss parallelism across different languages, we can similarly use the network-based approach to study trans-linguistic gloss parallelism, a phenomenon noted by scholars of vernacular glossing.[96]

After discussing the potential utility of the network-based approach to glossing, it is fitting to offer remarks on its general limits and specific avenues for improvement. First, it is hard to assess how robust the conclusions that can be obtained via this approach are in light of the loss of historical material on which they are built, a problem common to historical network research.[97] The fragmentary survival of annotated manuscripts should make us particularly cautious about interpreting the absence of evidence (e.g., in the form of isolated components or weakly connected segments in a network) as the evidence of absence without sufficient support of extrinsic evidence, and to keep in mind that the observed results represent minimalistic conclusions (e.g., gloss parallelism and the extent of transmission can always be assumed to have been higher than observed).

We also need to remember that the co-occurrence networks explore gloss parallelism, rather than gloss transmission or the social networks that facilitated this transmission. While gloss parallelism may reflect the historical process of transmission, obtaining information relevant for establishing transmission networks from co-occurrence networks may not be possible, as the information they record is, by rule, not rich enough for this purpose and is represented in a manner not compatible with, for example, constructing a stemma as a particular type of transmission graph.[98] Furthermore, as was explained in section 2.2, gloss parallelism can have different causes, and distinguishing parallel from shared glosses must be done based on criteria that are external to network analysis and requires a substantial degree of domain knowledge. Even if additional modes of data preprocessing other than particularity ranking and gloss clustering were applied to an organic corpus of glosses, it is unlikely that we could filter out all of the noise from the co-occurrence networks that can be constructed from this data. On the contrary, the more restrictive the criteria, the more likely it is that we will also lose relevant information. It remains to be seen whether the method can be fur-

---

96 Moran, "Language Interaction in the St Gall Priscian Glosses," 134–39; Lambert, "L'étude des gloses"; Bauer, "Different Types of Language Contact in the Early Medieval Celtic Glosses."

97 Knappett, "Networks in Archaeology," 28–29.

98 The presence of complete graph elements in the co-occurrence network, in particular, is irreconcilable with a transmission network model.

ther improved to target noise more efficiently without diminishing the quality of the data.

As for some of the blind spots of the method as described in this article, as we have seen in section 6.2, it may be valuable to pay attention to the gloss distribution within collections of annotation, as a substantially diffused distribution pattern is consistent with spontaneous composition and random gloss parallelism, and may therefore provide an argument for assuming gloss parallelism due to processes other than transmission. The method could be further developed to account for the relative position of glosses within a collection of annotations and to incorporate information about the paleography of the glosses to better represent the layered nature of certain collections of annotations. The current method also does not work with the temporal aspect of gloss parallelism, even though such information is available and could be used to create network graphs that account for this property, in the same way that the geographical aspect of gloss parallelism was explored above.[99] Finally, edges in the co-occurrence networks constructed and examined in this article were made undirected. However, it should be possible to incorporate directionality into network analysis and visualization, provided it is made clear that it does not represent the direction of transmission of material from one specific manuscript to another, but rather a general direction of transmission of material.

8. Conclusion

This article outlines a network-based approach that allows us to study organic corpora of glosses in their complexity. In contrast to traditional scholarship, which emphasizes particular forms of sequential textuality and transmission facilitated by copying from an exemplar to an apograph, the network-based approach allows us to operate on the subtler level of individual glosses and to take gloss parallelism, rather than transmission (or a specific type of transmission), as a point of departure. The chief advantage of the network-based method is that it allows researchers to work with historical material in the form in which it came down to us, without having to either adopt preliminary assumptions about that corpus (e.g., that all instances of philological similarity within the corpus are due to transmission or that glosses were transmitted as standard texts), or to discard some information insofar as it cannot be fitted into the narrow criteria imposed by traditional methods. At the same time, by adopting the strategies described in section 2 of this article (particularity ranking and gloss clustering), the network-based approach can account for transmission as a specific historical process of interest and larger textual units than glosses (clusters). As a result, we can fully

99 On temporality in historical network research, see Knappett, "Networks in Archaeology," 67–70.

map the internal structure of a corpus of glosses, keeping its multilayered character and heterogeneity in the picture and not sacrificing certain elements of the corpus just because they cannot be considered gloss traditions, families, or commentaries. A network can even serve as an editorial model, and a network graph can provide an alternative visualization strategy to a stemma.[100]

As for the specific conclusions that can be drawn about the glossing of the first book of the *Etymologiae* following the analysis carried out in sections 5 and 6, the most significant properties of this corpus seem to be its heterogeneity and regionality. As far as the surviving evidence can be assumed as being broadly representative of the character and circumstances of the glossing of the *Etymologiae* in the early Middle Ages, there appears not to have been any dominant gloss family or tradition transmitted by a large number of witnesses (although at least one identified cluster appearing in two manuscripts, F, stands out due to the large number of glosses it contains). Rather, we discerned thirteen different clusters of glosses that seem to reflect distinct glossing efforts undertaken in the ninth and the tenth centuries, and seven micro-clusters, which are probably remnants of glossing predating the Carolingian period.

Most clusters seem to have circulated regionally, such as E and G in the insular world, A and I in France, P in Spain, and D in northern Italy. Only the micro-clusters are diffused much more widely across the early medieval Latin-writing world, which is likely due to their substantial age and fossilized state. The intensity of the glossing seems to have been highest in northern France and lowest in the German area, from which no gloss cluster originating in the ninth and the tenth centuries survives. Importantly, the regionality of the glossing of the first book of the *Etymologiae* is also matched by transmission patterns. In both northern Italy and Spain, we can only evidence the transmission of glosses by copying from an annotated exemplar to its apographs; in the insular world and Brittany, the presence of three hubs (IRHT342, Harley3941, Paris7585) revealed in sections 5.4 and 6.3 is consistent with a preference for collecting glosses into a manuscript that serves as their depositories; and in northern France, the transmission of glosses may have been driven by instructional needs, as it has a different network pattern. A specific place in the landscape of the glossing of the first book of the *Etymologiae* should be accorded to Brittany, for which glosses circulating both in the insular world and in northern France can be shown to have been available and collected. Given the age and character of Harley3941, we should assume that Brittany benefited from glossing taking place in both northern France and the insular world in the previous 150 years.

Some of the methodological and theoretical points made in this article that deserve emphasizing are:

100 Steinová and Boot, "Editing Glosses as Networks: Exploring the Explorative Edition."

- **As gloss parallelism cannot always be attributed to transmission, it is necessary to engage in data pre-processing to reduce noise before carrying out network analysis.** The distortive effect of spontaneous composition and random gloss parallelism is demonstrated in section 5.4. It could be useful to model how much gloss parallelism should be taken as a baseline due to spontaneous composition and randomness in a co-occurrence network constructed following the principles outlined in this article. Given the information quality of data used for constructing Rnk23-clustered-noX in sections 5 and 6, the particularity ranking and gloss clustering outlined in section 2 seem to be efficient strategies for noise elimination.
- **While trivial parallel glosses cannot be considered shared by default, many of them were likely transmitted.** The corpus studied in this article, as many organic corpora of glosses, is constituted mostly by glosses too trivial to be treated as transmitted by default (i.e., assigned particularity ranks 1 and 2 in section 3.1). Gloss clustering can help determine whether they may have been part of a transmitted package of glosses. Even if we cannot claim that every trivial gloss in a given cluster must have been transmitted, they can be considered broadly indicative of transmission as long as they: a) feature in clusters constituted primarily or also by glosses particular enough to be considered transmitted (i.e., assigned particularity rank 3 in section 3.1); or b) appear in a cluster in volumes too large to be explainable by spontaneous composition and randomness alone; or c) we possess concrete extrinsic evidence that substantiates their transmission (as in the case of cluster G in Harley3941).
- **In highly organic corpora, glosses can be assumed to have circulated on their own or in very small units.** This has been shown in sections 5.6 and 6.2, particularly on micro-clusters C1–C7. Overall, the examination of the corpus of glosses to the first book of the *Etymologiae* has shown that glosses could be transmitted in relatively small units, i.e., in the range of five to ten glosses. This is probably partially an effect of the loss of evidence, as identifiable clusters may correspond to what had once been larger batches of glosses. Nevertheless, some glosses to the first book of the *Etymologiae* were evidently transmitted in the early medieval Latin-writing Europe in very small units or isolation (e.g., clusters A and Q). Therefore, we should be wary of overfocusing on large or easily detectable clusters just because they are more prominent or visible to the human eye. This should also make us wonder what the circulation of glosses in small units or isolation reveals about the character of glossing and gloss transmission in the Middle Ages and the potential 'attrition' of gloss clusters due to manuscript loss.
- **Some gloss clusters are poorly visible or invisible using traditional methods.** While several gloss clusters in the corpus used as a demonstrative case in this article could be detected and partially described via close reading (e.g., E, F, and G), some clusters are likely to escape traditional methods because of their relatively small size, low particularity, small number of witnesses, and dispersed gloss distribution in manuscripts. These include cases that involve such large volumes of parallel glosses (clusters I and N) that they can-

not be dismissed as phantoms conjured by spontaneous composition or random processes. The network-based approach described in this article may be particularly valuable for identifying 'invisible' gloss clusters. It is noteworthy that within the corpus studied in this article, these 'invisible' clusters predominate in the region of northern France, in which the extrinsic evidence suggests that glosses to the first book of the *Etymologiae* circulated in an instructional context. The two features may be interconnected, indicating that traditional methods may be blind to transmission processes that are particularly interesting to study.

- **Co-occurrence networks with a high concentration of nodes with high degrees may provide evidence for the process of the accumulation of glosses.** In section 5.4, it was shown that the co-occurrence networks constructed from data introduced in section 3 have relatively high average and median degrees, and in section 6.3 that nodes that appear as hubs in the constructed co-occurrence networks correspond to manuscripts that bear extrinsic signs of having been designed or used for collecting glosses. In one case (Paris7670), recognition of hubs may have even identified a depository manuscript, for which we lack extrinsic clues. It was proposed that the two properties may relate to the multilayered character of the corpus. If this hypothesis can be substantiated, the identification of hubs could be a quick way to trace annotated manuscripts in which glosses were collected, and the degree distribution could help us to establish to what extent the accumulation of glosses played a role in the evolution of a gloss corpus.
- **Different transmission processes may generate different network patterns.** In sections 5.4 and 6.1–2, we distinguished four network patterns that may be associated with different transmission processes. First, the copying of glosses from an annotated exemplar to its apograph (a process resembling the transmission of standard texts) corresponded to relatively thick edges between a small number of manuscripts that are isolated from other manuscripts due to their non-cumulative character. Second, the transmission of glosses within the main text due to their fossilization generated large, complete graph components with edges of minimal thickness. Third, the collection of glosses and copying of batches of glosses from one manuscript to another generated hubs. Finally, the transmission of glosses in an instructional context was connected with a network pattern in which many manuscript pairs or triplets are mutually connected with relatively light edges. These and perhaps other network patterns need to be examined further to ascertain whether they can be used to detect specific transmission processes with network analysis.

The network-based approach to glossing has much to offer. In the future, the network-based approach described and demonstrated in this article will hopefully be extended to new corpora of glosses, and thus its utility tested. In the process, the validity of the conclusions articulated here shall be ascertained. The more corpora that are probed with network-based methods, the more we are likely learn about the historical processes of generation and transmission of glosses,

and the better we will understand the character of glossing in Latin-writing medieval Europe and beyond.

9. References

Andrée, Alexander. "Anselm of Laon Unveiled: The *Glosae svper Iohannem* and the Origins of the *Glossa ordinaria* on the Bible." *Mediaeval Studies* 73 (2011): 217–60.

Barabási, Albert-László. *Network Science.* Cambridge: Cambridge University Press, 2017. http://www.networksciencebook.com/.

Bauer, Bernhard. "Different Types of Language Contact in the Early Medieval Celtic Glosses." *Proceedings of the Harvard Celtic Colloquium* 37 (2017): 33–46.

— "The Celtic Parallel Glosses on Bede's 'De Natura Rerum.'" *Peritia* 30 (2019): 31–52.

— "The interconnections of St Gall, Stiftsbibliothek, MS 251 with the Celtic Bede manuscripts." *Keltische Forschungen* 8 (2019): 31–48.

— "Venezia, Biblioteca Marciana, Zanetti Lat. 349. An Isolated Manuscript? A (Network) Analysis of Parallel Glosses on Orosius' Historiae Adversus Paganos." *Études celtiques* 45 (2019): 91–106.

Bellettini, Anna. "Il codice del sec. IX di Cesena, Malatestiano S. XXI.5: le *Etymologiae* di Isidoro, testi minori e glosse di età ottoniana." *Italia medioevale e umanistica* 45 (2004): 49–114.

Bischoff, Bernhard. "Die europäische Verbreitung der Werke Isidors von Sevilla." In *Isidoriana. Collección de estudios sobre Isidoro de Sevilla,* edited by Manuel Cecilio Díaz y Díaz, 317–44. León: Centro de estudios San Isidoro, 1961.

Brughmans, Tom, Anna Collar, and Fiona Coward. "Network Perspectives on the Past: Tackling the Challenges: Introduction." In *The Connected Past: Challenges to Network Studies in Archaeology and History,* edited by Tom Brughmans, Anna Collar, and Fiona Coward, 3–20. Oxford: Oxford University Press, 2016. https://eprints.soton.ac.uk/432433/.

Buringh, Eltjo. *Medieval Manuscript Production in the Latin West: Explorations with a Global Database.* Leiden: Brill, 2011.

Calloni, Carlo Giovanni. "Allegorizzare le 'Etymologiae': l'irlandese Probo e gli estratti esegetici del codice Laon BM 447." *Filologia Mediolatina* 29 (2022): 113–48.

Chiesa, Paolo. "The Genealogical Method: Principles and Practice." In *Handbook of Stemmatology: History, Methodology, Digital Approaches,* edited by Philipp Roelli, 74–87. Berlin: De Gruyter, 2020.

Conti, Aidan. "A Typology of Variation and Error." In *Handbook of Stemmatology: History, Methodology, Digital Approaches,* edited by Philipp Roelli, 242–52. Berlin: De Gruyter, 2020.

Contreni, John J. "The Carolingian Renaissance: Education and Literary Culture." In *The New Cambridge Medieval History 2: c. 700–c. 900,* edited by Rosamond McKitterick, 709–57. Cambridge: Cambridge University Press, 1995.

— "The Pursuit of Knowledge in Carolingian Europe." In *The Gentle Voices of Teachers: Aspects of Learning in the Carolingian Age,* edited by Richard Sullivan, 106–41. Columbus, OH: Ohio University Press, 1995.

Dionisotti, Anna Carlotta. "On the Nature and Transmission of Latin Glossaries." In *Les manuscrits des lexiques et glossaires de l'antiquité tardive à la fin du moyen âge: Actes du colloque international (Erice, 23–30 septembre 1994),* edited by J. Hamesse, 202–52. Leuven: Fédération internationale des instituts d'études médiévales, 1996.

Fernández Riva, Gustavo. "Network Analysis of Medieval Manuscript Transmission. Basic Principles and Methods." *Journal of Historical Network Research* 3 (2019): 30–49.

Ganz, David. "Book Production in the Carolingian Empire and the Spread of Caroline Minuscule." In *The New Cambridge Medieval History 2: c. 700–c. 900,* edited by Rosamond McKitterick, 786–808. Cambridge: Cambridge University Press, 1995.

Godden, Malcolm R., and Rohini Jayatilaka. "Counting the Heads of the Hydra: The Development of the Early Medieval Commentary on Boethius's Consolation of Philosophy." In *Carolingian Scholarship and Martianus Capella: Ninth-Century Commentary Traditions on* De Nuptiis *in Context,* edited by Mariken Teeuwen and Sinéad O'Sullivan, 363–76. Cultural Encounters in Late Antiquity and the Middle Ages 12. Turnhout: Brepols, 2011.

Guidi, Vincenzo, and Paolo Trovato. "Sugli stemmi bipartiti. Decimazione, asimmetria e calcolo delle probabilità." *Filologia italiana* 1 (2004): 9–48.

Hoenen, Armin. "The Stemma as a Computational Model." In *Handbook of Stemmatology: History, Methodology, Digital Approaches,* edited by Philipp Roelli, 226–41. Berlin: De Gruyter, 2020.

Hoey, Michael. *Textual Interaction: An Introduction to Written Discourse Analysis.* London: Routledge, 2001.

Holtz, Louis. "Glosse e commenti." In *Lo spazio letterario del Medioevo 1: Il Medioevo latino, 3: La ricezione del testo,* edited by Guglielmo Cavallo, Claudio Leonardi, and Enrico Menestò, 59–111. Rome: Salerno, 1995.

Jeauneau, Édouard. "Le Commentaire érigénien sur Martianus Capella (De nuptiis, Book I)." In *Quatre thèmes érigéniens,* 101–66. Montréal: Institut d'études médiévales Albert-le-Grand, 1978.

Kapitan, Katarzyna Anna. "Perspectives on Digital Catalogs and Textual Networks of Old Norse Literature." *Manuscript Studies* 6 (2021): 74–97.

Knappett, Carl. "Networks in Archaeology: Between Scientific Method and Humanistic Metaphor." In *The Connected Past: Challenges to Network Studies in Archaeology and History,* edited by Tom Brughmans, Anna Collar, and Fiona Coward, 21–33. Oxford: Oxford University Press, 2016. https://doi.org/10.1093/9780198748519.003.0007.

Lambert, Pierre-Yves. "Les commentaires celtiques a Bède le vénérable." *Études celtiques* 20 (1983): 119–43.
— "Les commentaires celtiques à Bède le vénérable." *Études celtiques* 21 (1984): 185–206.
— "L'étude des gloses: méthodes et instruments." *Britannia Monastica* 19 (2017): 45–82.
Lapidge, Michael. "The Study of Latin Texts in Late Anglo-Saxon England: The Evidence of Latin Glosses." In *Latin and the Vernacular Languages in Early Medieval Britain,* edited by Nicholas Brooks, 99–140. Leicester: Leicester University Press, 1982.
Lemercier, Claire, and Claire Zalc. *Quantitative Methods in the Humanities: An Introduction.* Translated by Arthur Goldhammer. Charlottesville, VA: University of Virginia Press, 2019.
Moran, Pádraic. "Language Interaction in the St Gall Priscian Glosses." *Peritia* 26 (2015): 113–42. https://doi.org/10.1484/J.PERIT.5.108317.
Moulin, Claudine. "Paratextuelle Netzwerke: Kulturwissenschaftliche Erschließung und soziale Dimensionen der althochdeutschen Glossenüberlieferung." In *Verwandtschaft, Freundschaft, Bruderschaft. Soziale Lebens- und Kommunikationsformen im Mittelalter,* edited by Gerhard Krieger, 56–82. Akten des Symposiums des Mediävistenverband 12. Berlin: De Gruyter, 2009.
Newman, Mark E. J. *Networks.* 2nd ed., Oxford: Oxford University Press, 2019.
Nievergelt, Andreas. "Glossen aus einem einzigen Buchstaben." In *The Annotated Book in the Early Middle Ages. Practices of Reading and Writing,* edited by Mariken Teeuwen and Irene van Renswoude, 285–304. Utrecht studies in medieval literacy 38. Turnhout: Brepols, 2018.
O'Sullivan, Sinéad. "Problems in Editing Glosses: A Case Study of Carolingian Glosses on Martianus Capella." In *The Arts of Editing Medieval Greek and Latin: A Casebook,* edited by Elisabet Göransson, Gunilla Iversen, Barbara Crostini, Brian M. Jensen, Erika Kihlman, Eva Odelman, and Denis Searby, 290–310. Studies and Texts 203. Toronto: Pontifical Institute of Mediaeval Studies, 2016.
— "Text, Gloss, and Tradition in the Early Medieval West: Expanding into a World of Learning." In *Teaching and Learning in Medieval Europe: Essays in Honour of Gernot R. Wieland,* edited by Greti Dinkova-Bruun and Tristan G. Major, 3–24. The Journal of Medieval Latin Publications 11. Turnhout: Brepols, 2017.
Palumbo, Giovanni. "The Genealogical Method: Criticism and Controversy." In *Handbook of Stemmatology: History, Methodology, Digital Approaches,* edited by Philipp Roelli, 88–109. Berlin: De Gruyter, 2020.

Peeples, Matthew A., Barbara J. Mills, Randall W. Haas Jr, Jeffery J. Clark, and John M. Roberts Jr. "Analytical Challenges for the Application of Social Network Analysis in Archaeology." In *The Connected Past: Challenges to Network Studies in Archaeology and History,* edited by Tom Brughmans, Anna Collar, and Fiona Coward, 57–84. Oxford: Oxford University Press, 2016. https://doi.org/10.1093/9780198748519.003.0010.

Pitman, Jim. *Probability.* New York: Springer, 1993.

Roelli, Philipp. "Definition of Stemma and Archetype." In *Handbook of Stemmatology: History, Methodology, Digital Approaches,* edited by Philipp Roelli, 209–25. Berlin: De Gruyter, 2020.

Rossi, Guido, ed. *Atti del Convegno internazionale di studi Accursiani.* Milano: Giuffrè, 1968.

Ryan, Martin J. "Isidore amongst the Islands: The Reception and Use of Isidore of Seville in Britain and Ireland in the Early Middle Ages." In *A Companion to Isidore of Seville,* edited by Jamie Wood and Andrew Fear, 424–56. Brill's Companions to the Christian Tradition 87. Leiden: Brill, 2020.

Schiegg, Markus. *Frühmittelalterliche Glossen: Ein Beitrag zur Funktionalität und Kontextualität mittelalterlicher Schriftlichkeit.* Germanistiche Bibliothek 52. Heidelberg: Winter, 2015.

Schindel, Ulrich. *Die lateinischen Figurenlehren des 5. bis 7. Jahrhunderts und Donats Vergilkommentar.* Abhandlungen der Akademie der Wissenschaften in Göttingen. Philologisch-Historische Klasse 91. Göttingen: Vandenhoeck & Ruprecht, 1975.

— "Zur Datierung des Basler Figurentraktats (cod. lat. F III 15d)." *Göttinger Forum für Altertumswissenschaft* 2 (1999): 161–78.

Singer, Karen B. "Random Intersection Graphs." Ph.D., The Johns Hopkins University, 1995. https://www.proquest.com/docview/304306614/abstract/B7DC2EB06B3D4803PQ/1.

Stagni, Ernesto. "Nell'officina di Paolo Diacono? Prime indagini su Isidoro e Cassiodoro nel Par. lat. 7530." *Litterae Caelestes* 4 (2012): 9–105.

Steinová, Evina. "Annotation of the Etymologiae of Isidore of Seville in Its Early Medieval Context." *Archivum Latinitatis Medii Aevi* 78 (2020): 5–81.

Steinová, Evina, and Peter Boot. "Editing Glosses as Networks: Exploring the Explorative Edition." *A Companion to Digital Editing Methods* 1 (2022): forthcoming.

— "The Glosses to the First Book of the Etymologiae of Isidore of Seville: A Digital Scholarly Edition." Amsterdam: Huygens Institute, 2021. https://db.innovatingknowledge.nl/edition/#left-home.

Teeuwen, Mariken. "Marginal Scholarship: Rethinking the Function of Latin Glosses in Early Medieval Manuscripts." In *Rethinking and Recontextualizing Closses. New Perspectives in the Study of Late Anglo-Saxon Glossography,* edited by Patrizia Lendinara, Loredana Lazzari, and Claudia Di Sciacca, 19–37. Textes et Études Du Moyen Age 54. Turnhout: Brepols, 2011.

— "The Impossible Task of Editing a Ninth-Century Commentary: The Case of Martianus Capella." *Variants: The Journal of the European Society for Textual Scholarship* 6 (2007): 191–208.
— "Writing in the Blank Space of Manuscripts: Evidence from the Ninth Century." In *Ars Edendi Lecture Series,* vol. IV, edited by Barbara Crostini, Gunilla Iversen, and Brian M. Jensen, 1–25. Stockholm: Stockholm University Press, 2016.
Trovato, Paolo. "Neo-Lachmannism: A New Synthesis?" In *Handbook of Stemmatology: History, Methodology, Digital Approaches,* edited by Philipp Roelli, 109–38. Berlin: De Gruyter, 2020.
Tura, Adolfo. "Essai sur les *marginalia* en tant que pratique et documents." In *Scientia in margine: études sur les Marginalia dans les manuscrits scientifiques du Moyen Âge à la Renaissance,* edited by Danielle Jacquart and Charles S. F. Burnett, 261–387. Sciences historiques et philologiques. Hautes études médiévales et modernes 88. Genève: Droz, 2005.
de Valeriola, Sébastien. "Can Historians Trust Centrality? Historical Network Analysis and Centrality Metrics Robustness." *Journal of Historical Network Research* 6, no. 1 (2021): 85–125. https://doi.org/10.25517/jhnr.v6i1.105.
Valleriani, Matteo, Florian Kräutli, Maryam Zamani, Alejandro Tejedor, Christoph Sander, Malte Vogl, Sabine Bertram, Gesa Funke, and Holger Kantz. "The Emergence of Epistemic Communities in the Sphaera Corpus: Mechanisms of Knowledge Evolution." *Journal of Historical Network Research* 3 (2019): 50–91. https://doi.org/10.25517/jhnr.v3i1.63.
Wieland, Gernot R. "The Glossed Manuscript: Classbook or Library Book?" *Anglo-Saxon England* 14 (1985): 153–73.
— *The Latin Glosses on Arator and Prudentius in Cambridge University Library, MS Gg. 5.35.* Studies and Texts 61. Toronto: Pontifical Institute of Mediaeval Studies, 1983.

Appendix I: Overview of the manuscripts containing glosses to the first book of the *Etymologiae*

Full Manuscript shelfmark	Shortened label	Date of glossing	Place and region of glossing	All glosses	Parallel glosses	Rank 1	Rank 2	Rank 3
Orléans, Bibliothèque municipale, MS 296 (pp. 1–32)	Orleans296	9th c., 1/2	Paris or Fleury (northern France)	768	294	29	201	64
Leiden, Universiteitsbibliotheek, Voss. Lat. O 41	VLO41	10th c.	Fleury (northern France)	682	190	25	136	29

Full Manuscript shelfmark	Shortened label	Date of glossing	Place and region of glossing	All glosses	Parallel glosses	Rank 1	Rank 2	Rank 3
London, British Library, Harley 3941	Harley3941	9th/10th c. and 10th c.	unknown (Brittany)	535	309	20	169	120
Paris, Bibliothèque nationale de France, Lat. 7670	Paris7670	9th c.	Paris (northern France)	353	126	19	79	28
Reims, Bibliothèque municipale, MS 426 (fols. 1–117)	Reims426	9th c.	Reims (northern France)	345	127	14	93	20
Paris, Bibliothèque nationale de France, Lat. 7490	Paris7490	9th c.	Paris or Fleury (northern France)	241	65	12	36	17
Paris, Bibliothèque nationale de France, Lat. 7585	Paris7585	10th c., 2/2	Canterbury (England)	225	129	7	46	76
Paris, Bibliothèque nationale de France, Lat. 7671	Paris7671	9th c.	unknown (northern France)	135	37	3	23	11
Paris, Bibliothèque nationale de France, Lat. 7559	Paris7559	9th c.	Paris (northern France)	116	42	11	22	9
Trier, Bibliothek des Bischöflichen Priesterseminars, MS 100 (fols. 1r–16)	Trier100	9th c.	unknown (France)	74	14	1	8	5
Leiden, Universiteitsbibliotheek, Voss. Lat. F 82	VLF82	9th c.	Paris (northern France)	71	27	1	18	8
Oxford, Bodleian Library, Junius 25 (fols. 134–151)	Junius25	9th c.	Murbach (German area)	60	17	2	15	0
Bologna, Biblioteca Universitaria, MS 797	Bologna797	9th c.	Area of Reims (northern France)	55	24	4	15	5

Full Manuscript shelfmark	Shortened label	Date of glossing	Place and region of glossing	All glosses	Parallel glosses	Rank 1	Rank 2	Rank 3
Paris, Bibliothèque nationale de France, Lat. 11278	Paris11278	9th c., 1/2	unknown (southern France?/ northern Italy?)	48	28	3	10	15
Institut de recherche et d'histoire des textes, BVMM, Collections privées, digitisation of CP 342	IRHT342	12th c.	unknown (France?)	47	35	3	10	22
Montecassino, Archivio dell'Abbazia, MS 320 (pp. 5–398)	Montecassino320	beginning of the 10th c.	unknown (Italy)	43	8	2	6	0
Gotha, Forschungsbibliothek, Membr. I 147	GothaI147	9th c., 2/4	unknown (Brittany)	42	34	0	1	33
Madrid, Real Academia de la Historia, MS 76	RAH76	c. 946	San Millán de la Cogolla? (northern Spain)	38	28	1	25	2
Oxford, Queen's College, MS 320	Queen320	end of the 11th c./ beginning of the 12th c.	Canterbury (England)	38	28	0	6	22
Montpellier, Bibliothèque interuniversitaire, H 53 (fols. 5–265)	MontpellierH53	10th/ 11th c.	unknown (eastern France)	33	14	4	5	5
Munich, Bayerische Staatsbibliothek, Clm 6411	Clm6411	9th c., 1/4	Passau? (German area)	30	8	2	6	0
Chartres, Bibliothèque municipale, MS 16	Chartres16	11th c.	unknown (France)	29	8	3	4	1
Madrid, Real Academia de la Historia, MS 25	RAH25	c. 954	San Pedro de Cardeña (northern Spain)	29	28	1	25	2

Full Manuscript shelfmark	Shortened label	Date of glossing	Place and region of glossing	All glosses	Parallel glosses	Rank 1	Rank 2	Rank 3
Vatican, Biblioteca Apostolica Vaticana, Barb. Lat. 477 (fols. 4–123)	BarbLat 447_4	Beginning of the 11th c.	unknown (France)	29	3	2	1	0
Vatican, Biblioteca Apostolica Vaticana, Pal. Lat. 1746	PalLat1746	9th c.	Lorsch (German area)	27	6	4	2	0
Paris, Bibliothèque nationale de France, Lat. 7583	Paris7583	9th c., 2/2	unknown (northern France)	25	6	0	3	3
Cesena, Biblioteca Malatestiana, S.XXI.5	Cesena SXXI5	9th c. and 10th/11th c.	unknown (northern Italy)	20	15	1	3	11
Munich, Bayerische Staatsbibliothek, Clm 6250	Clm6250	9th c. and 10th/11th c.	Freising (German area)	15	9	1	3	5
Vatican, Biblioteca Apostolica Vaticana, Vat. Lat. 5763 (fols. 3–80)	VatLat5763	9th c.	Bobbio? (northern Italy)	15	5	1	3	1
Bern, Burgerbibliothek, MS 101	Bern101	9th c., 1–2/3	Loire area (France)	13	9	3	2	4
Paris, Bibliothèque nationale de France, n.a.l. 2633 (fols. 18–19)	ParisNAL2633	9th c., 4/4	unknown (France)	12	2	0	1	1
Venice, Biblioteca Marciana, II 46	VeniceII46	11th/12th c.	unknown (northern Italy)	11	11	0	0	11
London, British Library, Cotton Caligula A xv (fols. 3–38, 42–64, 73–117)	CotCalAxv	12th c.	Canterbury (England)	10	7	0	1	6
Munich, Bayerische Staatsbibliothek, Clm 4541	Clm4541	9th c., 3/3 and 11th c., 2/2	Benediktbeuern (German area)	10	8	1	3	4

Full Manuscript shelfmark	Shortened label	Date of glossing	Place and region of glossing	All glosses	Parallel glosses	Rank 1	Rank 2	Rank 3
Schaffhausen, Stadtbibliothek, Min. 42	Schaffhausen42	9th c., 2/2	Mainz and St. Gallen (German area)	8	6	0	1	5
Oxford, Bodleian Library, Auct. T.2.20 (fols. 2–124)	AuctT2.20	9th c., 3/4	Auxerre (northern France)	7	2	0	2	0
Laon, Bibliothèque Suzanne Martinet, MS 447	Laon447	9th c., 2/4	Mainz (German area)	7	5	1	2	2
London, British Library, Arundel 129	Arundel129	unknown	unknown	6	1	0	0	1
Wolfenbüttel, Herzog August Bibliothek, Weiss. 64	Wolfenbuttel64	9th c.	unknown (France?)	6	4	0	1	3
Berlin, Staatsbibliothek, Ham. 689	Ham689	11th c.	unknown (northern Italy)	4	2	0	0	2
Leiden, Universiteitsbibliotheek, BPL 122	BPL122	9th c., 4/4	Lyon (southern France)	3	0	0	0	0
Paris, Bibliothèque nationale de France, Lat. 10293	Paris10293	9th c., 3/4	Reims (northern France)	3	2	1	0	1
Milan, Biblioteca Ambrosiana, L 99 sup.	MilanL99sup	8th c., 2/2	Bobbio (northern Italy)	3	2	0	1	1
London, British Library, Harley 3099	Harley3099	12th c., 2/3	Munsterbilsen (German area)	2	1	0	1	0
Leiden, Universiteitsbibliotheek, Voss. Lat. O 15	VLO15	11th c., 1/2	Limoges (Southern France)	2	0	0	0	0
Paris, Bibliothèque nationale de France, Lat. 7588	Paris7588	unknown	unknown	2	2	0	2	0
Vatican, Biblioteca Apostolica Vaticana, Reg. Lat. 1953	RegLat 1953	9th c., 1/4	Orléans (northern France)	2	0	0	0	0

Full Manuscript shelfmark	Shortened label	Date of glossing	Place and region of glossing	All glosses	Parallel glosses	Rank 1	Rank 2	Rank 3
Vatican, Biblioteca Apostolica Vaticana, Barb. Lat. 477 (fol. 3)	Barb-Lat447_3	beginning of the 11th c.	unknown (France)	1	0	0	0	0
Bern, Burgerbibliothek, MS 611 (fols. 42–93)	Bern611	8th c., 1/2	Bourges (southern France)	1	0	0	0	0
Brussels, Koninklijke Bibliotheek, II 4856	Brussels-II4856	end of the 8th c.	Corbie (northern France)	1	1	0	0	1
Cologne, Dombibliothek, MS 123 (fols. 76–80)	Cologne123	9th c., 4/4	unknown (eastern France)	1	1	0	1	0
London, British Library, Harley 2713 (fols. 1–34)	Harley-2713	9th c., 4/4	unknown (northern France)	1	0	0	0	0
London, British Library, Harley 5977 (fol. 71)	Harley-5977_71	unknown	unknown	1	0	0	0	0
Reims, Bibliothèque municipale, MS 425	Reims425	mid-9th c.	Reims (northern France)	1	1	0	0	1
Total:	–	–	–	4,286	1,732	182	993	557

For Appendix II (Edge) and Appendix III (Node) please see https://zenodo.org/record/8146577.

The Journal of HISTORICAL NETWORK RESEARCH

KATHARINA KASKA

Scribal networks

Visualizing twelfth-century Cistercian book production through network analysis

Journal of Historical Network Research 9 (2023) 101–129

Keywords Paleography, Cistercians, library history, Magnum Legendarium Austriacum, Medieval Studies, Manuscript Studies

Abstract This paper tests the application of network analysis to the visualization and analysis of paleographical data. In recent years, the twelfth-century scriptoria of the Austrian Cistercian monasteries of Heiligenkreuz, Zwettl and Baumgartenberg have been thoroughly investigated. A vast amount of data on scribes and their contributions to various manuscripts has been published in analog publications, as well as online on the website www.scriptoria.at, run by Alois Haidinger. The presentation of the data, mainly in the form of lists, makes it difficult for researchers to appreciate the possibilities that this groundbreaking work provides. For this paper, the data is instead presented as networks of codicological units and networks of scribes within and between the monasteries. These networks highlight the development and interconnectedness of twelfth-century Cistercian book production, point out potential research questions (e.g., for the *Magnum Legendarium Austriacum*), and aid in disseminating the results to a wider audience.

C²DH LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

UNIVERSITY OF LUXEMBOURG

Introduction

Recent research has provided a vast amount of data on the scribes who copied books in Austrian monasteries in the twelfth century. By evaluating their collaboration and tracking single hands across many manuscripts, it has become possible to reconstruct the scriptoria of three Cistercian monasteries (Heiligenkreuz, Zwettl and Baumgartenberg). In this paper, the available data is used to showcase how network analysis can help to present these reconstructions in a reader-friendly form and point out further research questions.

In the first part, the three monasteries chosen for this research are introduced, along with their libraries. In part two, the data sources and the steps towards network analysis with Gephi are discussed. Part three addresses questions of data quality and completeness. Part four presents network diagrams for the Heiligenkreuz scriptorium, as well as for all three scriptoria combined. These include networks of scribes and of codicological units. Finally, part five discusses the application of network analysis for investigating manuscript and scribal transfer, using the example of *Magnum Legendarium Austriacum* and *De sacramentis* by Hugh of St. Victor.

1. Historical background

The arrival of monks from Morimond in the Viennese Woods, along with the foundation of Heiligenkreuz in 1133, led to a quick expansion of the Cistercian order in northern Austria. In 1138 the first daughterhouse, Zwettl, was founded, followed by Baumgartenberg in Upper Austria in 1141/2.[1] The large number of still extant manuscripts shows that all three monasteries started to establish new libraries shortly after their foundation.[2]

For Heiligenkreuz, a book list compiled during the abbacy of its first abbot Gotschalk (1133–1147) provides detailed information on about 70 manuscripts

Acknowledgements: I like to thank the libraries and librarians in Heiligenkreuz (P. Roman Nägele), Zwettl (Andreas Gamerith) and the State Library of Upper Austria. Without their continuous support, my work would not be possible.
Corresponding author: Katharina Kaska, katharina.kaska@onb.ac.at

1 There is no modern overview on the history of the Cistercian order in Austria. Short entries for each monastery can be found in Zák, *Österreichisches Klosterbuch,* 109–39. Some aspects of the early history of Heiligenkreuz and Zwettl are discussed in Lutter, "Zisterzienser."

2 Manuscripts from Heiligenkreuz are today kept in the monastery's own library as well as the Austrian National Library, manuscripts from Zwettl are kept in Zwettl, while manuscripts from Baumgartenberg can be found in the State Library of Upper Austria as well as the Austrian National Library.

that formed the monastery's early library.[3] The list does not mention liturgical books, which clearly existed but were probably kept in another room and only survive as fragments.[4] In the late fourteenth century, another detailed book list was compiled that yet again makes it possible to identify extant books as coming from Heiligenkreuz.[5] Overall, about two-thirds of the manuscripts mentioned in the book lists survive.[6]

The high medieval book lists for Zwettl are less detailed and therefore more difficult to interpret. A list from the late twelfth century mentions works by Augustine with no clear indication that those works were actually part of the library. It seems very likely, however, since the manuscript that contains the list can be identified as its last entry.[7] This entry also shows another shortcoming of the list: it does not indicate which works were combined into one volume. Another list from the first half of the thirteenth century poses a similar challenge. It clearly states that the books were part of the library, but only mentions one text per volume.[8] If each entry stands for one manuscript, then the list contains over 100 volumes. As in Heiligenkreuz, liturgical books are not included. Finally, another contemporary list contains all the works by Augustine held in Zwettl.[9] There are discrepancies in the number of works by Augustine between all lists, which sheds some doubt on their reliability.[10] The identification of extant manuscripts with entries in the book list is referenced in the modern manuscript catalogue, but has not been reevaluated for this study.[11]

The library of Baumgartenberg was much smaller than those of its motherhouse and sisterhouses. An early thirteenth-century book list provides detailed descriptions of nearly 70 non-liturgical manuscripts (if multi-volume manu-

3 Edited in Gottlieb, *MBKÖ 1,* 18–21. A detailed discussion of discrepancies and issues with identification, as well as a table of identified manuscripts, can be found in Haidinger and Lackner, *Bibliothek,* 10–18.

4 See for instance Cod. 176, fol. I, or a group of fragments of a breviary called Fragmentgruppe Cod. 7, at: https://www.scriptoria.at/cgi-bin/scribes.php?ms=AT3500-FragmC7 (accessed May 25, 2022).

5 Gottlieb, *MBKÖ 1,* 34–74. At about the same time, an inventory of all bookcases was also drawn up (edited Ibid., 22–33).

6 The exact number is difficult to determine since some of the entries cannot firmly be identified with extant manuscripts.

7 Gottlieb, *MBKÖ 1,* 510–11.

8 Gottlieb, *MBKÖ 1,* 511–14.

9 Gottlieb, *MBKÖ 1,* 514–16.

10 Discussed by Gottlieb.

11 Ziegler, *Zisterzienserstift Zwettl.* No new attempts to identify the entries in the book list have been carried out since Ziegler. An upcoming publication on the Zwettl scriptorium by Alois Haidinger will most likely shed more light on this question (see footnote 59).

scripts are counted separately).[12] In addition, several liturgical manuscripts are listed that are lost today.[13] Contrary to Heiligenkreuz and Zwettl, Baumgartenberg is no longer active; it was closed in 1784 as part of the dissolution of monasteries during the reign of emperor Josef II. Its library was dispersed and partly sold off in the late 1780s. Most, if not all, parchment manuscripts that were kept in it at the time of the dissolution are today part of the collections of the State Library of Upper Austria and the Austrian National Library. However, less than 50% of the manuscripts mentioned in the book list are still extant, which points to earlier losses.[14]

Recent paleographical studies, discussed in more detail below, have shown that a large number of extant manuscripts from Heiligenkreuz and Zwettl, and to a lesser extent from Baumgartenberg, that are mentioned in the book lists were produced by the respective scriptoria.[15] These investigations also point towards exchange processes between the scriptoria, such as manuscript transfer and, more importantly, the transfer of scribes. Overall, these manuscripts therefore provide a perfect case study for possible interactions between motherhouses, daughterhouses and sisterhouses as far as book production is concerned, as well as for the development of scriptoria in twelfth-century Cistercian monasteries in general.[16]

12 Some entries are slightly later additions. The list is edited in Paulhart, *MBKÖ 5,* 14–18 (with manuscript identifications).

13 A few fragments survived in manuscript bindings, see e.g., Vienna, Austrian National Library, Cod. 671.

14 A very short introduction to the history of Baumgartenberg's library can be found in Paulhart, *MBKÖ 5,* 13–14; Kaska, "Schreiber und Werke," 63–64. Early modern shelfmarks indicate that there were more paper manuscripts extant at the time of the dissolution than survive today. This topic will be discussed in more detail in future publications on the history of the library of Baumgartenberg.

15 For more detailed information on the number of surviving books from each monastery see section 3.

16 For more details on possible exchange processes and how to determine them see Kaska, "Schreiber und Werk"; and Haidinger and Lackner, *Bibliothek.* While only scribal networks are discussed in this paper, the investigation is part of a larger project on the interaction between paleographical and philological networks in twelfth-century book production with a focus on the library of Baumgartenberg. For a short description, see: http://www.iter-austriacum.at/kodikologie/texttransfer-und-buchaustausch-netzwerke-monastischer-handschriftenproduktion-am-beispiel-des-zisterzienserstifts-baumgartenberg-in-oberoesterreich/ (accessed May 25, 2022).

2. Available data and processing

For my study, I use paleographical data from two different sources: most information comes from an online publication by Alois Haidinger, while additional material is provided by my own research on the library and scriptorium of Baumgartenberg.[17]

In 2010, Alois Haidinger and Franz Lackner started to catalogue the medieval manuscripts in Heiligenkreuz. In addition to detailed descriptions of manuscript content and codicological features,[18] considerable effort was put into identifying all the scribes, rubricators and correctors. The results of this detailed paleographical analysis have been continuously published online at www.scriptoria.at since 2013 (see figure 1). A short analysis of the early scriptorium was also published in printed form in 2015.[19]

In recent years, the scope of Haidinger's research has expanded to other monasteries whose manuscripts can be connected to the Heiligenkreuz scriptorium. Intensive work has gone into analyzing the early scriptorium of Zwettl. To a lesser extent, data on twelfth-century manuscripts from the Cistercian monastery of Rein, and on a few manuscripts from Baumgartenberg that are today in the Austrian National Library,[20] is also available online. In a future update, scriptoria.at will also include an analysis of the illuminations in all manuscripts.[21]

The basic unit of reference for the paleographical analysis is not manuscripts but codicological units. Manuscripts can consist of one or more codicological units, potentially copied at different times periods or in different production contexts. Multiple units were sometimes combined into one manuscript shortly after

17 On how to identify scriptoria in the high Middle Ages see Garand, "Manuscrits monastiques"; Egger, "Suche," 377–88. On the practical aspects of scriptorium research see e.g., the discussions in Cohen-Mushlin, *A Medieval Scriptorium,* 53–55; Golob, *Cistercian Manuscripts,* 64–68.

18 The earliest manuscripts are published in Haidinger and Lackner, *Bibliothek.* Further manuscript descriptions can be found at: https://www.scriptoria.at/cgi-bin/sc_desc.php (accessed May 25, 2022).

19 Haidinger and Lackner, *Bibliothek,* 21–35.

20 The data is not always complete and was reevaluated and expanded by my own studies.

21 See the announcement on the website scriptoria.at: "Eine wesentlich erweiterte Version von scriptoria.at wird in der ersten Jahreshälfte 2022 online gestellt werden. Zu jeder Schreiberhand sollen nicht nur wie bisher Schriftspecimina in Form von Abbildungen, sondern zusätzlich sukzessive Zusammenstellungen ihrer Schriftcharakteristika in Form von Abbildungen von Kürzungen, Ligaturen, Einzelbuchstaben und Wörtern beziehungsweise Wortteilenwerden geboten werden. Darüber hinaus wird zu jeder in scriptoria.at genannten Handschrift auch deren Buchschmuck (einschließlich der niederrangigen Elemente wie Majuskelinitialen und Lombarden) analysiert werden." (accessed May 25, 2022).

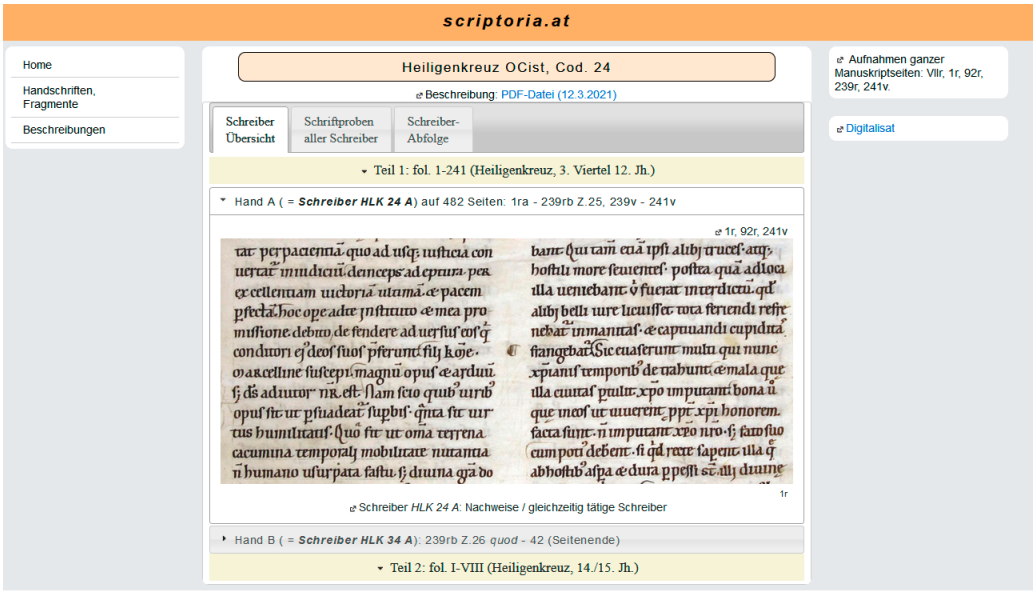

Fig. 1 Screenshot of paleographical data for Heiligenkreuz Cod. 24 as presented in scriptoria.at

their production, but may also have been combined only centuries later when books were rebound. If the difference between their times of production is immediately obvious, e.g., in the case of a high medieval codicological unit bound together with late medieval codicological units, this is acknowledged even in the most basic manuscript catalogues. Many of the manuscripts investigated in this study, however, seem quite uniform at first glance. All pages were copied at approximately the same time, and sometimes even by the same scribes. Even in these manuscripts it is possible to differentiate between codicological units by comparing the quire marks and the ruling, or by investigating the quire structure. Some of these manuscripts have not changed their composition since their time of production, as can be shown by comparisons with medieval book lists. They might indeed have been planned that way and only copied in several codicological units for practical reasons. Others had parts added to them at a slightly later period. Since in many cases the date of compilation cannot be decided with certainty, the codicological unit remains the best reference unit for research.[22]

22 For an extensive discussion of the issue of codicological units on a more general level see Andrist et al., *La syntaxe*. For the monasteries discussed in this paper, the rearrangement of codicological units is obvious if medieval book lists and modern manuscript catalogues are compared. For Heiligenkreuz see the discussion of this topic throughout Kaska, "Untersuchungen."

In scriptoria.at, scribal hands are identified by letters and can be referenced using the shelfmark, for example "Heiligenkreuz Cod. 19 Hand A".[23] If a scribe can be identified in more than one codicological unit, it is named after an important manuscript in his oeuvre in a similar fashion, e.g., scribe "HLK 19 A". In a few cases the actual name of the scribe is known, such as Udalricus or Heinricus. These two types of identification are then used whenever a scribe appears in a codicological unit.

For each scribal hand, detailed information on its contribution to a manuscript is available on the website, which makes it possible to display the sequence of hands in the codicological unit (tab "Schreiber-Abfolge" on scriptoria.at). Similarly, all correctors and rubricators are identified and labeled, e.g., "Heiligenkreuz Cod. 19 Korrektor/Rubrikator A", with their contributions listed.

In my own project, I collected paleographical data on all manuscripts from Baumgartenberg mentioned in the thirteenth-century book list and labeled it according to Haidinger's convention. The data is stored locally in tabular form.

Scriptoria.at already provides a basic analysis of the paleographical data. For each scribe, all codicological units to which he contributed, as well as his role in these manuscripts – text scribe, corrector, rubricator – are listed and given a first impression of his importance within the scriptorium. In a separate list, all hands that collaborated with the scribe in question within the same codicological unit are collected as a first step to constructing a network of scribes.[24] Both lists are useful for researchers with a more in-depth knowledge of the scriptorium, but can be overwhelming for non-experts. It therefore seemed reasonable to use network display as a tool for better visualization and further analysis.

To investigate the relation between the scriptoria of motherhouses, daughterhouses and sisterhouses, I retrieved all published data on twelfth-century manuscripts[25] from Heiligenkreuz, Zwettl and Baumgartenberg from scriptoria.at by manually transferring it into tabular form (Excel sheet). The cut-off point of 1200 was chosen, as for Zwettl no sufficient data for the thirteenth-century scriptorium is available, and my own research on Baumgartenberg mainly focuses on the earliest manuscripts. The network therefore excludes data on the later development of the Heiligenkreuz scriptorium that is already available on scriptoria.at, as well as data on some early thirteenth-century Baumgartenberg manuscripts. Manuscript fragments were also not included, even if they can be dated to the twelfth

23 For a description of the naming system and the database in general see Haidinger and Lackner, *Bibliothek,* 21–22.

24 For the example of HLK 19 A, see the entry at: https://www.scriptoria.at/cgi-bin/rel_scribes.php?scribe_name=HLK%2019%20A (accessed May 25, 2022).

25 This includes manuscripts dated "um 1200" in the database.

Shelf-mark	Provenance	Date	Role	Name	Identifi-cation	Pages	Details
OÖLB Cod. 328	Baumgartenberg	1142–1175	text scribe	OOeLB 328 A	HLK 19 A	100	26ra–76va Z 8 (est); ...

Tab. 1 Basic data structure.

century. Some of them were written by known scribes, but since in general they lack information on scribal collaboration, they could distort the network.

The scriptoria.at data was then structured by using the "text to column" command in Excel. For the table, I followed the conventions on scriptoria.at and used the categories shelfmark, provenance, date, role of scribe, name of scribe, identification of scribe, pages and details (see Table 1). Nearly 1200 rows of data could thus be retrieved from Alois Haidinger's online publication. About 130 further rows were contributed by my own research, mainly on manuscripts in the State Library of Upper Austria.

Not all columns are relevant for the network analysis performed in this study; they were included in the original table to maintain a complete set of data from scriptoria.at for reference purposes and further studies.

The category "role" has the values text scribe, rubricator and corrector, and therefore indicates the type of contribution within the manuscript. Only text scribes were included in the analysis for this paper. Corrections to a manuscript can happen after the initial production and even after manuscripts were transferred from their place of origin; in fact, rubrics are not always entered immediately. Text scribes, on the other hand, are responsible for the very first step in copying a manuscript, i.e., copying the main part of the text from an exemplar.

In most codicological units, the "date" and "provenance" categories are solely based on recent paleographical observations.[26] This therefore provides a point of comparison for the results obtained by network analysis, but should not be used to construct a network. The category "pages" indicates the importance of the contribution of each scribe to a codicological unit.[27] and can be useful to differentiate

26 A Boolean category could be added for manuscripts/codicological units mentioned in the various booklists. This can narrow the date for manuscripts from Heiligenkreuz that are mentioned in the earliest book list and are therefore datable before the middle of the twelfth century.

27 "Details" adds a more detailed description of the parts of the manuscript each scribe wrote.

between different types of text scribes. Short contributions in many codicological units can, for instance, point towards "teachers" who write a few lines as exemplars for their pupils.[28] These categorizations require additional in-depth investigations of scribal features, which are not within the scope of this paper. For the basic network presented here, the page count was therefore not used in the analysis.

The main categories are thus "shelfmark" and "identification", which serve as the source and target for the edge table used for all network analyses. The shelfmark is the ID for a manuscript or codicological unit, while "identification" denotes the ID of a scribe who contributed to a particular codicological unit. If the scribe is known, i.e., he contributed to more than one codicological unit, the entry in the identification column follows the conventions detailed above. If a scribe cannot be identified in another codicological unit and the identification field would be empty, I used the "name" tag[29] with an additional hyphen for the identification column (in the example in Table 1 this would be OOeLB 328-A). This convention allows one to also include singular scribes in the network analysis, and helps to quickly differentiate known and unknown scribes.

For further processing, a reduced table was produced with information on all text scribes (target) in all codicological units (source) from Heiligenkreuz, Zwettl and Baumgartenberg before 1200. Additionally, a similar table that only features information on codicological units from the library of Heiligenkreuz was compiled in order to test the validity of the network projection as described in the following section.

Since the dataset lists contributions to codicological units by text scribes, the edges in this representation connect two different types of nodes: manuscript/codicological unit-nodes (category "shelfmark"), and scribe-nodes (category "identification"). It is therefore a two-mode network. The first, larger table contains 1103 edges and 935 nodes. Of these nodes, 365 are manuscripts, and 570 are scribes. The data for Heiligenkreuz alone contains 564 edges and 448 nodes (166 codicological unit nodes and 282 scribal nodes).

In a next step, both tables were imported into Gephi. Before analysis, the two-mode network needs to be transformed into two separate one-mode networks: a network with scribe-nodes that shows the connections (i.e., collaborations) between various scribes, and a network with codicological unit-nodes that brings

28 For a discussion of these "praescriptiones" in Carolingian scriptoria, see Tibbetts, "Praescriptiones."

29 The category "name" names the scribes as A, B, C, etc. as they appear in the manuscript without attempting to identify them (as discussed above).

together codicological units that share scribes. This was achieved by using the Gephi Plug-in MultiMode Networks Projections.

Different evaluation techniques were then applied to the network to obtain the following network properties:

Degree centrality

The parameter degree expresses the number of nodes to which a node is connected in a network. In this study, where nodes are scribes or codicological units, the degree expresses how well-connected certain scribes or codicological units are. In a network of codicological units, a node with a high degree indicates a codicological unit that shares scribes with many other codicological units. In the network of scribes, a node with a high degree indicates a scribe that worked together with many other scribes and is therefore highly connected within the scriptorium. However, the degree of a node is also increased if a scribe contributed to a codicological unit that many other scribes also contributed to. The degree value can therefore be high even if the scribe only contributed to a few codicological units, if these are multi-scribe codicological units. Since the number of pages each scribe wrote in each codicological unit is not included in this basic network model, the degree does not give an indication of the scribal output and therefore the scribe's importance for book production overall. The degree value is therefore a very basic tool of analysis, but has shown interesting results for the scriptorium of Heiligenkreuz. Nodes with a higher degree are larger in figure 3, which is the only graph where the degree is used for evaluation.

Betweenness Centrality

As a measure of centrality, "betweenness centrality" was chosen. This measures how often a node lies on the shortest path of connection between two points in the network. It therefore gives an indication of how important a node is to connect different parts of the network. In the scribal network, the betweenness centrality is especially high for scribes that work in more than one scriptorium. Each scriptorium forms its own cluster in the network that is connected by these scribes. Nodes with higher betweenness centrality are larger in the graph.

Clusters

A cluster is a group of nodes that is more densely connected to each other than to other nodes. In this study, these are either groups of codicological units that share the same scribes or groups of scribes who collaborate in several codicological units. To divide the network into clusters, the modularity of the graph is calculated in Gephi. Nodes belonging to the same clusters are given the same color. A major issue with modularity is its resolution limit. Some of the clusters identified by modularity optimization, such as those used by Gephi, might actually be

combinations of smaller clusters.[30] Small clusters in large networks thus remain hidden. To obtain a complete partition of the network, the major clusters have to be reexamined to determine if they themselves contain clusters. However, for the investigations in this paper that focus on large clusters related to separate scriptoria, this issue seems to be negligible (see the results in section 4).

3. Data quality

As with all medieval sources, several issues with data quality need to be addressed before attempting an analysis. The main issues are the incompleteness of the data and the methodological challenges when using results from paleographical research.

Clearly, the data is not complete, since not all manuscripts have survived to the present day. As indicated in the introduction, for Heiligenkreuz the earliest book list from the middle of the twelfth century indicates a loss of about one third of non-liturgical books. The survival rate of liturgical books from this period is close to zero – only a few fragments are still extant. Overall, the losses of the very early products of the scriptorium are therefore clearly higher, but cannot be quantified due to a lack of information. The loss rate for the latter part of the twelfth century can only be extrapolated by using the fourteenth-century book list as a general guideline.[31] Again, about two thirds of non-liturgical books mentioned in the list have survived, which would point to a rather constant rate of loss. However, some uncertainties remain. The loss rate can depend on the text transmitted in the manuscripts, as was shown for liturgical manuscripts. Different periods of production in the scriptorium might focus on different types of text, which in turn influence the rate of loss. During the very early stages of the library, Heiligenkreuz mainly collected works by the Church Fathers such as Augustine, Jerome and Gregory the Great, but barely included modern theologians in its library.[32] This is even true for Bernard of Clairvaux (*ca.* 1090–1153), one of the most important Cistercian authors of the time. By the middle of the twelfth century, only his *Apologia* was available in Heiligenkreuz, although in the following decades several manuscripts of his works were copied. Similarly, only two texts by Hugh of St. Victor (*ca.* 1097–1141), another highly popular contemporary theologian, are mentioned in the book list,[33] but more works were

30 See the discussion in Fortunato and Barthélemy, "Resolution limit."

31 There is no indication that more than one library existed in Heiligenkreuz in the fourteenth century, i.e., almost all extant manuscripts can be identified in the book lists. It therefore gives a good indication of the books possessed by the monastery at this time. As stated previously, liturgical manuscripts are the exception. For the distribution of books throughout the monastery see e.g., Gottlieb, *Über mittelalterliche Bibliotheken,* 303–9.

32 For a similar development in Aldersbach (OCist) see Frioli, "Antichi manoscritti," 212–13.

33 Haidinger and Lackner, *Bibliothek,* 16.

copied shortly after its compilation. These works were either not deemed immediately necessary for monastic life or, being rather recent works, might only have become available for copying in the second half of the twelfth century. Overall, one can assume a slightly different acquisition profile for this later time period. Does this then indicate a difference in losses between the time period covered by the first book list and slightly later copies? Cross checking with the extensive fourteenth-century book list, the only other source of information for the medieval library, does not show a clear pattern. Most works by Augustine or Gregory the Great are still extant, while many of Jerome's work are missing. Of 11 manuscripts with works by Bernard of Clairvaux, three are missing today and similar numbers can be established for Hugh of St. Victor. One category where the losses seem to be higher than average are commentaries on books of the Bible. This also explains the higher rate of loss for Jerome, whose commentaries were a prominent part of the early library. These books might have been replaced by more modern works as the genre evolved. They are an example of different loss patterns for different types of text, which still need to be investigated in more detail.

Another reason for the loss of very early copies might have been manuscript quality. As the scriptorium becomes more experienced and better or more complete exemplars became available, older copies may have been discarded. Two such cases will be discussed in section 5.

While these factors clearly influence the completeness of the data and can distort the analysis, it is not easy to quantify their influence. Of even greater concern is the difference in survival rate between the monasteries. The rate of loss at Baumgartenberg, about 50%, is much higher than that of Heiligenkreuz, which is in turn possibly higher than that of Zwettl. According to scriptoria.at, about 160 codicological units from prior to 1200 survive from Heiligenkreuz, and about 170 from Zwettl, although not all of these can be attributed to the respective scriptoria. For Baumgartenberg, only 27 codicological units can be attributed to this time period. Due to a lower rate of survival, as well as a smaller library to begin with, Baumgartenberg therefore differs considerably from the other two monasteries in the absolute number of codicological units available for investigation.

Paleographical data furthermore differs from other forms of data due to its subjective nature. Discerning different scribes in a codicological unit, or identifying the same hand in various codicological units, is done by visual comparison. The outcome of this greatly depends on the ability of the paleographer to recognize scribal features. Results can be disputed and different paleographers can form different opinions on the same raw material, i.e., the same corpus of manuscripts. Alois Haidinger, who is responsible for most of the data used in this study, is an expert on twelfth-century handwriting, and thus far no doubts have been raised on his observations. For both Heiligenkreuz and Zwettl, about one

third of the scribes could be identified in only one codicological unit.[34] At least in the early phase of the Heiligenkreuz scriptorium, these isolated hands often only contributed a few lines to a codicological unit. Since having only a small sample makes a paleographical comparison more difficult, it is quite possible that the number of unique hands is in fact lower.[35] Until automated scribal identification is better established and proven to be superior to the traditional work of paleographers, this issue cannot be resolved.

For Baumgartenberg the situation is more challenging. Alois Haidinger did investigate a few Baumgartenberg manuscripts, nowadays kept in the Austrian National Library, but most data was provided by my own research. While identifying scribes within the Baumgartenberg corpus is no different from Haidinger's work on Heiligenkreuz and Zwettl, identifying connections between the scriptoria is to a certain extent affected by my previous research. As part of my master's thesis, I compiled lists of the Heiligenkreuz scribes in manuscripts mentioned in the earliest book list, in order to identify hitherto unknown Heiligenkreuz manuscripts in the Austrian National Library.[36] This knowledge helped me considerably in identifying these Heiligenkreuz scribes in early Baumgartenberg manuscripts, without consulting additional manuscripts or images. By contrast, for the latter part of the twelfth century, as well as for all manuscripts from Zwettl, identification is only possible by comparing scribal hands with the data available on scriptoria.at. However, the website does not provide any search options, or even lists of scribes that can be scrolled through. This means clicking through hundreds of manuscripts and thousands of images in the hope of finding matching scribes. It is therefore very likely that some connections between Baumgartenberg and other scriptoria were missed.

For Baumgartenberg, the incompleteness of data is most likely a more serious issue than for Heiligenkreuz and Zwettl, due to the nature of its early book production. Many of the early Baumgartenberg manuscripts are closely connected to the Heiligenkreuz scriptorium, i.e., they were at least partly written by prominent Heiligenkreuz scribes.[37] At the same time, there are not many scribes that can be identified in more than one Baumgartenberg manuscript but not in Heiligenkreuz. A number of manuscripts cannot be connected to other Baumgartenberg manuscripts at all. It therefore seems likely that until the latter part of the twelfth century, there was no productive, independent scriptorium in Baumgar-

34 This number was taken from the data available on scriptoria.at. For the early phase of the Heiligenkreuz scriptorium, Alois Haidinger reports similar numbers; see Haidinger and Lackner, *Bibliothek,* 28. Manuscripts that were clearly not written in the scriptorium were excluded. For imported manuscripts, see the filter "Entstehungsort" in scriptoria.at.

35 Haidinger and Lackner, *Bibliothek,* 28–29, especially n. 76.

36 Kaska, *Neu identifzierte.*

37 For more details and examples, see Kaska, *Schreiber und Werke*.

tenberg.[38] Manuscripts were produced in close collaboration with Heiligenkreuz and also transferred from the motherhouse.[39] Since philological investigations show close textual connections to Zwettl for several texts,[40] it seems reasonable to assume similar exchange processes for scribes that I have so far been unable to identify by paleographical comparisons. Overall, therefore, solely due to limitations in the data collection, the network will be more incomplete for Baumgartenberg than for Heiligenkreuz and Zwettl.

4. Network properties and analysis

After this more general introduction on data acquisition and quality, the following paragraphs will discuss several types of network diagrams obtained from the complete dataset, as well as data on individual scriptoria.

To compare the results from the traditional paleographical investigations discussed above with those obtained by network analysis, I started the investigation by looking solely at the scriptorium of Heiligenkreuz. For its early period, Alois Haidinger has already provided his interpretation of the paleographical data by dating the codicological units and naming important scribes. The early book list also allows the grouping of manuscripts based on external written evidence. For the first network diagram, therefore, only data on Heiligenkreuz manuscripts up to 1200 was included (figure 2). The nodes in this graph are codicological units that are connected via shared scribes (edges). The node size indicates the betweenness centrality, while the colors denote the clusters of codicological units obtained by calculating the modularity in Gephi.[41]

Overall, the graph contains 9 such clusters and 37 isolates, labeled in gray. These are paleographically isolated codicological units. Their scribes could not be identified in any other Heiligenkreuz manuscripts, and they thus appear as islands on the graph. A further two clusters only contain two codicological units each, while one clusters contains three. This indicates that they share scribes, but these scribes cannot be identified in any other manuscript in the main bulk of the network. Two of these clusters are most likely "artifacts" in the sense that they are dated to around 1200 and might be related to later manuscripts not included in the study. However, HLK Cod. 289, ÖNB Cod. 830 part 1 and HLK

38 The situation changes in the early thirteenth century, see Simader, "Österreich," 346–48 for details on the development of book illumination in Baumgartenberg.

39 One example is ÖNB Cod. 726. For another possible example see section 5.

40 Research is still ongoing. For one example see Kaska, "Zur hochmittelalterlichen Überlieferung."

41 The resolution was kept at the default value of 1 and the edge weight was taken into account, which lead to a modularity of 0.399. The distance is a result of the layout algorithm (Force Atlas).

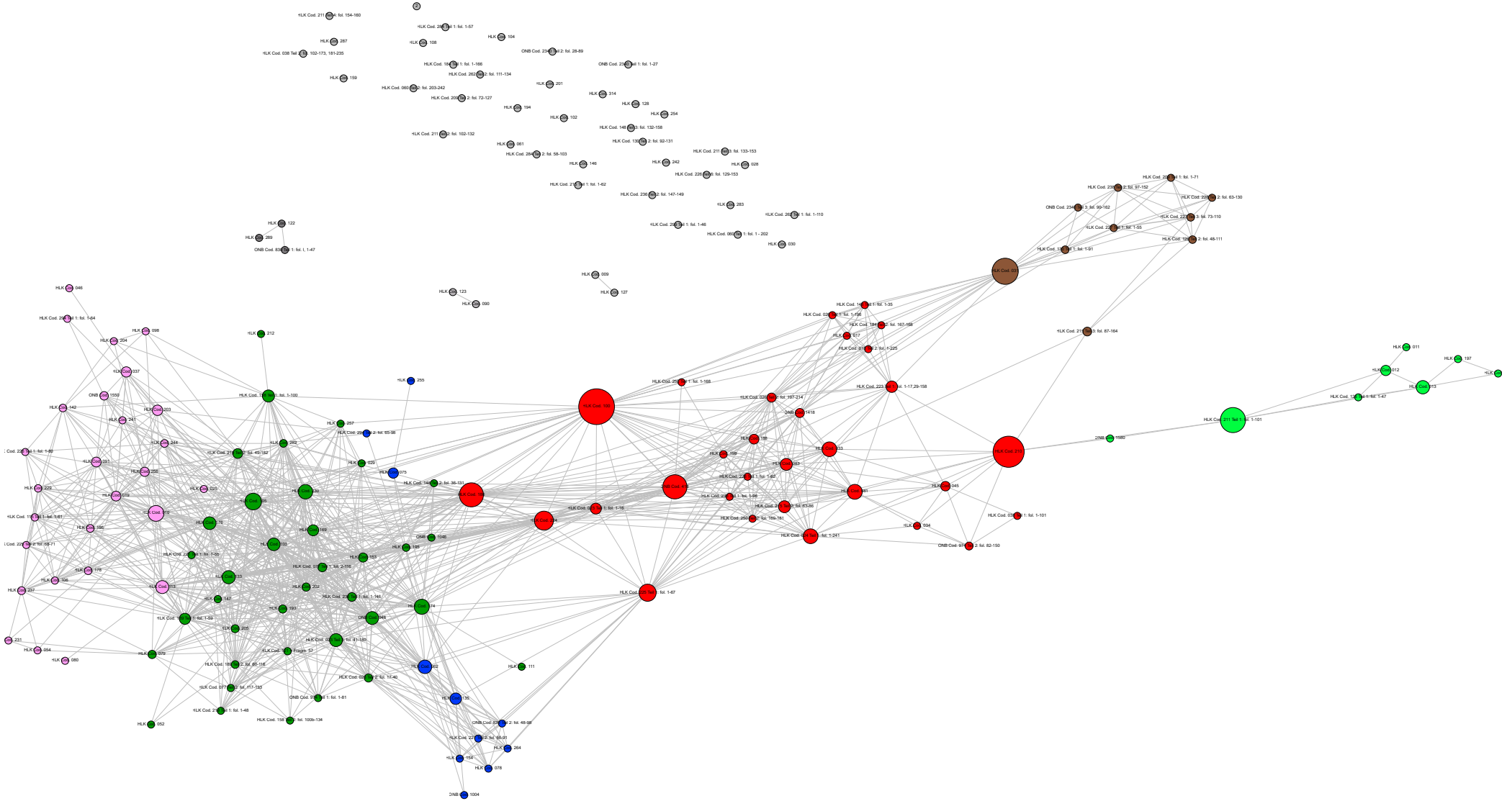

Fig. 2 Network of twelfth-century manuscripts from Heiligenkreuz (node size: betweenness centrality, colors: modularity)

eISSN: 2535-8863
DOI: 10.25517/jhnr.v9i1.192

Cod. 122, which form the third cluster, are mentioned in the earliest book list and would be expected to be part of the main bulk of the network. Closer inspection shows that all three manuscripts show scribal characteristics typical of manuscripts from France (Burgundy?). Nevertheless, they were most likely written in Heiligenkreuz, since scribe HLK 122 A copied the monastery's earliest extant charter.[42] They also share rubricators and correctors with other Heiligenkreuz manuscripts. Since these scribal roles were not included in the networks presented in this paper, they appear to have no connection to the main component of this graph.

This leaves six clusters (modularity classes) with between 8 and 37 elements each that constitute the main bulk of the network. A highly interconnected group forms the center (dark green). This is surrounded by slightly less interconnected groups (dark blue and pink). All three contain manuscripts mentioned in the earliest book list, as well as some not mentioned in the list. This doesn't come as a surprise, since there is no indication of the scriptorium stopping to produce manuscripts after the compilation of the book list. The book list is therefore a snapshot of a still-growing library, and many scribes continued their work into the third quarter of the twelfth century. Further paleographical investigation could help to determine if the division of the main component into three distinct groups can be related to the inner organization of the scriptorium, or any evolution in manuscript production. At this stage it would also be necessary to check the validity of the modularity algorithm, i.e., to determine if these clusters are constructed of smaller clusters and thus if they are hiding further communities (as discussed in the previous section).

A clear evolution can be observed in the clusters on the right-hand side of the graph. A large cluster (red) is dated to the third quarter of the twelfth century on scriptoria.at, and none of the manuscripts are mentioned in the book list. A few codicological units are dated to the middle of the twelfth century (e.g., HLK Cod. 224 or HLK Cod. 225 part 1: fol. 1–67). Considering their close association with slightly later codicological units in this network, a reevaluation of these dates might be necessary. To the far right of the diagram, a small group of codicological units (brown) is mainly connected to the scriptorium through Cod. 31. Also, for another group (light green), only one connection to other codicological units from Heiligenkreuz has so far been established (HLK Cod. 211 part 1). This is due to a single scribe active in this codicological unit (HLK 211 A) who also contributed to HLK Cod. 210, a manuscript highly connected within the large "light green cluster", as well as to ÖNB Cod. 1580. All manuscripts within the "light green cluster" are dated to the late twelfth century. These include the

42 Kaska, "Untersuchungen," 23–24. For French scribes in Heiligenkreuz, see Haidinger and Lackner, *Bibliothek,* 23–26.

so-called *Magnum Legendarium Austriacum* that will be discussed in more detail in section 5.

Overall, the network partition visualizes the development of the scriptorium over time, and shows that the Gephi modularity function does indeed provide reliable results at this level. There is a flush of activity in the early phase spanning from the foundation in 1133/4 to some point after the compilation of the book list (1147 at the latest). The scriptorium then enters – without a clear break – a new phase at some point in the third quarter of the twelfth century, where a new generation of scribes takes over (red). Towards the end of the century, two more distinctive groups can be distinguished that might be connected to later manuscripts not included in this study.

In his book on the early Heiligenkreuz scriptorium, Alois Haidinger identifies a list of the main scribes up to the third quarter of the twelfth century who also participated in the production of the manuscripts mentioned in the book list compiled before 1147. These scribes contributed a large number of pages to the manuscripts and are thus present in a large number of codicological units. In a scribal network where the number of copied pages is not taken into account, one parameter for the importance of a scribe for the scriptorium is the number of other scribes he collaborated with.[43] Figure 3 therefore shows a network of scribes where the size of the node is correlated with its degree. The largest nodes are the scribes Udalricus and Leonhardus, followed by Heinricus, HLK 19 A, HLK 10 A, HLK 230 A, HLK 202 A and HLK 23 C. The ranking is similar to the list published by Haidinger, even though it is based on a larger dataset and uses a different ranking parameter.[44] In the case of the Heiligenkreuz scriptorium, the interconnectedness of a scribe is therefore a good indication of his importance for the scriptorium.

Overall, for Heiligenkreuz all results from network analysis agree well with published databases from traditional paleographical investigations. This implies that the network analysis, and especially the algorithm used for network partition in this study, can indeed serve as a tool to investigate scriptoria where raw data is available but limited analyses have been published, as is the case for Zwettl or for the interaction between Heiligenkreuz, Baumgartenberg and Zwettl. The next step was therefore to include all manuscripts from Heiligenkreuz, Baumgartenberg and Zwettl up to 1200 (for which paleographical data was available) in a network of codicological units (figure 4).

43 The issue of multi-scribe manuscripts discussed in section 2 has to be kept in mind when using the degree parameter.

44 In Haidinger's table, only the number of codicological units to which the scribes contributed is taken into account.

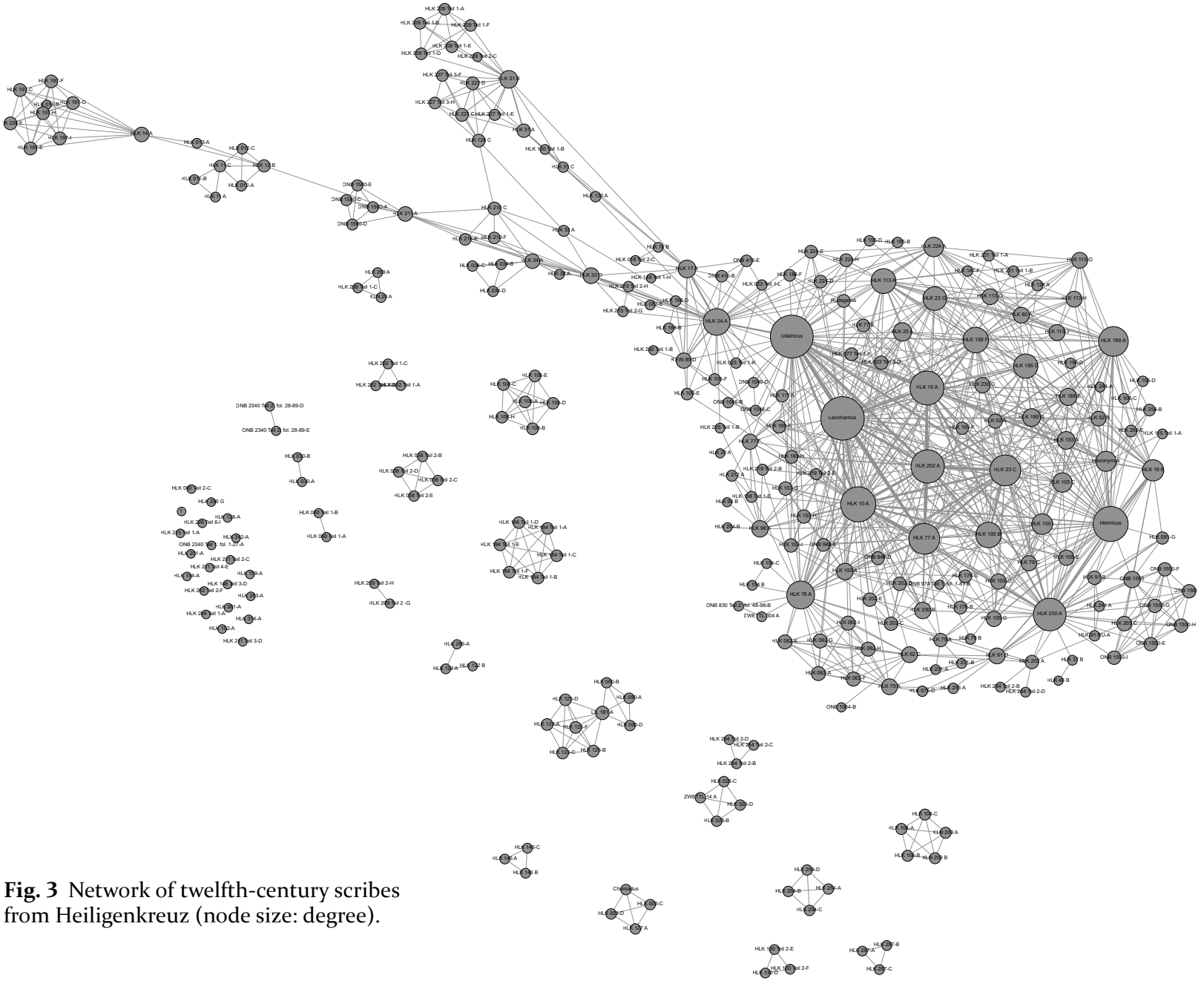

Fig. 3 Network of twelfth-century scribes from Heiligenkreuz (node size: degree).

eISSN: 2535-8863
DOI: 10.25517/jhnr.v9i1.192

The 365 nodes in this graph are codicological units, while the edges are shared scribes. There are eight interconnected clusters with seven to 79 elements each[45] that include 268 nodes in total. These larger clusters are represented by different colors in figure 4. The remaining nodes are part of very small clusters of two to three codicological units (16 nodes) or isolated codicological units[46] (grey in figure 4). 2182 edges connect these nodes.

The basic layout for the main component of the early Heiligenkreuz scriptorium (green) and its later development (red) is similar to the data that can be seen from Heiligenkreuz alone in figure 2. A small group of codicological units from Baumgartenberg is also part of the green cluster (e.g., OOeLB Cod. 318, OOeLB Cod. 328, OOeLB Cod. 319). This comes as no surprise, since several Heiligenkreuz scribes contributed to these codicological units. However, from additional scribal contributions as well as philological investigations it seems likely that they were in fact written in Baumgartenberg.[47] They are the result of a scribal exchange between motherhouses and daughterhouses that is not obvious from network analysis alone. A further small group of codicological units from Baumgartenberg at the bottom of the graph (light pink) is directly connected to the earlier phase of the Heiligenkreuz scriptorium.

Of particular interest are Heiligenkreuz's connections to Zwettl, which can be found in various groups. Zwettl Cod. 91 is an integral part of the earliest Heiligenkreuz group (green) and it is assumed that the manuscript was in fact written in Heiligenkreuz.[48] Considering the results for Baumgartenberg, this should be confirmed by philological or art historical analysis if possible.[49] The main bulk of the Zwettl scriptorium is correctly identified by the Gephi algorithm and represented by the blue group in the upper half of the graph. Attached to this group are further groups of codicological units from Zwettl, which are in turn connected to codicological units from Heiligenkreuz (brown, pink and yellow). There is a clear connection from the early Heiligenkreuz scriptorium (green) via a small group of codicological units (pink) to Zwettl. As will be shown in the scribal network, they share the scribe HLK 98 that worked for both scriptoria. A similar role is played by scribe HLK 203 A for the "brown group". The "pink group" is also connected through Zwettl Cod. 6 to another distinct group of the Zwettl scriptorium

45 For calculating the modularity the resolution was kept at the default value of 1 and the edge weight was taken into account, which lead to a modularity of 0.601. The distance is a result of the layout algorithm (Force Atlas).
46 For a discussion of these islands see previous paragraphs.
47 See in detail Kaska, "Schreiber und Werke."
48 See www.scriptoria.at under the shelfmark.
49 Zwettl Cod. 293 part 3 is also part of the same group. Only one scribe contributed to the manuscript (HLK 10 A), who worked on both Heiligenkreuz and Zwettl manuscripts, which makes any attribution less certain.

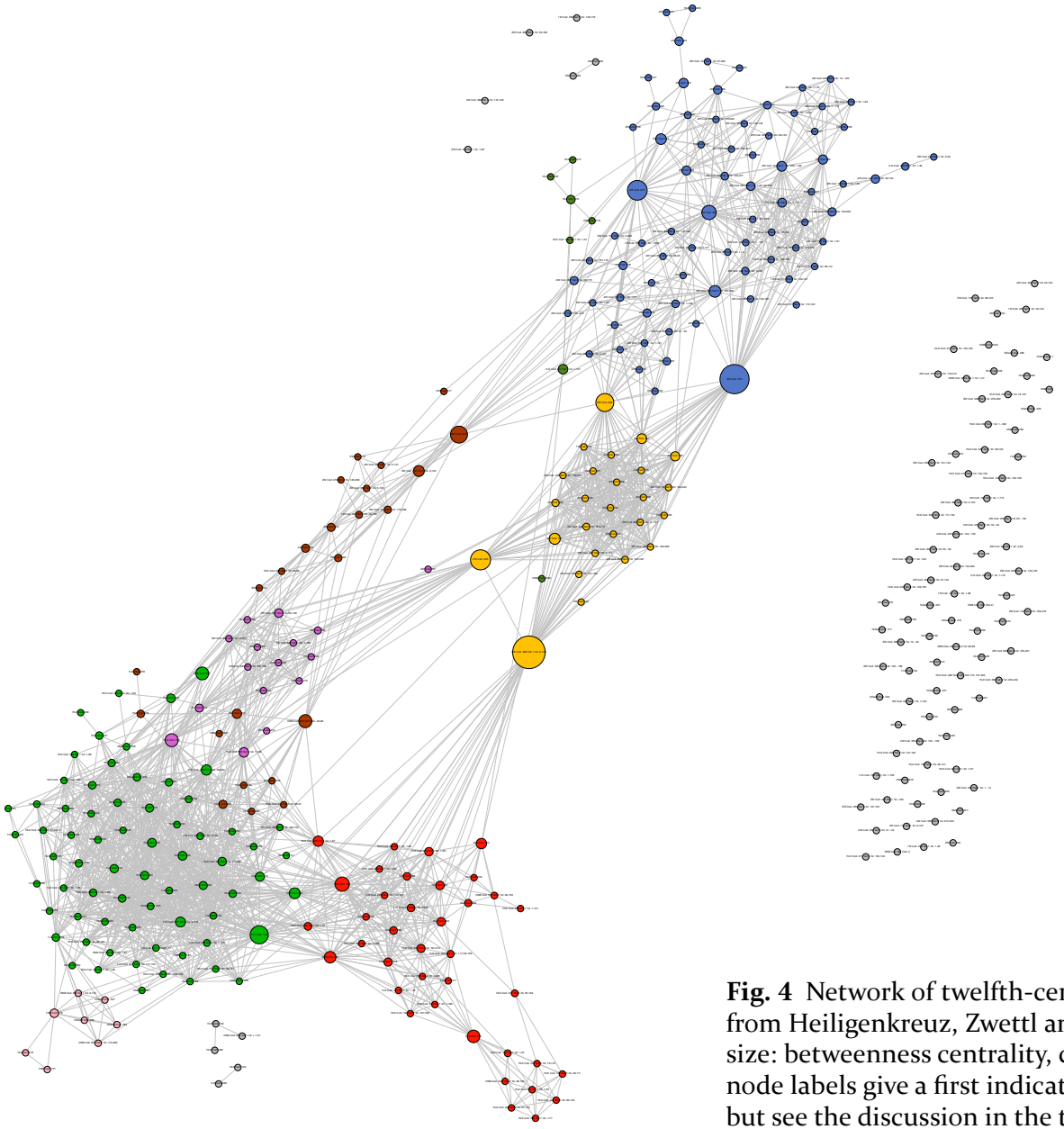

Fig. 4 Network of twelfth-century codicological units from Heiligenkreuz, Zwettl and Baumgartenberg (node size: betweenness centrality, colors: modularity). The node labels give a first indication of the place of origin, but see the discussion in the text.

eISSN: 2535-8863
DOI: 10.25517/jhnr.v9i1.192

(yellow). This group is in turn connected to the later Heiligenkreuz scriptorium (red) through the first part of Zwettl Cod. 299. This connection relies solely on the scribe HLK 24 A, one of the most prolific Heiligenkreuz scribes of that period. Apart from Cod. 299, he only contributed to one other codicological unit from the library of Zwettl, which may have been written in Heiligenkreuz (Cod. 295 part 1).[50] The connection between the later Heiligenkreuz and the Zwettl scriptorium is therefore quite weak. The network graph implies that there was more interaction between the motherhouse and the daughterhouse in the middle and perhaps the beginning of third quarter than towards the end of the twelfth century. This confirms recent investigations and manuscript datings and contradicts the assertion in older literature that book production in Zwettl only really started in the early 1170s.[51] It could also point towards a development of the Zwettl scriptorium over time towards greater independence. The results will have to be investigated in greater detail again and correlated with additional data and analyses of the manuscripts from Zwettl.

The most important change to the network by including Zwettl manuscripts occurs in the group around the *Magnum Legendarium Austriacum* (light green group in figure 2). In figure 4, some of these codicological units are part of a group (olive) that is closely connected to the Zwettl scriptorium (blue) in the upper part of the graph, but also has connections to Heiligenkreuz (e.g., Cod. 13 and 14).[52] Others even become part of the main group (blue) of the Zwettl scriptorium (e.g., HLK Cod. 11 and 12). This peculiarity will be discussed in more detail later.

A second option to learn about the interaction between the scriptoria is to look at the interaction of scribes instead of manuscripts/codicological units, i.e., to construct a network with scribe-nodes and edges that are codicological units from the complete data and then calculate the modularity for clustering (figure 5).[53] The number of nodes (570) is much larger than figure 4, which shows the network of codicological units, while the number of edges is lower (1819).

---

50 See the list of HLK 24 A's contributions at: https://www.scriptoria.at/cgi-bin/rel_scribes.php?scribe_name=HLK%2024%20A (accessed May 25, 2022).

51 Rössl, "Schriftlichkeit."

52 A few other Heiligenkreuz manuscripts dated to the period around 1200 also moved to the outskirts of the Zwettl network. Here, the inclusion of thirteenth-century manuscripts might change the network properties. Also, Heiligenkreuz Cod. 299 part 1 is, in this representation, closely connected to the Zwettl scriptorium (on the edge of the blue group). However, the manuscript was actually written in the Augustinian monastery of Klosterneuburg and is only connected via one scribe from Klosterneuburg that also worked in Zwettl. It would be interesting to further investigate this interaction between the Cistercian scriptoria and the Augustinian house.

53 For calculating the modularity the resolution was kept at the default value of 1 and the edge weight was taken into account, which lead to a modularity of 0.683. The distance is a result of the layout algorithm (Force Atlas).

The large number of nodes corresponds to the large number of scribes that contributed to the codicological units (see discussion in section 3). Since in most cases we deal with multi-scribe manuscripts, it is not surprising that the number of communities detected in the data is also much higher than for the network of codicological units. Several scribes in one codicological unit already constitute a community in this representation. In total, 106 clusters were found, eleven of which include eight or more scribes. The largest cluster, which represents the early scriptorium of Heiligenkreuz, includes 96 scribes. All clusters with less than eight scribes were represented in grey in the graph. Many of these are isolated clusters, i.e., none of the scribes in a manuscript could be identified with known scribes from Heiligenkreuz, Zwettl or Baumgartenberg. They are arranged around the central part of the network in no particular order.

Both the scriptoria of Heiligenkreuz (bottom part with green and red main groups) and Zwettl (top with blue and purple main group) include several scribal communities with overlapping groups in the middle of the graph where scribes worked for both scriptoria. At the very bottom (brown), the scriptorium of Baumgartenberg is visible. These communities can give an insight into the inner workings of the scriptoria, highlighting which scribes regularly collaborated, and thus might also show a relative chronology of the scribes. Answering these questions in full would require a more in-depth investigation of the manuscripts and the addition of further data from rubrics and correctors, which is not within the scope of this study.

One parameter that can point towards exchange processes, however, can be easily determined. While in figure 2 the Heiligenkreuz scribes with the highest number of collaborations (highest degree) were highlighted, in figure 4 the most important scribes for connecting various part of the network, i.e., those with a high betweenness centrality, are visible. Higher betweenness centrality correlates with larger node size.

For the early Heiligenkreuz scriptorium (green) the largest nodes (and therefore the most important scribes) for network construction are: Udalricus, Heinricus and Leonardus, as well as the anonymous hands HLK 10 A and HLK 19 A. Udalricus and Heinricus are also among the most important scribes in the early phase of the scriptorium when it comes to the number of codicological units to which they contributed,[54] and have high degree values. This indicates that all three values can be useful for determining the importance of scribes within a scriptorium. In future analyses, this finding can be expanded to determine if a combination of these values can help to investigate the inner structure of the scriptorium.

54 Haidinger and Lackner, *Bibliothek,* 29.

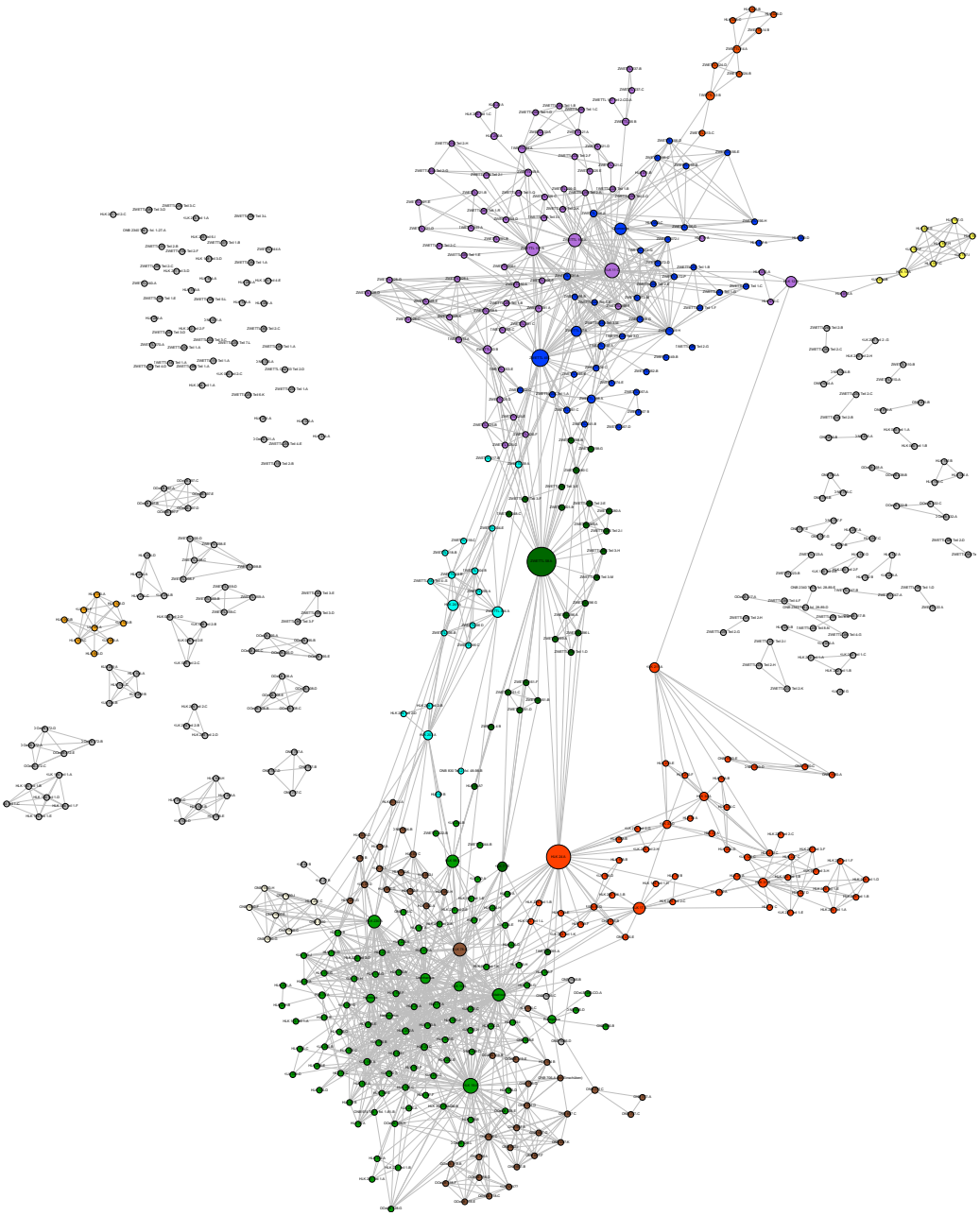

Fig. 5 Network of twelfth-century scribes from Heiligenkreuz, Zwettl and Baumgartenberg (node size: betweenness centrality, colors: clusters). The node labels give a first indication of the scribes' location, but see the discussion in the text.

From this network it appears that the most important connectors to the scriptorium of Baumgartenberg (brown group) are HLK 19 A and HLK 78 A. HLK 19 A contributes a few lines to several Baumgartenberg manuscripts, which might point towards a role as a manager or teacher.

For the later phase of the Heiligenkreuz scriptorium (red), the main connectors are HLK 24 A and HLK 17 A. HLK 17 A mainly connects to a group of scribes around HLK 31 B, while HLK 24, as discussed previously, brings together parts of the Heiligenkreuz scriptorium with Zwettl.

The turquoise group is even more closely connected to Zwettl than is obvious from the diagram. The scribes HLK 241 A and HLK 203 A actually contributed to fewer codicological units from Heiligenkreuz than from Zwettl. The main connector between the Heiligenkreuz main group (green) and this turquoise group is HLK 98 A, who contributed roughly equal numbers of codicological units to both scriptoria. Another large node in the intersection between Zwettl and Heiligenkreuz is ZWETTL 58 A as the center of the olive group. This scribe was already determined as a main scribe of the scriptorium in an older publication by Charlotte Ziegler.[55]

Other main members of the Zwettl scriptorium are found in the "blue group" as well as the "purple group". Central to this part of the network is ZWETTL 49 L, who is active in 19 codicological units as a text scribe. His importance to both the scriptorium and the network is even greater if one considers his contributions as a corrector and rubricator, which bring the number of his codicological units up to 49.[56] Further important connectors are Zwettl 77 A (blue), Zwettl 101 A (purple), and Zwettl 194 A (purple), who all contribute to several codicological units. Of similar importance is the scribe HLK 11 C (purple) named after HLK Cod. 11, one of four surviving volumes of the Heiligenkreuz *Magnum Legendarium Austriacum*.

5. Using the network

Using network analysis for paleographical data is different from many other applications of the technique in that considerable analytical work is needed to acquire the data. By comparing dozens of manuscripts and even more scribal hands by traditional means, a paleographer usually gains deep insight into the inner working of a scriptorium, i.e., who were the most important scribes, which

55 Ziegler, *Zisterzienserstift Zwettl,* xvi.

56 See his profile at: https://www.scriptoria.at/cgi-bin/rel_scribes.php?scribe_name=ZWETTL%2049%20L (accessed May 25, 2022). He is the main corrector for the Zwettl scriptorium.

scribes worked together etc., even before compiling the data necessary for network analysis.[57] This is especially true if the data is compiled into a database – a necessary step if one is to compare many hundreds of manuscripts. The basic network analysis preformed in this study therefore did not yield striking new results, but rather visualizes information already available to the researcher.

The situation is quite different for the user of a paleographical database such as scriptoria.at. For the earliest Heiligenkreuz manuscripts, Alois Haidinger has published some of his findings in a reader-friendly printed form.[58] He focuses on French influence in Heiligenkreuz – a topic that the database cannot cover – as well as on some of the most important scribes of this period. A similar analysis will be published on the Zwettl scriptorium.[59] For later manuscripts as well as many of the other scribes, researchers are left to work with scriptoria.at, where information can only be accessed through lists of manuscript shelfmarks (see introduction). This is where a network diagram becomes useful, as it allows the reader to quickly work through a vast amount of data about scribes in an accessible visual form. It shows which manuscripts/codicological units or scribes are related and, in the case of Heiligenkreuz, even shows the development of the scriptorium over time. It can also visually confirm hypotheses on manuscript identification and dating, and even point out further areas of research, as the following examples will show.

For two items from the Heiligenkreuz book list, it is argued by paleographers and art historians that the entries in the book list cannot be identified with extant copies of the text in the monastery's library. These manuscripts are said to have been written later, replacing earlier copies. One such case is Hugh of St. Victor's *De sacramentis*. We possess a manuscript with this text, Heiligenkreuz Cod. 100, dated to the latter part of the twelfth century, while the copy mentioned in the book list must have been copied before the middle of the twelfth century.[60] It is possible that the earlier copy made at Heiligenkreuz was given to the library of its newly established daughterhouse of Baumgartenberg shortly after the book list was compiled and survives in the Oberösterreichische Landesbibliothek (State Library of Upper Austria) in Linz. OÖLB Cod. 319 was already part of the library of Baumgartenberg in the Middle Ages, but was mainly copied by scribes from Heiligenkreuz. It shows the same *lacunae* as Heiligenkreuz Cod. 100, which cannot be found in any other Austrian copy of the text.[61] Similarly, Augustine's *Confessiones* is mentioned in the Heiligenkreuz book list, but today only survives in a manuscript from the third quarter of the twelfth century (Heiligenkreuz

57 See section 2 and especially note 17 for how to do scriptorium research.
58 Haidinger and Lackner, *Bibliothek*.
59 Expected as a volume in the series *Codices manuscripti et impressi. Supplementum*.
60 Haidinger and Lackner, *Bibliothek,* 16, n. 25.
61 Kaska, "Schreiber und Werke," 78–83.

Cod. 24).[62] It may be a copy of the codex mentioned in the book list that did not survive.

Indeed, both Cod. 100 and Cod. 24 are part of a group (red in figure 4) that shows the Heiligenkreuz scriptorium after its initial phase of production, and clearly cannot have been part of the library in the middle of the twelfth century. In turn, OÖLB Cod. 319 is closely connected to the first phase of production in Heiligenkreuz and therefore might well be the lost Heiligenkreuz copy. Network analysis therefore confirms previous assumptions about book loss and transfer based on traditional paleographical and philological methods.

As mentioned previously, another paleographically interesting group of codicological units and scribes centers on the Heiligenkreuz copy of the so-called *Magnum Legendarium Austriacum* (MLA, Heiligenkreuz Cod. 11–14). Recent work by Diarmuid Ó Riain has given deeper insight into the compilation and distribution of this voluminous hagiographical collection.[63] The multi-volume legendary survives in late twelfth to early thirteenth-century copies of varying completeness in the Benedictine monasteries of Melk (Lower Austria) and Admont (Styria), the Cistercian monasteries of Heiligenkreuz, Zwettl and Lilienfeld (all Lower Austria), as well as the Augustinian house of St. Pölten[64] (Lower Austria).. A few fragments of yet another copy were recently found in the Benedictine monastery of Göttweig in Lower Austria.[65] Ó Riain proposes that the MLA was compiled in Admont, where a direct copy of this "Ur-MLA" still exists. For the other manuscripts, a single intermediary copy β is proposed that served as the exemplar for the Heiligenkreuz copy. From β, a lost copy γ derives that then serves as the exemplar for the manuscripts from Melk, St. Pölten and Zwettl (Cod. 13–15 and 24). According to Ó Riain, the copies from Heiligenkreuz and Zwettl are therefore not directly related.[66]

Paleographical investigation plays a lesser role in Ó Riain's publications and is only used for localizing and dating the manuscripts. However, it might be interesting to look at the scribes and their collaborators again in the future and gain a deeper insight into the role of the MLA within the scriptoria of Heiligenkreuz and Zwettl, as well as their collaboration. The network diagram (figure 4) firmly places Heiligenkreuz Cod. 11 and 12 in the Zwettl scriptorium, while Cod. 13 and Cod. 14 are part of a small group connecting the scriptoria of the motherhouse and the daughterhouse. Overall, using the data available on scriptoria.at at the

62 Haidinger and Lackner, *Bibliothek,* 12, n. 11.
63 Ó Riain, "Magnum Legendarium"; Ó Riain, "Neue Erkenntnisse." For a list of manuscripts see: http://mla.dingbat.at/ (accessed May 25, 2022).
64 Today kept in the Austrian National Library.
65 Ó Riain, "Neue Erkenntnisse," 3–6.
66 For the stemma, see Ó Riain, "Magnum Legendarium," 153 and 141.

point of the writing of this article, the scribes of the Heiligenkreuz MLA are more closely connected to Zwettl than to Heiligenkreuz. This is at odds with the art historical findings that see part of the initials in the Heiligenkreuz MLA indebted to a group of Heiligenkreuz manuscripts from the fourth quarter of the twelfth century, and therefore assume Heiligenkreuz as the place of production.[67] It is possible that Heiligenkreuz scribes simultaneously worked for both Heiligenkreuz and Zwettl as is most likely the case for Heiligenkreuz scribes working for Baumgartenberg in an earlier period. It is likewise possible that we can see a transfer of personnel with e.g., scribe HLK 11 C moving from Heiligenkreuz to Zwettl, or vice versa. For firm conclusions, the interaction between the MLA scribes and other scribes should be investigated in more detail and correlated with information on the illuminators and their work, not only in the MLA but also in the other manuscripts from the two scriptoria. While this intriguing connection between the Heiligenkreuz MLA and Zwettl is visible from the lists provided on scriptoria.at, network analysis shows the extent of this interaction at a glance.

Despite all the challenges in data acquisition and interpretation, network analysis for paleographical data provides a valuable additional tool for accessible data visualization, and thus helps to disseminate paleographical research to a wider audience. It can encourage researchers to reevaluate previous results and identify new research questions. The results in this study also point towards the possibility of further, more advanced network analyses to learn more about scribal collaborations and the organization of a scriptorium that cannot be easily investigated by traditional methods alone.

6. References

Andrist, Patrick, Paul Canart, and Marilena Maniaci. *La syntaxe du codex. Essai de codicologie structurale.* Bibliologia 34. Turnhout: Brepols, 2015.

Cohen-Mushlin, Aliza. *A Medieval Scriptorium. Sancta Maria Magdalena de Frankendal*. Wolfenbütteler Mittelalter-Studien 3. Wiesbaden: Harrassowitz, 1990.

Egger, Christoph. "Die Suche nach dem archimedischen Punkt. Methodische Probleme der Erforschung von Scriptorien und Buchproduktion im 12. Jahrhundert am Beispiel von Admont." In *Scriptorium. Wesen – Funktion – Eigenheiten. Comité international de paléographie latine, XVIII. Kolloquium. St. Gallen 11.–14. September 2013,* edited by Andreas Nievergelt et al., 375–89. Munich: Beck, 2015.

67 See the short description by Susanne Rischpler at: http://mla.oeaw.ac.at/#/decorations (accessed May 25, 2022).

Fortunato, Santo, and Marc Barthélemy. "Resolution limit in community detection." *Proceedings of the National Academy of Sciences* 104, no. 1 (2007): 36–41.

Frioli, Donatella. "I più antichi manoscritti dello scriptorium di Aldersbach." In *Beiträge zur mittelalterlichen Geschichte des Klosters Aldersbach: Bericht zur interdisziplinären Tagung "Mittelalterliche Geschichte des Klosters Aldersbach" am 1. und 2. Oktober 2020,* edited by Robert Klugseder, 201–35. Studien und Mitteilungen zur Geschichte des Benediktinerordens und seiner Zweige. Ergänzungsband 5. St. Otilien: Eos, 2020.

Garand, Monique-Cécile. "Manuscrits monastiques et scriptoria aux XIe et XIIe siècles." In *Codicologica 3: Essais typologiques,* edited by Albert Gruys and Johan P. Gumbert, 9–33. Leiden: Brill, 1980.

Golob, Nataša, *Twelfth-century Cistercian manuscripts: the Sitticum Collection.* Ljubljana: Slovenska Knjiga, 1996.

Gottlieb, Theodor. *Mittelalterliche Bibliothekskataloge Österreichs 1: Niederösterreich.* Vienna: Holzhausen, 1915.

— *Über mittelalterliche Bibliotheken.* Leipzig: Harrassowitz, 1890.

Haidinger, Alois. "Bibliothek und Skriptorium des Stiftes Heiligenkreuz." Accessed May 25, 2022. https://www.scriptoria.at/cgi-bin/index.php.

Haidinger, Alois, and Franz Lackner. *Die Bibliothek und das Skriptorium des Stiftes Heiligenkreuz unter Abt Gottschalk (1134/1147).* Codices manuscripti et impressi. Supplementum 11. Purkersdorf: Holinek, 2015.

Kaska, Katharina. "Schreiber und Werke: Ein Vergleich paläographischer und textlicher Beziehungen am Beispiel der österreichischen Zisterzienserklöster Heiligenkreuz und Baumgartenberg als methodischer Zugang zur Untersuchung monastischer Netzwerke." In *Die Bibliothek – The Library – La Bibliothèque: Denkräume und Wissensordnungen,* edited by Andrea Speer and Lars Reuke, 61–94. Miscellanea Mediaevalia 41. Berlin: De Gruyter, 2020.

— "Untersuchungen zum mittelalterlichen Buch- und Bibliothekswesen im Zisterzienserstift Heiligenkreuz." MA Thesis, University of Vienna, 2014.

— "Zur hochmittelalterlichen Überlieferung von Maximus Confessor, Capita de caritate in der Übersetzung des Cerbanus." *Jahrbuch der Österreichischen Byzantinistik* 70 (2020): 221–48.

Lutter, Christina. "Zisterzienser und Zisterzienserinnen in und um Wien." *Historisches Jahrbuch* 132 (2015): 141–77.

Ó Riain, Diarmuid. "Neue Erkenntnisse zur Entstehung und Überlieferung des Magnum Legendarium Austriacum." *Mitteilungen des Instituts für österreichische Geschichtsforschung* 128, no. 1 (2020): 1–21.

— "The Magnum Legendarium Austriacum: a new investigation of one of medieval Europe's richest hagiographical collections." *Analecta Bollandiana* 133 (2015): 87–165.

Paulhart, Herbert. *Mittelalterliche Bibliothekskataloge Österreichs 5: Oberösterreich.* Vienna: Böhlau, 1971.

Rössl, Joachim. "Schriftlichkeit, Buchkunst und geistige Kultur." In *Stift Zwettl und seine Kunstschätze,* edited by Karl Kubes and Joachim Rössl, 29–33. St. Pölten: Pressehaus, 1979.

Simader, Friedrich. "Österreich." In *Geschichte der Buchkultur Bd. 4/2: Romanik,* edited by Andreas Fingernagel, 327–77. Graz: Adeva, 2007.

Tibbetts, Sidney. "Praescriptiones, student scribes and the Carolingian scriptorium." In *La collaboration dans la production de l'écrit médiéval. Actes du XIIIe colloque du Comité international de paléographie latine (Weingarten, 22–25 septembre 2000),* edited by Herrad Spilling, 25–35. Paris: Librairie Droz, 2003.

Zák, Alphons. *Österreichisches Klosterbuch. Statistik der Orden und Kongregationen der katholischen Kirche in Österreich.* Vienna: Kirsch, 1911.

Ziegler, Charlotte. *Zisterzienserstift Zwettl. Katalog der Handschriften des Mittelalters.* 4 vols. Vienna: S.O.Cist.Zwettl, 1985.

The Journal of HISTORICAL NETWORK RESEARCH

DOMINIQUE STUTZMANN/
LOUIS CHEVALIER

Books of hours as codified compilations of compilations

Textual networks and hybrid liturgical uses

Journal of Historical Network Research 9 (2023) 130–183

Keywords Books of hours, Middle Ages, Manuscript studies, Liturgy, Philology, Low Countries, Hours of the Virgin, Office of the Dead

Abstract The present contribution aims to change how we study the textual content of books of hours by tackling the most common texts at a large scale. Intended for lay people and imitating the model of liturgical books, books of hours contain a core of votive offices and appear to have a very standardized content. The choice and the order of the chants, readings and prayers may vary within the offices according to not only the liturgical destination, but also the place of production, the target export market, and the choices of the client. Variations are therefore difficult to characterize and analyze. Here, we focus on an analysis of the Hours of the Virgin and the Office of the Dead as both compilations and networks of compilations. At the level of texts and liturgical uses, we highlight and study textual commonalities based on geography or other historical links (e.g., Germany, the Dominican order and southern France for the Hours of the Virgin, Flanders and Scandinavia, Poitiers and Bordeaux, Auxerre and Bayeux for the Office of the Dead). Since patrons or copyists could also modify the expected contents, our last part analyses the uses of Utrecht and Bruges, and how uses specific to one institution may either be faithfully reproduced or give way to hybridization. This phenomenon is characterized, for example, by inserting pieces from another use into a well-identified set.

C²DH
LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

UNIVERSITY OF LUXEMBOURG

1. Introduction

A map to navigate the haystack? We are not there yet, but textual network analysis can help us to understand the composition of a very large rick, the numerous books of hours, and how the very many tiny bits and pieces of text are assembled within them. The aim of this paper is to introduce some tools of network analysis in the study of books of hours to help drawing general maps of textual and liturgical proximities.

Although they are known to all Medievalists, especially to art historians, books of hours are often overlooked or disdained as a standard mass production, and their texts are more often than not left to one side. Yet, as the bestseller of the Middle Ages, they played an important role in the cultural life of their time period and transmitted different sets of texts, so that they now constitute precious sources on medieval devotion, liturgy, social representation, and textual circulation in Western medieval society. Finding textual commonalities and connections in such a large body of sources is a challenge. Through an in-depth analysis, scholars have been able to reveal the correlations between illuminators and copyists, resulting in some more or less coherent groups of manuscripts. Textual commonalities may be linked to local habits or originate from workshop practices, whether in the choice of suffrages or prayers, or the compilation and configuration of variant pieces.[1] The most flexible section is the calendar, and the variations in "strictly regional feasts",[2] as well as the arbitrary parts of "full" calendars, may be both daunting and highly revealing, as shown by J. Plummer, G. Clark, S. van Bergen, M. Hülsmann and T. Kren.[3] Thanks to the work of J. Plummer and

Acknowledgements: This research was funded by the French National Research Agency under the grant agreement ANR-17-CE38-0008. The datasets and code to produce the tables and figures are available at https://doi.org/10.5281/zenodo.6671462. The authors wish to very warmly thank Saskia van Bergen, Martin Morard and the anonymous reviewers for their comments, suggestions and contributions.

Corresponding author: Dominique Stutzmann, dominique.stutzmann@irht.cnrs.fr

1 For the region of particular focus in this paper, see Bergen, 'De Meesters van Otto van Moerdrecht. Een onderzoek naar de stijl en iconografie van een groep miniaturisten, in relatie tot de productie van getijdenboeken in Brugge rond 1430'. In section I.4 on the collaboration of 'librarians', copyists and illuminators, she studies the material features (dimensions), hands and their changes for different texts, in relation to texts and liturgical uses. In section I.5, she highlights networks of illuminators, but also based on textual similarities in calendars, which are supported by appendix 12 on clustering the calendrical information.

2 Clark, 'Table of Strictly Regional Feasts in Calendars Printed or Written for Southern Netherlandish and Contiguous Dioceses, Abbeys, and Colleges before 1559'; Clark, 'Appendix III. Qu'est-ce qu'un calendrier parisien?'

3 Plummer, '"Use" and "beyond Use"'; Hülsmann, 'Text Variants in the Utrecht Calendar. A Help in Localizing Dutch Books of Hours'; Hülsmann, 'De wisselende samenstelling van de Utrechtse heiligenkalender: een onderzoek naar de taakverdeling bij het afschrij-

G. Clark, we also know that other sections bear traces of their textual connections, with a philological basis for the prayers *Obsecro Te* and *O intemerata,*[4] or with regard to the choice of versicles in the litanies.[5]

Scholarly interest in the text of the books of hours has sharply increased in recent decades. In a recent book, K. Rudy has scrutinized the textual features of Netherlandish books of hours and analysed how the Vernacular/Latin bilingualism played a role in the perceived efficiency of prayers, not only from the perspective of the copyist's literacy, but also in terms of the ability of the readers to articulate the texts with devotion, or to understand and act according to the prescriptions.[6] She addressed several specific texts (e.g., Mass of Saint Gregory or prayers to the Virgin), as well as how images would underline the expected outcome and indulgences of reading the Hours of the Virgin or the Office of the Dead, including examples from manuscripts containing the offices in Latin.

Still, more often than not, the textual contents of the main offices remain beyond the scope of investigation. The present article introduces the concepts and tools of network analysis in order to address the corpus at a larger scale and aims at revising how we identify textual circulations. We also discuss the notion and process of hybridization to explain the state of certain manuscripts and/or texts. Inspired by works on the most variable parts of books of hours, such as calendars and litanies,[7] our main contribution is to investigate the parts of books of hours which are the most common and deemed the most stable, namely the Hours of the Virgin and the Office of the Dead.

We also define and embrace books of hours as codified compilations of compilations. Not only do books of hours as a genre gather offices in a complex fashion, but offices and suffrages have, by their very nature, the quality of compilations of individual pieces, composed according to specific liturgical rules. Here we address the question of textual networks at the level of offices and hours, but this enquiry could also be extended to the level of manuscripts. Our study of networks and hybridisation will have a specific focus on the Low Countries to better

---

ven en het decoreren van handschriften'; Kren, 'Seven Illuminated Books of Hours Written by the Parisian Scribe Jean Dubreuil, c. 1475–1485'.

4 Plummer, '"Use" and "beyond Use"'; Clark, 'Beyond Jacquemart Pilavaine, Simon Marmion, and the Master of Antoine Rolin: Book Painting in the Hainaut in the Penultimate Decade of the Fifteenth Century', 396–97.

5 Clark, 'Beyond Saints: Variant Litany Readings and the Localization of Late Medieval Manuscript Books of Hours. The D'Orge Hours'.

6 Rudy, *Rubrics, Images and Indulgences in Late Medieval Netherlandish Manuscripts.*

7 For a database of calendars in books of hours, for now without exploitation, see also Macks, 'CoKL: Corpus Kalendarium'. For the network of liturgical calendars, but not connected to books of hours, see Heikkilä and Roos, 'Quantitative Methods for the Analysis of Medieval Calendars'.

understand how the interdependence of devotion, commercial imperatives, and mass production gave birth to hybrid texts.

Section 2 provides some definitions and characterizes books of hours as multi-layered compilations, combining different sections in different orders, and associating offices for different liturgical uses, which are themselves compilations of texts and pieces. Section 3 investigates anew the available data on the Office of the Dead and the Hours of the Virgin. First, we will reassess the results from a relatively recent study on the Office of the Dead using techniques from network analysis, and then expand to the Hours of the Virgin. We will demonstrate how network analysis can reveal unexpected proximities between uses. In section 4, we will then address some specifics of book production in the Low Countries and discuss the process of hybridisation. Specifically, we will re-evaluate the links between the uses of the Dominican order and of Utrecht, and analyze a liturgical use previously thought to be for Bruges in order to demonstrate that it can be understood as a hybrid.

2. Definitions

2.1 Offices and liturgical uses

Books of hours assemble pieces of different text genres and contain several texts. "Offices" or "hours" are at their core. They are the sets of prayers that monks and secular clerks are supposed to duly recite (hence the Latin name *"officium",* duty) in order to sanctify the day with prayers, marking the passing time of each day and night, generally distributed in eight hours (Matins, Lauds, Prime, Terce, Sext, None, Vespers, Compline – hence the Latin name *"horae",* hours). The description of the content of each office is called *"cursus"* (i.e., succession of texts).

Each of the liturgical offices that compose a book of hours is in itself a compilation, built from hundreds of different pieces, and these pieces pertain to specific categories. In the following pages, we will use the liturgical vocabulary to denominate these categories (e.g., antiphons, benediction or blessing, canticle, chapters, hymns, lessons or reading, psalms, responsories, versicles), but their definition is not necessary to comprehend the present article. Suffice to say that the structure of the overall cursus is very hierarchical and granular, with fixed arrangements for the sections, subsections and pieces, so that the position of each text category is almost always identical and texts are not distributed randomly.[8]

8 Daille et al., 'Transcription automatique et segmentation thématique de livres d'heures manuscrits', 18.

Before the Council of Trent in the sixteenth century, the liturgy of the Church was very diverse and each diocese, congregation, and major church used a different cursus for the same office. The structure of the office is the same across different uses, while the set of prayers and chants differs. An office is said to be "for the (liturgical) use of" an institution if its text concords with the textual specificities of the cursus of the institution. Almost all dioceses and major abbeys chose their texts from a rather limited corpus, but did not use them at the same moment during the day. The selection and order are typical for specific liturgical uses, as was first recognized at the start of the twentieth century by G. Beyssac, then by F. Madan and V. Leroquais.[9] The aforementioned scholars used the combinatory diversity and relative stability of known sets to identify their multiple liturgical uses.

Crucial to the present study is that a manuscript containing an office for the use of some church may have been produced in another context, and both the place of origin and the first place of use may differ from the liturgical use. For example, books of hours were produced in Flanders, for the use of Rome and a commissioner in Paris.

In this context, many scholars and cataloguers overuse the word use for calendars and other texts, or for the entire volume. Calendars in books of hours may indeed integrate some liturgical components, such as the hierarchy of feasts rendered in gold, blue, red or other colors. However, they are not liturgical calendars *stricto sensu* and we cannot label them in such a precise way as for liturgical pieces, as is demonstrated by the CoKL database.[10] Likewise, some descriptions infer a theoretical use for the complete volume corresponding to the place of origin or the place of first destination. For example, MS. Paris, BnF, Bibl. Arsenal, Ms-569 was called "Heures à l'usage de Gand" by H. Martin at the end of the nineteenth century and this title is still in the online catalogue, despite some revisions.[11] In reality, both the Hours of the Virgin and the Office of the Dead in this manuscript are for the use of Rome.

Books of hours contain mainly so-called votive offices, whose contents are fixed and do not change along with the calendar, as the collective celebration within the church does. They are therefore adequate for lay people with less knowledge about liturgy. The Hours of the Virgin and Office of the Dead are the main votive

9 Madan, 'Hours of the Virgin Mary (Tests for Localization)'; Madan, 'The Localization of Manuscripts'; Leroquais, *Les sacramentaires et les missels manuscrits des bibliothèques publiques de France;* Leroquais, *Les Livres d'heures manuscrits de la Bibliothèque nationale;* Beyssac, 'Moyen Court'.

10 Macks, 'CoKL: Corpus Kalendarium'.

11 Bibliothèque nationale de France and Torres, 'Ms-569. Livres d'heures, en latin, à l'usage de Gand'.

offices at the core of book of hours. The practice of adding the private recitation of these offices to the participation in the main collective liturgy emerged and developed in the high Middle Ages.[12] In the eleventh century, Peter Damian was an important advocate for the daily recitation of the Hours of the Virgin, and while the Carthusian order included it from the start, some orders resisted (e.g., the Cistercian order until 1194).[13] The evolution of both offices was not always parallel: Cluny adopted the liturgy for the Dead from an early date, but due to its already high burden, resisted the introduction of the Marian office.[14] Thankfully for our analysis, even if they are fixed and shorter, votive offices retain the diversity of the collective liturgy, so that their liturgical use can be distinguished and ascertained. Section 3 will show how seriality can be introduced to explore textual networks.

2.2 Books of hours as a compilatory genre

Books of hours constitute a distinct category in literary history. Each of the manuscripts falling in this category contains several parts and diverse offices. These constitute a genre rather than a work *per se,* and each of them is a compilation, produced with similar choices, but creating a different result. Indeed, there is no consensus on which combination of components defines a book of hours. For instance, V. Leroquais distinguished between "essential", "secondary", and "accessory texts".[15] The essential elements are, according to Leroquais, those that were drawn from the breviary, i.e., the calendar, the Hours of the Virgin, the penitential psalms and litanies, the suffrages, and the Office of the Dead. Secondary are the four Gospel lessons, two prayers named *Obsecro Te* and *O intemerata,* the Hours of the Cross, and the Hours of the Holy Spirit. Until this point, the list coincides with the eight essential components listed by R. Wieck (the Hours of the Cross and of the Holy Spirit are listed as one item, as are *Obsecro Te* and *O intemerata*),[16] but V. Leroquais also records the Passion according to John, the *Joys of the Virgin* (*Quinze joies de la Vierge*), and the *Seven requests to our Lord* (*Sept requêtes à notre Seigneur*) among the secondary texts, while they are considered accessories by R. Wieck. Other texts may then be added *ad libitum.* More recently, while underlining that there are "numerous hybrid forms", J. Hamburger lists only "Calendar, Little Office of the Virgin Mary, Shorter Hours [Cross and Holy Spirit], Office of the Dead" among the "standard sections", thus omitting even the penitential psalms and litanies.[17]

12 Batiffol, *Histoire du bréviaire romain,* 184–87.
13 Laporte, *Aux sources de la vie cartusienne. Dl. 4/2,* 529–33.
14 Ibid., 529–33.
15 Leroquais, *Les Livres d'heures manuscrits de la Bibliothèque nationale,* vol. 1, p. xiv.
16 Wieck, *Time sanctified,* 27–28.
17 Hamburger, 'Another perspective: the book of hours in Germany', 97.

At the level of essential texts, a peculiar consequence of the compiling and customizing process is that one manuscript book of hours may contain offices for different liturgical uses. For instance, MS. Amsterdam, Universiteitsbibliotheek, XXV C 26 was produced in Ghent and contains the Hours of the Virgin for the use of Rome and an Office of the Dead for the use of Tournai.[18] This phenomenon is not rare, but will not be further analyzed in this article.

The discrepancies in the expected contents highlight the compilatory nature of books of hours. As they contain offices, which are themselves already codified compilations of many texts of different natures, they form a second-grade compilation that is also subject to rules and local variations. Two manuscripts may have exactly the same texts, but in a different order: e.g., MS. Auxerre, Trésor de la cathédrale, 14, for the use of Troyes, and MS. Paris, BnF, Bibl. Arsenal, Ms-637, copied in Normandy for the use of Rome. Both contain a calendar, Gospel lessons, the Hours of the Cross, of the Holy Spirit, and of the Virgin, the penitential psalms and litanies, the Office of the Dead, and the prayers *Obsecro Te* and *O intemerata;* however, while this is the order in the former, the latter places both prayers before the Hours of the Virgin, and the Hours of the Cross and of the Holy Spirit after the Hours of the Virgin.[19]

D. Porter has highlighted how much the order of sections may correlate with the region of production,[20] corresponding to some remarks by G. Baroffio[21] or, on a smaller geographical scale and for vernacular Dutch manuscripts, by M. Hülsmann.[22] For instance, observing only books of hours for the use of Rome, D. Porter could evidence that the Hours of the Cross and the Holy Spirit come before the Hours of the Virgin in Flemish books, as the Gospel lessons also do at times, contrary to the rest of European book production.[23] The "modular method", used to build books of hours and studied by K. Rudy,[24] allows for "changing the structural order of the quires",[25] but such changes would require a new binding and only a detailed analysis could help to trace the re-ordering in some manuscripts.

In recent literature, networks of manuscripts containing the same texts have been used to support genre analysis and text classification, and to understand

18 Clark, 'AmsterdamUBxxvC26 (228)'.
19 Hazem et al., 'Hierarchical Text Segmentation for Medieval Manuscripts', 6243.
20 Porter, 'Books of Hours as Transformative Works'.
21 Baroffio, 'Testo e musica nei libri d'ore', 21 n. 16.
22 Hülsmann, 'Variation in Page Layout'.
23 Email from 2 to 6 June 2021.
24 Rudy, *Piety in Pieces,* 15–57.
25 Ibid., 12.

the reshuffling of short texts and poems.[26] Authors produced bi-modal networks containing two sorts of entities, namely manuscripts and texts, but suppress the notions of seriality and of granularity. For books of hours, such an approach without seriality could help to highlight when and where certain texts were more popular, but it is not adequate to analyze macro-structures such as those uncovered by D. Porter.

In the following pages, section 3 will introduce the techniques of social network analysis for the textual contents of the offices. We will address the notion of seriality in the textual network analysis, starting with the dataset published by Knud Ottosen to re-assess the validity of his interpretation, and then continuing with a study of the Hours of the Virgin.

3. Offices as compilations and seriality in textual networks

3.1 Office of the Dead

In the early 1990s, K. Ottosen published an extensive study on the Office of the Dead.[27] K. Ottosen's study is fundamental, both because of the breadth of his corpus and the quality of his liturgical analysis, as he was able to investigate the liturgical corpus in a philological manner as well as proving relations and dependencies between liturgies. First, we will summarize K. Ottosen's contribution to the knowledge on how the Office of the Dead developed, and how responsories are apt sources for a liturgical analysis. We will then use his dataset to introduce the methodology of network analysis on liturgical offices and propose a more fluid approach.

3.1.1 *K. Ottosen's contribution*

Based on the previous discovery that the liturgical diversity could be captured through the "responsories", he focused on these pieces, which are chanted in the three nocturns of Matins as a response to the lessons (or readings), and which function as an interpretation and exegesis of the biblical readings, whose meaning they may even contradict.[28]

---

26 Julien, 'Délier, lire et relier'; van der Heijden, 'Or Ai Ge Trop Dormi'; Fernández Riva, 'Network Analysis of Medieval Manuscript Transmission'.

27 Ottosen, *The Responsories and Versicles of the Latin Office of the Dead*.

28 Ibid., 249. K. Ottosen also distinguishes two main families of readings, and twelve additional sets of readings known through a small number of witnesses. His data sheets also contain the versicles, and he suggests that studying them, especially the versicles to the *"Libera me"*, would be fruitful. In the following, we focus on his dataset of responsories.

Accompanying his study and including the results of G. Beyssac and M. Huglo, he published a comprehensive census of responsories attested in 2047 sources, and later published the full dataset online.[29] A masterpiece of open data scholarship, this highly valuable database can be subjected to new enquiries; this subsection is thus entirely based on his data, only providing new observations in section 3.1.2. Following his example, we will name the positions of the responsories in the liturgy "R1" to "R12", and record the texts that are chanted with the code given by K. Ottosen as ranging from "1" to "99". Both in secular and monastic institutions, the Matins during which the responsories are chanted are divided into three nocturns, but in secular institutions each nocturn has only three lessons and their three responsories, while monastic institutions usually have four lessons and four responsories in each nocturn.[30] To compare both kinds of cursus, K. Ottosen named the three-by-three responsories of secular houses R1–R3, R5–R7, R9–R11 (and sometimes used the positions R4, R8, R12 to record additional responsories at the end of the third lesson of each nocturn). As we focus on secular institutions, we will generally record only nine responsories, and in our tables we will generally omit the columns "R4", "R8" and "R12", which have no information.

K. Ottosen demonstrated how minimal differences may distinguish two liturgical uses either in different institutions, or before and after a liturgical reform, as in Subiaco, Jerusalem, Bayeux, and Zwiefalten. Table 1 shows how an ensemble of eleven responsories (12, 14, 24, 28, 32, 38, 40, 57, 68, 72, 82) suffice to build five similar, yet distinct series.

Location	Time	R1	R2	R3	R5	R6	R7	R9	R10	R11
Subiaco	970–1000	14	72	24	32	28	57	68	40	38
Jerusalem	1131	”	”	”	”	57	28	”	”	”
Zwiefalten	1100–1200	”	”	”	”	”	”	”	”	82
Bayeux	1200–1300	”	”	”	”	”	12	”	”	38
Zwiefalten	1380–1420	”	”	”	”	”	68	82	28	”

Tab. 1 Similar responsories in the Office of the Dead in several institutions (Subiaco, Bayeux, Jerusalem) or successive states of a liturgy for one institution (Zwiefalten).

29 Ottosen, 'Responsories of the Latin Office of the Dead'. The dataset we used was that published at: https://www-app.uni-regensburg.de/Fakultaeten/PKGG/Musikwissenschaft/Cantus/Ottosen/search.html. In the meantime, the site has moved to https://www.cantusplanus.de/databases/Ottosen/Ottosenseries.php.

30 Cluny, Moissac, Mont-Saint-Michel are among the exceptions and will be mentioned again later.

Location	Time	R1	R2	R3	R5	R6	R7	R9	R10	R11
Rome	1350–1500	14	72	24	46	32	57	68	28	38
Paris	1100–1200	72	14	32	57	24	68	28	46	38

Tab. 2 Similar responsories in the Office of the Dead ordered in different sequences (Rome, Paris).

It is not only Subiaco and Jerusalem that have the same ensemble of responsories, and the order is key to distinguishing between them. A prominent example in the medieval world are the uses of Paris and Rome, which have all nine responsories in common. Only the order is different, and, except for the last, none is chanted at the same place in the cursus (cf. Table 2).

In his study, K. Ottosen identified several sets of responsories which are characteristic of specific regions or liturgical families. His repertory was arranged by increasing the number of responsories and the series from R1 to R12. Therefore he could easily spot and study groups that were identical or shared a common beginning. The hierarchy and series of responsories are presented in their numerical order, thus largely according to the arbitrary alphabetical order. Table 3 provides an overview of what K. Ottosen called "types" and "sections", as well as their historical, geographic, or liturgical nature.

Some institutions shared the same liturgical use. Often, local churches use the same sets as the neighboring cathedral or a nearby monastery (Saint-Arnulf in Metz, Saint-Lo in Rouen and Saint-Lo in Coutances, and probably Saint-Thierry and Saint-Remi in Reims, with a different set than the cathedral and Saint-Denis). Most groups of locations attached to a common set make sense from a historical or geographical perspective and correspond to evident connections. To name a few, already discussed by K. Ottosen: the series 14, 72, 24, 32, 57, 40, 68, 82, 38 shows a link between Avranches and Rouen, and probably corresponds to the actions of bishop John of Avranches in the eleventh century;[31] while the series 44, 47, 58, 76, 83, 79, 1, 18, 38, is attached to a geographically coherent region including Magdeburg, Hildesheim, Brandenburg, Werden, Odense, Havelberg, Hamburg, Halle, and Halberstadt. Some connections may not be self-evident at first sight. For example, the series 14, 72, 24, 32, 57, 28, 68, 46, 38 was used by the Dominican order, Sainte-Croix (Paris), Teutonic order, Uppsala, Skara, Västeras, Abo, Mercedarians, and Dominican Vienna. The institutions in this list appear geographically scattered. However, the convent of Sainte-Croix in Paris is an establishment of the Canons Regular of the Holy Cross (or Crosiers), which followed the same liturgy as the Dominican order, as did the Teutonic order and the

31 Ottosen, *The Responsories and Versicles of the Latin Office of the Dead,* 249–50.

Types	Sections	Locations
2-6		Ambrosian
14-32		Cambrai
14-36		Lyon: 14, 36, 46, 51, 75, 95 ... • Carthusian: 14, 36, 46, 67, 51, 33 ... • Grandmontine: 14, 36, 67, 46, 51, 75 ... • Saint-Rufus and Coimbra: 14, 36, 72, 32 ...
14-68-72		Arles
14-72-21		Agde, Lerida, Marseilles
14-72-24	32-28-57	Strasbourg
14-72-24	32-57-12	Bayeux
14-72-24	32-57-28	Various variants, including Dominican, "Metz-Normandy-Sarum"
14-72-24	32-57-36	England
14-72-24	32-57-37	Nantes
14-72-24	32-57-38	Autun, Fleury
14-72-24	32-57-40	Avranches, Rouen, Dol
14-72-24	32-57-46	Various unrelated variants
14-72-24	32-57-51	Noyon, Tournai
14-72-24	32-57-56/62	Minor sections: Italy
14-72-24	32-68-57	Romano-Germanic pontifical, model of the following
14-72-24	32-57-68	Bamberg and derivations
14-72-24	46-32-57	Roman use, Franciscan
14-72-24	47-1	Marbach, Sion
14-72-24	56-46-58	Orléans
14-72-24	57-32-40	Premonstratensian
14-72-24	68-57-82	Utrecht, Windesheim
14-72-24	82-32-57	Reims, Laon
14-72-24	90-32-57	Cluny and derivations
14-72-24	90-32-68	Aniane, Bobbio, Monza
14-72-28		Mâcon
14-72-32	24-57-68	Chartres
14-72-32	57-24-28	Prüm
14-72-32	68-24*-46	Sens and derivations
14-72-40		Loches
14-72-46		Cistercian
14-72-47		Littlemore (Oxford)
14-72-51		Saint-Oyen (Jura)
14-72-56		Italy, Vallumbrosan order

Types	Sections	Locations
14-72-58		Flanders
14-72-68		Évreux, Albi, Aurillac
14-72-79		Cologne
14-72-82	1-28-93	Troyes
14-72-82	24-32-1	Vich
14-72-82	24-32-57	Nîmes
14-72-82	32	Toulouse, Schleswig
14-72-83		Münster
14-72-90		Old Roman, Italy
14-72-138		Saint-Ghislain (Hainaut)
14-82-72		Grenoble
72-14	24	Péronne
72-14	32	Paris, Meaux, Saint-Quentin, Saint-Pol-de-Léon
72-14	38	Arras
72-14	40	Auxerre
72-14	56	Amiens
72-14	68	Valence
72-82		Châlons-en-Champagne
82-72		Vienne (France)
82-83 and 83-82		Lund
83-25		Saint-Vaast (Arras), Bruges, Lille, Scandinavian dioceses
25		Major offices: Flanders, Utrecht, Windesheim
36		Major offices: Saint-Vanne (Verdun)
44		Major offices: North German dioceses, Odense, Constance
58		Major offices: St. Lambrecht (Austria)
70		Major offices: Southern Germany, Austria
79	1	Major offices: Trier, Murbach, Lobbes
79	10 and 27	Major offices: Nonnberg, Salzburg, Mondsee, Aquileia
79	44 and 58	Major offices: Trier, St. Emmeram, Passau, Erfurt
79	76	Major offices: Metz
79	82 and 83	Major offices: Bamberg and German institutions
85		Major offices: German institutions, Thérouanne
Extraordinary types		

Tab. 3 Overview of K. Ottosen's findings. Series named according to the order of responsories in the series and to the numbering of responsories by K. Ottosen.

Mercedarians. Therefore, it can be seen as a specifically Dominican cluster, and the proof of the Dominican influence on some Scandinavian dioceses.[32] Similarly, a large and (at first glance) incoherent group with the series 14, 72, 24, 90, 32, 57, 68, 28, 46 extends from Winchcombe and Evesham in England to Marseille and Montmajour in Provence. K. Ottosen points out the influence of Cluny on all the institutions concerned.[33]

Despite some diversity, the Low Countries are largely a coherent liturgical region. K. Ottosen identified some series of responsories that were in use in this area: three minor offices and two major ones. These are, on the one hand, the minor office starting with "14, 72, 58" and found mostly in psalter-hours linked to the region; the Premonstratensian office (14, 72, 24, 57, 32, 40); and the minor office of Utrecht and Windesheim (Utrecht: 14, 72, 24, 68, 57, 82, 93, 58, 29; Windesheim: 14, 72, 24, 68, 57, 82, 32, 58, 38); and, on the other hand, the major office "83-25" and "25" used in the Low Countries and Upper Rhine region.[34] The latter group is subdivided into "25-44" and "25-72". Table 4 shows the series of major offices with the three groups. These display a total of 8 different series with only 17 different responsories, in particular 9 that are reshuffled and are characteristic of the region.

The type "25", with responsory 25 in the first position, links the uses of Antwerp, Utrecht, Liège, and Brussels to the uses of Verdun and Cologne,[35] and type "25, 44, 47" corresponds to Utrecht, Windesheim and Corssendonck.[36] Type 83-25, which is not very large geographically, does not extend beyond the borders of Flanders, except for an excursion to Scandinavia.[37] It comprises not only Saint-Omer, Lille, Phalempin, Tournai, Saint-Donatian in Bruges, Henin-Lietard, Watten, Notre-Dame-du-Bon-Conseil, Mont-Saint-Eloi, but also Nidaros and Skalholt.

K. Ottosen's study has demonstrated the complexity of the liturgical history and opened a clear path to understanding liturgical evolutions, coherences and connections between distant places. However, his presentation, as well as some of the conclusions based on the order of the responsories, should now be revised.

---

32 Ibid., 239–42.

33 Moissac, Winchcombe, Jumieges, Mont-Saint-Michel, Saint-Benigne in Dijon, Saint-Thierry [Reims], Saint-Amand, Saint-Vaast, Bourgueil, Troarn, Saint-Sepulchre in Cambrai, Saint-Bertin [Saint-Omer], Saint-Ghislain, Saint-Victor in Marseille, Montmajour, San Benito el Real, Corbie, Evesham, Broholm, Vendome, Saint-Ouen, Saint-Germain-des-Prés [Paris], Saint-Maur-des-Fossés, Cluny, Saint-Remi [Reims], Villeneuve-les-Avignon, and others which are unidentified.

34 Ottosen, *The Responsories and Versicles of the Latin Office of the Dead,* 336–43.

35 Ibid., 343–346.

36 Ibid., 345.

37 Ibid., 336–40.

Location	Time	R1	R2	R3	R5	R6	R7	R9	R10	R11
Utrecht, Windesheim, Corssendonck	1300–1400	25	44	47	13	93	83	40	79	18
Verdun	1300–1400	”	72	18	”	58	”	93	82	38
Brussels	1350–1550	”	”	38	”	40	”	”	58	18
Liege	1100–1200	”	”	”	”	93	”	82	”	”
Antwerp	1496	”	”	”	32	57	13	68	”	”
Saint-Donatian of Bruges	1520	83	25	13	58	29	93	72	82	38
Lille	1200–1300	”	”	”	”	”	”	82	72	”
Saint-Omer	1270–1290	”	”	”	”	93	29	72	82	”

Tab. 4 Office of the Dead: uses of the major offices of the Low Countries (responsories numbered according to K. Ottosen).

3.1.2 *Re-reading K. Ottosen's data and studying a textual network*

For all the major input to liturgical research of his study, K. Ottosen did not provide general statistics on his corpus and did not try to interpret the full network of liturgical uses and manuscripts. Types are juxtaposed and analyzed based on their beginning, without any evaluation of their other shared responsories. While many clusters have been explained by K. Ottosen, we can now add statistical remarks on the overall corpus, introduce the (social) network analysis techniques for the textual contents, and come to new conclusions.

The dataset contains 2047 series recorded from primary sources and 557 different "locations" in which the liturgy was in force (including an unidentified blank location for 141 sources). There are 697 distinct series or responsories, and 1067 distinct pairs of one "location" and one series of responsories.[38]

The corpus of responsories is small. From Ottosen's list of seventy-five responsories (numbered in alphabetical order from 1 to 95, with additional numbers from 101 to 138), only twenty are used more than 200 times.[39] Some responsories are used almost universally at a certain position, e.g., responsories 14, 72, and 38

38 For these statistics, we take the absent responsory "0" as a discriminating feature and count all series recorded by K. Ottosen, even if incomplete.

39 When a responsory is not attested due to the specifics of the source or because it is a secular institution (for resp. 4, 8 and 12), he records "0".

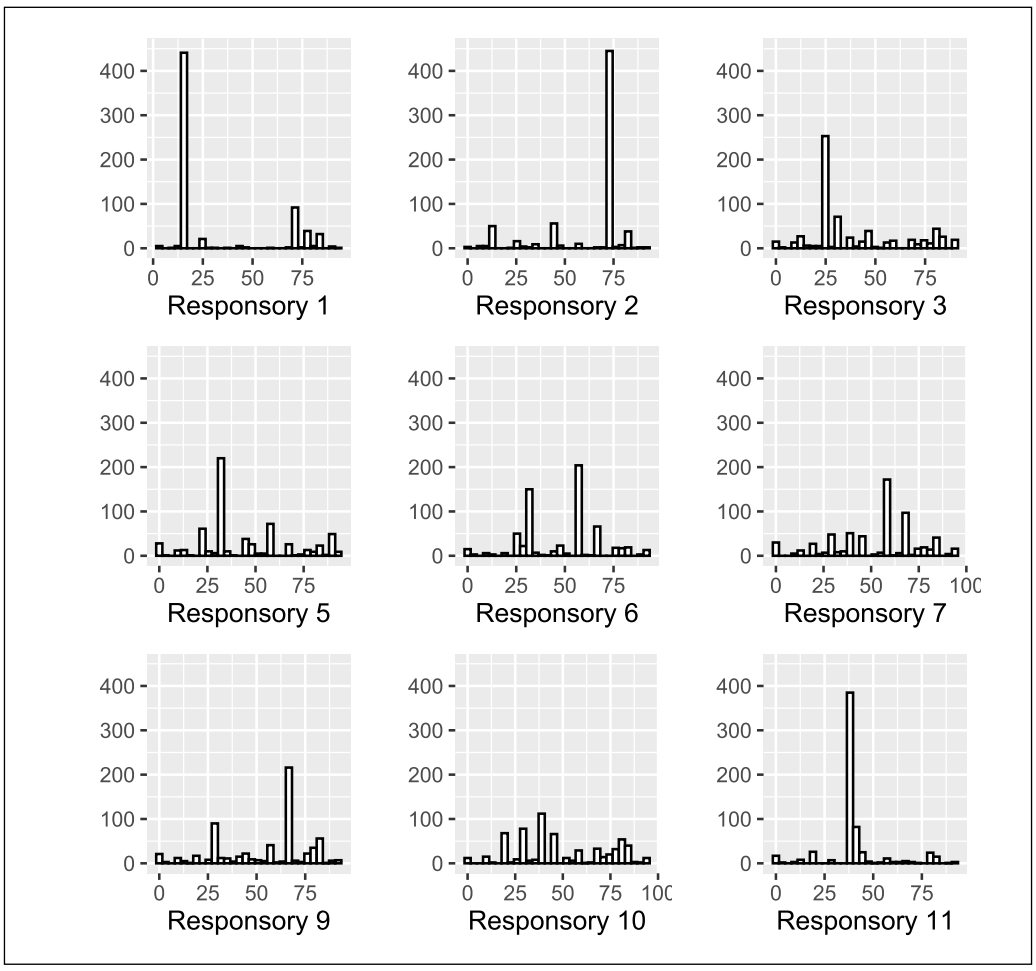

Fig. 1 Number of occurrences of each responsory across the different sets in the nine main positions (responsories numbered from 1 to 99 according to K. Ottosen). Source of the data: K. Ottosen.

in the first, second and penultimate positions, while there are more distributed possibilities in other positions, as illustrated in Figure 1.

As illustrated by Table 5, there are a large number of variant sets for some locations and this number is not strictly correlated with the number of witnesses examined by K. Ottosen. As evidenced by Table 6, manuscripts for the same use are not distributed evenly across the attested sets, but rather concentrate on certain sets of responsories. For the Roman use, almost all manuscripts bear the same text – the only difference is how an additional responsory is indicated at the end. For Paris, the same applies, yet the diversity is increased by the presence of hours with only one nocturn. In Rouen, it is once again the same, although there is a choice between resp. 28 and 40 at the end of the second nocturn (R8). Con-

Location	Witnesses	Sets of responsories
unidentified	141	109
Paris	52	11
Roman	70	9
Zwiefalten	9	8
Rouen	36	7
Tournai	16	7
Amiens	18	6
Bamberg	14	6
Saint-Denis	6	6
Utrecht	27	6
Weingarten	8	6
Ambrosian	5	5
Aquileia	8	5
Besancon	18	5
Braga	6	5
Chartres	13	5
Évreux	12	5
Hildesheim	6	5
Meaux	12	5
Münster	12	5
Passau	8	5
Poitiers	14	5
Sarum	27	5
Senlis	8	5
Soissons	7	5
Saint-Ghislain	5	5
Saint-Wandrille	6	5

Tab. 5 Number of different records and of different sets of responsories for the same "location" (table for more than 5 records and more than 5 sets).

Location	Series of responsories (R1 to R12)												Witnesses
Paris	72	14	32	na	57	24	68	na	28	46	38	na	30
	72	14	32	na	57	24	68	na	28	46	40	38	9
	72	14	38	na	na	na	na	na	na	na	na	na	5
Rome	14	72	24	na	46	32	57	na	68	28	40	38	34
	14	72	24	na	46	32	57	na	68	28	38	na	17
	14	72	24	na	46	32	57	na	68	28	38	40	6
	14	72	24	na	46	32	57	na	68	28	40	na	6
Rouen	14	72	24	na	32	57	40	na	68	82	38	na	17
	14	72	24	na	32	57	28	na	68	82	38	na	9
	14	72	24	na	32	57	28	40	68	82	38	na	3
	14	72	38	na	na	na	na	na	na	na	na	na	3

Tab. 6 Most frequent sets of responsories and number of witnesses for the use of Paris, Rome and Rouen (responsories numbered according to K. Ottosen, NA used for not applicable, rather than "0" used by K. Ottosen).

versely, the sets of responsories may be linked by K. Ottosen to up to 27 different locations, as illustrated in Table 7.

Beyond the clusters interpreted by K. Ottosen, some series currently resist interpretation, such as the set 14, 72, 24, 32, 57, 28, 68, 82, 38, which groups together the following locations in K. Ottosen's data: Nidaros, Carmelite [order], Hereford, Lincoln, Metz, Sarum, Templars, Reenes [sic for Rennes?], Oxford, Rennes, Rouen, Saint-Arnulf [Metz], Lisieux, Coutances, The Hague, Hospitallers, Hours of Marie de Bohun, Montieramey, Seez, Saint-Lo-de-Rouen, Saint-Jacut-de-la-Mer, Saint-Sauveur Toul, Saint-Lo-de-Coutances, Dol, Aversa, Saint-Maur Verdun. For such groupings, an interpretation is beyond the scope of the present article.

We now propose to use the same data to see if other connections can be highlighted. In order to study the Office of the Dead as a network, the modelling is not straightforward. K. Ottosen's census does not essentialize liturgical uses, and provides the manuscript evidence with a liturgical location. He also records additional, or "extra", responsories which may be used at the end of each nocturn to integrate some of the intrinsic variability of liturgical uses. For example, the last responsories for Rome may be not only 38 *"Libera me, Domine, de morte"*, but also 40 *"Libera me, Domine, de viis inferni"*; alternatively, the manuscripts may indicate both responsories, so that the extant witnesses provide four different sets, all

R1	R2	R3	R5	R6	R7	R9	R10	R11	R12	Different locations
14	72	24	32	57	28	68	82	38	na	27
14	72	24	90	32	57	68	28	46	na	27
14	72	24	46	32	57	68	28	40	38	21
14	72	24	90	32	57	68	28	46	38	17
14	72	24	32	57	68	28	40	38	na	15
14	72	38	na	na	na	na	na	na	na	15
14	72	24	90	32	57	68	28	38	na	12
79	1	18	47	58	83	10	76	38	na	11
14	72	24	32	57	28	68	46	38	na	10
14	72	24	46	32	57	68	28	40	na	10
14	72	82	32	57	40	68	46	38	na	10
44	47	58	76	83	79	1	18	38	na	10
14	72	24	32	57	28	68	82	40	na	9
72	14	32	57	24	68	28	46	38	na	9
14	72	24	32	57	28	68	40	38	na	8
14	72	24	32	57	28	68	82	38	40	8
14	72	24	32	57	68	79	40	38	na	8
14	36	72	32	68	24	51	56	38	na	7
14	72	24	32	57	68	82	83	38	na	7
14	72	24	90	32	57	68	28	40	na	6
14	72	24	na	na	na	na	na	na	na	6
14	72	82	32	57	38	68	28	40	na	6
72	14	38	na	na	na	na	na	na	na	6
14	72	24	46	32	57	68	28	38	40	5
14	72	46	32	57	40	68	28	38	na	5
25	44	47	13	93	83	40	79	18	na	5
83	25	13	58	29	93	82	72	38	na	5

Tab. 7 Sets of responsories and number of identified locations: table for more than 4 locations. The responsories are numbered according to K. Ottosen (NA is used for not applicable, rather than "0" used by K. Ottosen); responsories in R12 are variants of R11, as present in the source data from K. Ottosen. Statistics by the authors.

starting with 14, 72, 24, 46, 32, 57, 68, 28, and ending with 38, or 38+40, or 40, or 40+38 as illustrated in Table 6. For the following, we keep only the information in R1–R3, R5–R7, R9–R11, and only complete sets.

Building a network of responsories without taking into account the position would not render the liturgical process. The required formalization is not adequate to render the variability of "extra" responsories.

In a first approach, we can build a comprehensive network of manuscripts and responsories defined as follows:

- bimodal network with entities of type [Manuscript] and [Responsory]
- [Responsory] entities combine seriality and text, with the number given by K. Ottosen or with the incipit (e.g., "R1_14" or "R1_Credo quod"); [Responsory] entities corresponding to a missing information (e.g., "R4_0") should then be discarded
- liturgical locations are an additional piece of information

A second step is to project this multimodal network onto a unimodal one and establish links directly between manuscripts and weight edges according to the number of common readings. This can be done either through a dedicated library, e.g., MultimodeNetworksTransformationPlugin in Gephi, or by creating a specific edge table. For the present article, we use the R libraries "network" and "ggnetwork".

The network is very dense and concentrates around the small number of popular responsories. This can be amended in the second approach with a cut-off, for example by limiting the creation of edges for manuscripts which share at least four responsories.

Figure 2 provides a visualization with the Fruchterman-Reingold force directed algorithm, allowing us to evaluate how close the manuscripts and the uses they instantiate are from one another. Many clusters represent liturgical affinities, such as the one which groups Paris, Beauvais, Meaux, Thérouanne, and Saint-Pol-de-Léon, or the one grouping dioceses in southern Germany. Other cliques represent only one isolated, albeit popular use (e.g., Cambrai, Bourges).

Even with the correction of the cutting point, the network is biased by the popularity of some uses. Links from many manuscripts for one use (e.g., Rome, Cambrai, Paris) to all the others for the same use abet the creation of cliques and dense clusters. Although this may well represent how medieval people perceived liturgical uses, popular responsories, and their order, it precludes the correct perception of liturgical affinities. Therefore, we decided to remove the duplicates in K. Ottosen's table in order to keep a single instance of a location and a corresponding set of responsories.

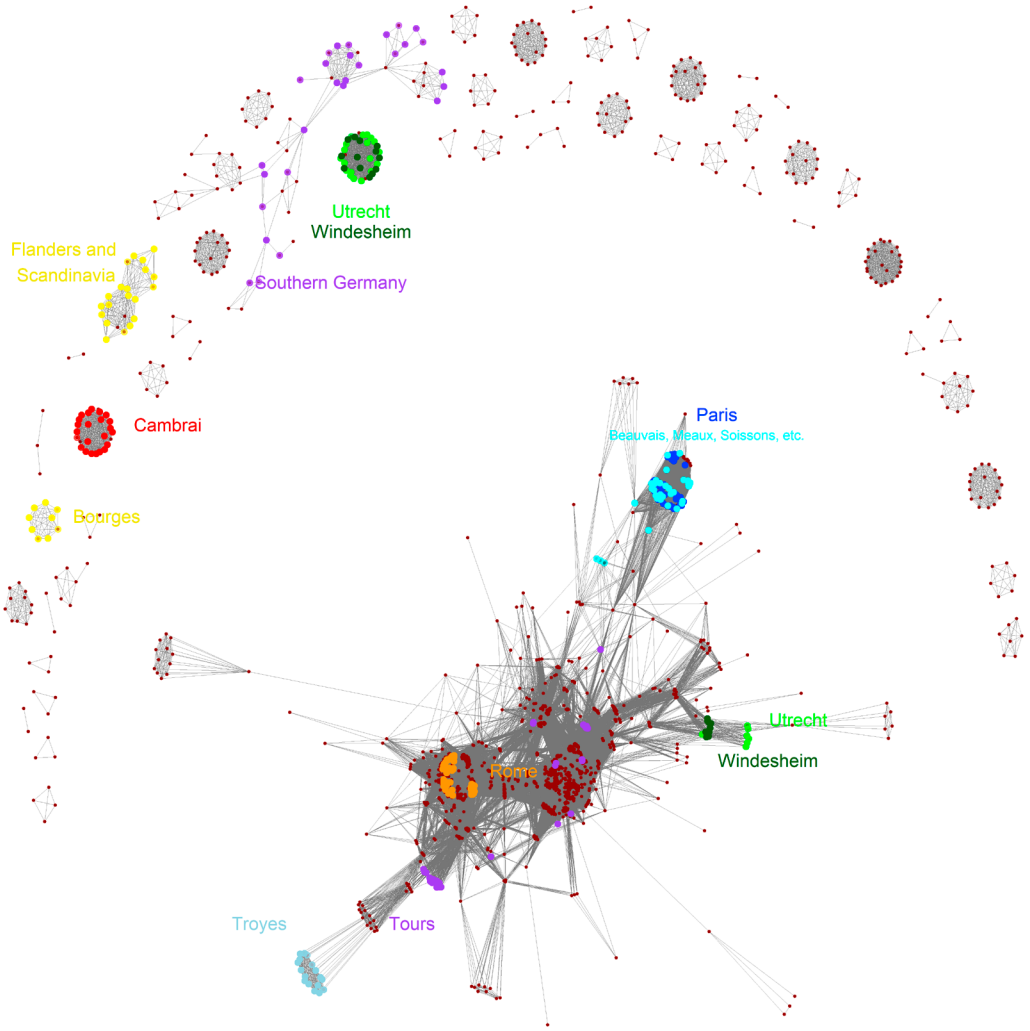

Fig. 2 Office of the Dead: network of manuscripts according to their liturgical uses (edges defined by more than five common responsories). Source of the data: K. Ottosen. Parameters and network graph by the authors, using R libraries "network" and "ggnetwork", applying the Fruchterman-Reingold algorithm.

Going back to a network approach to enquire if proximities may be spotted on a purely liturgical basis (i.e., without the help of the most copied and popular uses) proves to be useful. It does not change the overall perspective, but highlights links between different locations and helps to spot links for rarer uses, which would otherwise remain unnoticed.

In this instance, network analysis can shed light and give a broader perspective on a corpus that had already been the subject of a detailed analysis. It allows

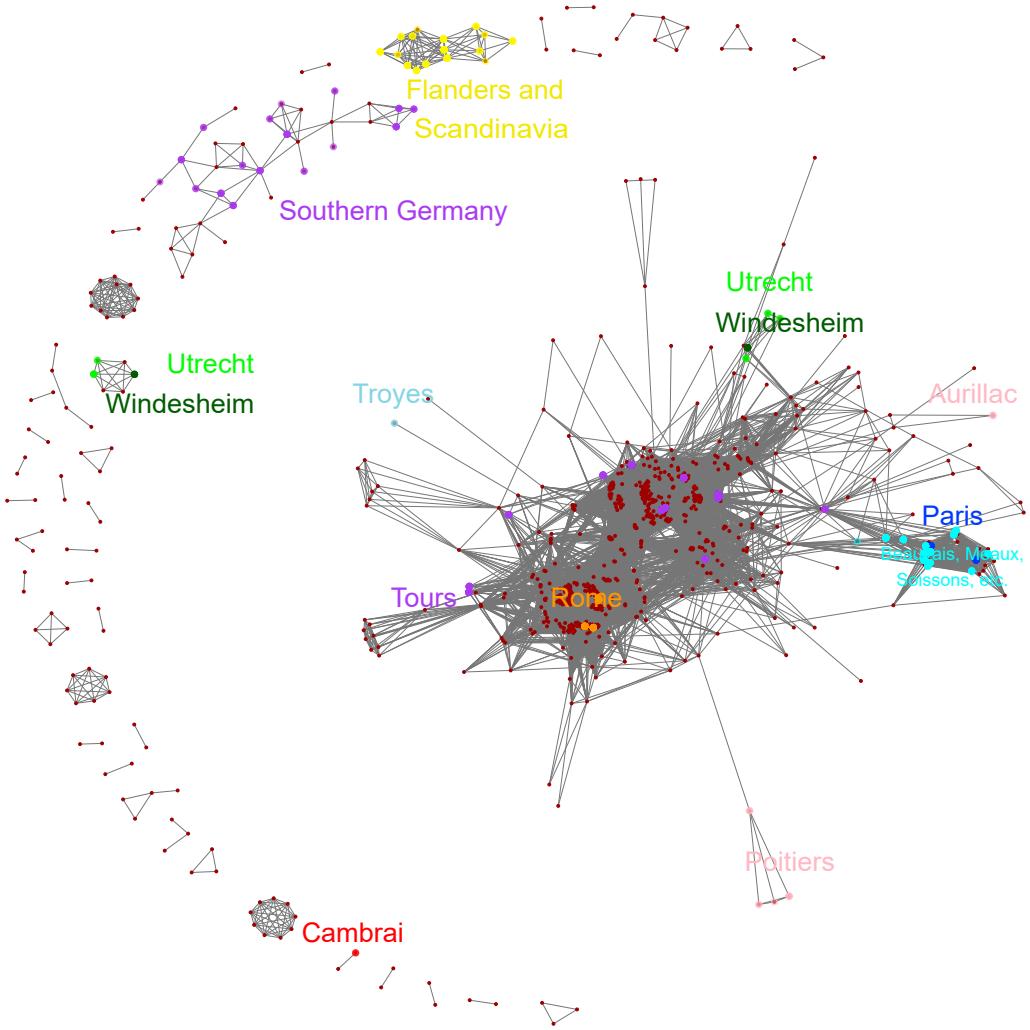

Fig. 3 Office of the Dead: network of liturgical uses (edges defined by more than five common responsories). Source of the data: K. Ottosen. Parameters and network graph by the authors, using R libraries "network" and "ggnetwork", applying the Fruchterman-Reingold algorithm.

for a clearer view of links and proximities of uses on a large scale, but also in the detail of local clusters. This is not the place to comment and analyze in detail certain unexpected connections in the liturgy of the Office of the Dead, which K. Ottosen knew of, such as the fact that Caen, Troarn and Saint-Wandrille in Normandy are in the cluster of southern German dioceses and institutions, and similarly that Saint-Pol-de-Léon and Quimper belong to the Parisian cluster.

With regard to the Low Countries, we may formulate two preliminary remarks. First, this network provides an excellent view of the singularity of major offices.

These stand out by themselves, while the minor offices are closer to the large continent (i.e., biggest connected component) of uses that are more or less related to Rome. In this continent, the minor offices are closer to dioceses in German territories (Würzburg, Basel, Weingarten), but also, less expectedly, to Angers and Saint-Jean-en-Vallée near Chartres.

The series of the major offices do not build a common group. The Flemish-Scandinavian cluster of the 83-25 group is not connected to others in our Figure 3, which demonstrates that no other use shares more than five responsories in the same position. It is therefore not only the choice of the first two responsories that is idiosyncratic, but the whole arrangement of the nine responsories. These connections between Flanders and Scandinavia are obviously not a random phenomenon. They are currently being studied by S. Myking.[40]

On a more profound level, the criterion of the first responsories should be discussed. On a first sight, it proves to be efficient. The liturgical uses and their connections as studied by K. Ottosen do not appear randomly, and the role and place of the first two responsories is probably a very profound liturgical mark. Nevertheless, the coincidences between the first responsories turn out to be greatly overestimated by the Danish scholar. Our new approach allows us to distinguish the kind of specificities and to measure the closeness of liturgies.

For example, Amiens (72, 14, 56, 24, 32, 57, 68, 40, 1/38) has been associated with some French uses, based on the first two responsories, whereas we can observe that the set of responsories in Amiens is very close to the series of the Vallumbrosian order (*14, 72,* **56**, **24**, **32**, **57**, **68**, 28, 46). There is only the inversion of the first two responsories and a change in the last two responsories.

Cambrai and Bourges are both isolated in the network, whereas K. Ottosen treats them in a very different manner: Cambrai as a type, because it starts with 14-32, while Bourges (14, 72, 82, 36, 46, 68, 57, 32, 38) is only mentioned in passing, as a "remote derivation from the Nîmes series" (14, 72, 82, 24, 32, 57, 68, 56, 40).[41] Bourges has fewer than five responsories in common with any of the other uses, but , nevertheless, was not analyzed by Ottosen as a separate type. However, Bourges was the see of an archbishopric and the responsory 36 in position R5 is very rare, and common to the Ambrosian rite and a witness of an Old Roman rite – the latter being more flexible than most.[42] The parallel between Bourges and Nîmes based on responsory 83 as R3, as happens in a large geographical zone, does not seem to be proof that Bourges took its liturgy from Nîmes. On the con-

40 Myking, 'Ter Doest, Lund, and the *Legendarium Flandrense*'; Myking, 'Les livres français dans la Norvège médiévale'.
41 Ottosen, *The Responsories and Versicles of the Latin Office of the Dead,* 321.
42 Ibid., 322–28.

trary, the prominent and vast see of Bourges, whose province in the high Middle Ages covered the dioceses of Albi, Cahors, Clermont-Ferrand, Le Puy, Limoges, Mende, and Rodez, may well be a partial witness to early or independent liturgical habits.

Yet again, the liturgy of Troyes does not truly stand out in K. Ottosen's presentation, as it starts with three common responsories (14, 72, 82, 1, 28, 93, 53, 68, 38). The second nocturn and the first responsory in the third nocturn are highly idiosyncratic, close to the common liturgical mainstream, but still isolated. One of the closest sets is a late series from Passau (14, 72, 82, 24, 28, 32, 68, 57, 38). This outlier position in the network should be the basis for new enquiries.

Even more peculiar is the presentation of Saint-Ruf and the congregation of Coimbra (14, 36, 72, 32, 68, 24, 51, 56, 38), almost in passing, because in Saint-Ruf "the two initial responsories indicate Lyonese influence".[43] In Figure 3, these seven locations build a separate clique. The series of Lyon (14, 36, 46, 51, 75, 95, 19, 94, 69) only coincidentally has the two initial responsories in common with Saint-Ruf, but there is not a single other shared responsory at the same position, and only one (responsory 51) shared in the ensemble. On the contrary, Narbonne (**14**, *72,* 82, **32**, **68**, 21, **51**, 79, **38**) or Arles (**14**, *68,* **72**, **32**, 79, 90, 40, 83, **38**), in which archdiocese Saint-Ruf is located, seem much more likely connections thanks to five (Narbonne) or four (Arles) responsories in common at the same positions, and one additional in the ensemble.

Another example shows how misleading it is to present according to the order of responsories and their incipits. The diocese of Auxerre (72, 14, 40, 32, 8, 12, 68, 36, 38) is analyzed in a separate section, but de facto inserted between Arras, Amiens, and Châlons-en-Champagne. Auxerre is isolated on our graph. Beyond the threshold of five identical responsories, one of the most similar sets is that of Bayeux (14, 72, 24, 32, 57, 12, 68, 40, 38), with the inversion of the first two responsories, and differences in R3, R6, R10. K. Ottosen described the use of Auxerre and stressed how late the first witnesses are (the end of the fourteenth century).[44] We have no further indication, and we should not presume relations or liturgical reforms too lightly. Nevertheless, there are clues that point towards some connections between Auxerre and Bayeux. Pierre de Villaines, bishop of Auxerre for a short period of time (1344–1347), was born in Bayeux and applied to be transferred from Auxerre to the see of Bayeux, where he was then bishop until his death in 1360.[45] His biography in the *gesta episcoporum* does not mention a li-

43 Ibid., 224.
44 Ibid., 332.
45 Lebeuf, *Mémoires concernant l'histoire ecclésiastique et civile d'Auxerre, tome premier,* 454–56; Lebeuf, *Mémoires concernant l'histoire civile et ecclésiastique d'Auxerre et de son diocèse,* 517–18.

turgical reform.[46] Some years later, Guillaume d'Estouteville, bishop of Auxerre at the time of the production of the first witnesses, was born in Normandy, was bishop of Évreux until 1376, then of Auxerre until 1382, then of Lisieux until his death in 1415.[47] In this latter instance, there is no proof that bishop Guillaume had connections to Bayeux. The second link is weak, but the first is direct, and our statistical view on the liturgy, which is not based on the first responsories, invites a fresh look at the liturgical evidence.

Finally, the set of Poitiers is well established (14, 72, 68, 57, 32, 46, 24, 51, 40/38). K. Ottosen treats it as a "derivation" from a type that is very difficult to characterize, with its earliest witness in Aurillac, but mostly attested in Provence with some outliers in Évreux in Normandy and Soignies in northern France.[48] Despite the common beginning (14, 72, 68), our Figure 3 shows Poitiers has no connection with Aurillac or the other manuscripts from the group. The only set that shares at least six responsories in common with Poitiers is from Ivrea (**14**, **72**, 24, **57**, **32**, 68, 67, **51**, **38**), while Aurillac only has three (**14**, **72**, **68**, 32, 38, 79, 67, 40, 64). The responsory 51 in position R10, common to Poitiers and Ivrea, is very rare. Among the very few other examples is Bordeaux (**14**, **72**, 24, 46, **32**, 57, 68, **51**, **40**). This latter use has only five responsories in common if we take the position into account, but also has exactly the same ensemble of responsories. The difference between Poitiers and Bordeaux consists of only two interchanges: responsories 24 and 68 in R3 = R10 and responsories 57 and 46 in R5 = R7. There is an evident connection, since Poitiers is a suffragan diocese of the Archdiocese of Bordeaux. This link is missing in our network (Figure 3), because we chose to create edges for more than five responsories in common.

From our examples, we can draw three methodological conclusions. First, a positive conclusion – the linear presentation and exploitation by K. Ottosen is unable to capture proximities, and even historically proven and basic links are hidden if one takes the responsories in their sequence. This is where network analysis based on the entire set of responsories provides an enhanced access to comprehending the differences between liturgical uses. The second conclusion is that a greater statistical proximity is not always proof of a closer link (Bordeaux is a better match than Ivrea). Third, it would probably be beneficial to introduce a hybrid analysis of the responsories as ensemble and as series at a further stage, in order to better evaluate and weight the existing differences. All these observations do not directly contradict K. Ottosen's highly nuanced conclusions and hypotheses. They do, however, show how (social) network analysis techniques may

46 Sot, *Les gestes des évêques d'Auxerre, Tome III,* 30–32.
47 Chartraire, 'Auxerre. IV. Liste des évêques d'Auxerre'.
48 Ottosen, *The Responsories and Versicles of the Latin Office of the Dead,* 312–15.

be applied to the same source data and offer an additional view that allows us to better understand proximities, to spot links that a linear presentation obfuscates, and to position liturgical series in relation to one another.

3.2 Hours of the Virgin

Unlike in the case of the Office of the Dead, with the in-depth study by K. Ottosen, there is no equivalent for the Hours of the Virgin, and there is no analysis of correspondences and influences between uses, despite repertories of varying pieces within the liturgy across different uses. We will demonstrate that representing the textual findings as a network is also helpful in this case.

3.2.1 *Data*

The Hours of the Virgin are composed of 400 to 500 different pieces, distributed in eight hours, each divided into five sections. As is the case for the responsories of the Office of the Dead, the possible variations are immense, including the absence of specific pieces at specific moments of the liturgy.

Scholars have used the variants to identify liturgical uses. For instance, F. Madan used only four pieces among the four hundred, namely the antiphons and chapters of Prime and None.[49] V. Leroquais used a much larger set of texts in his unpublished notes on 198 distinct uses,[50] and E. Drigsdahl makes a similar selection of nearly thirty texts for 85 distinct uses.[51] There is much variation that Leroquais's and Drigsdahl's data cannot capture, but their data are sufficient to nourish a liturgical or historical analysis of the Hours of the Virgin, and we have decided to initiate a database with the available information and to enhance it, namely to disambiguate similar incipits and to check if absent texts were only missing in the records, or also from the source manuscripts (recorded as "Expected but not attested" even if some positions are very rare). We then expanded the data under review, and we now have more than one hundred recorded texts for certain manuscripts. For 160 uses, we have recorded more than 55 different texts, excluding "Expected but not attested" (cf. Figure 4).

The uses with few recorded texts have little impact on our study. They are incomplete either because of the original documentation or through an incomplete previous recording. For example, the use of Tarentaise is partially described

49 Madan, 'The Localization of Manuscripts'.
50 MS. Paris, BnF, NAL 3162, available on Gallica since January 2018 (https://gallica.bnf.fr/ark:/12148/btv1b10033474t).
51 Drigsdahl, 'Late Medieval and Renaissance Illuminated Manuscripts – Books of Hours 1300–1530'.

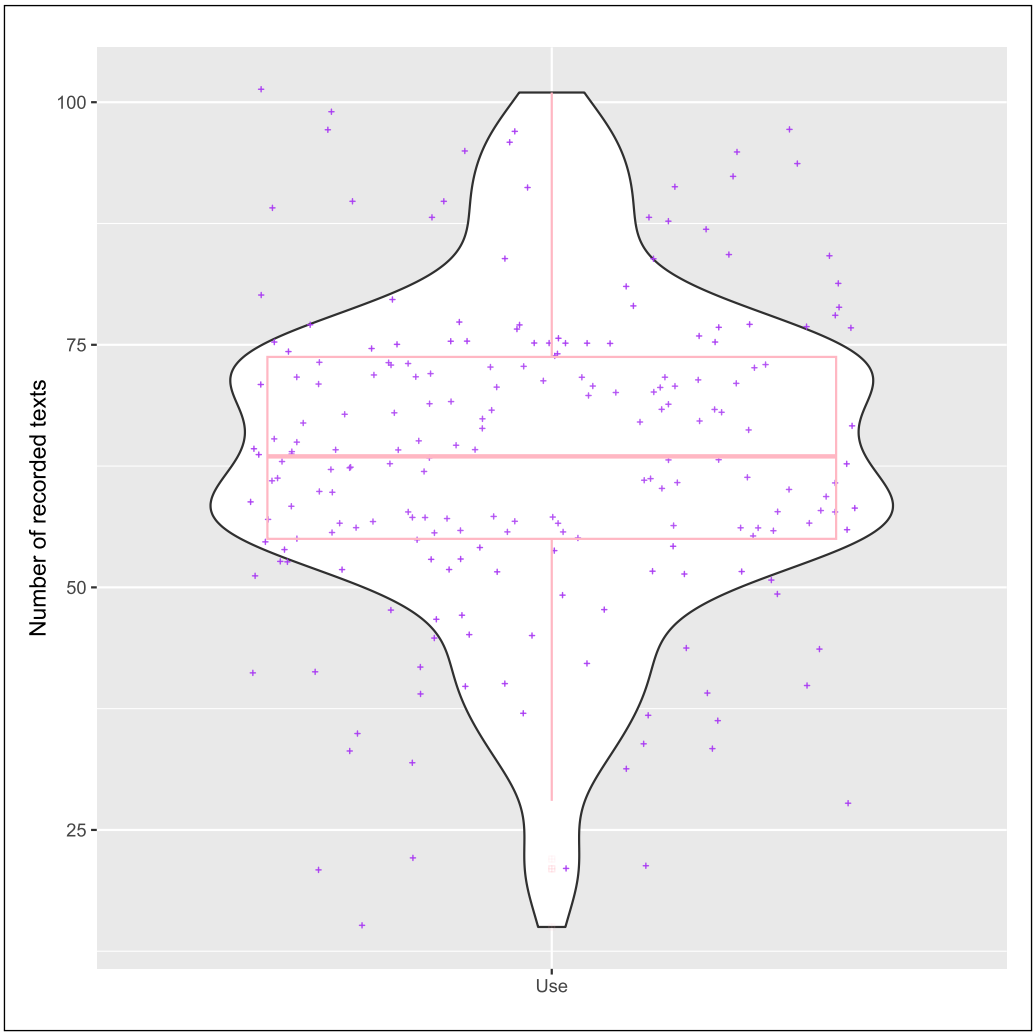

Fig. 4 Number of recorded attested texts by use.

by V. Leroquais, who stopped after Matins with a note stating that the text was useless; however, we have integrated this within in our data for future comparisons.[52] The uses of Uzès and Sisteron are documented by early printed breviaries in which the cursus for Prime, Terce, Sext and None is limited to the short lesson (*capitulum*). Most uses with fewer than 41 recorded texts are monastic, and therefore marginal in our study.

52 "Ce début d'office est intitulé: 'Incipit officium beate Marie in magno tempore'. Ce n'est pas l'*officium parvum*" (Paris, BnF, NAL 3162, f. 41r).

Text	#
Laetatus sum | Psalm 121	348
Nisi Dominus | Psalm 126	311
Ave maris stella | Cantus ID 8272.1	215
Ad te levavi oculos | Psalm 122	212
Veni creator Spiritus | AH 27066	209
Qui confidunt in Domino | Psalm 124	208
In convertendo | Psalm 125	207
Memento salutis auctor | Cantus ID 8248z = AH 5047	205
Nisi quia Dominus | Psalm 123	203
Dominus regnavit | Psalm 92	200
Benedicta tu in mulieribus | Cantus ID 001709 | Luke 1:42	192
Domine, Dominus noster | Psalm 8	190
O gloriosa domina | Cantus ID 008375e	189
Caeli enarrant gloriam Dei | Psalm 18	178
Quem terra, pontus, aethera| AH 22290 = Cantus ID 008375	178
Domini est terra | Psalm 23	176
Levavi oculos meos | Psalm 120	176
Ad Dominum cum tribularer | Psalm 119	174
Beati omnes qui timent Dominum | Psalm 127	170
Sancta et immaculata virginitas | Cantus ID 007569	161
Sancta Maria succurre miseris | Cantus ID 004703.1	161
Domine, non est exaltatum cor meum | Psalm 130	157
Sicut cynamomum et balsamum | Ecclesiasticus 24:20	156
Saepe expugnaverunt me a juventute mea | Psalm 128	151
Beata es Maria quae Dominum portasti | Cantus ID 006163	150

Tab. 8 Most frequent texts in the Hours of the Virgin across 214 liturgical uses recorded in the HORAE database (here: texts linked to 150 or more records).

As illustrated by Table 8, some texts appear more than 300 times across our 214 liturgical uses, which means they are frequently read twice a day in the same cursus. Psalm 121, for example, is commonly part of the psalm section of Terce and of Vespers, but also, albeit less frequently, in Prime and Vespers as in Agde, Aix-en-Provence or Béziers. Psalm 126 is similarly often read more than once a day, generally at None and Vespers, but also at Sext and Vespers in some dioceses.

3.2.2 *Hours of the Virgin as a network*

The inner complexity of the liturgical cursus for the Hours of the Virgin may be studied with the same methods as for the responsories in the Office of the Dead. Figure 6 represents a network of the liturgical uses as a one-mode network. Each dot represents one of the uses declared by V. Leroquais and E. Drigsdahl, for which they may have several sources, and to which we have added our own observations on medieval manuscripts and early printed books.

Edges and weights are created and calculated according to the percentage of common pieces within common recorded positions. Our index is not an IoU or Jaccard index, since we may have different levels of completeness in the description for different uses and it would distort the placing of less well described uses. Let's imagine a use that is known only by a fragment with a series of thirty pieces and which is in complete agreement, for this part, with another use for which we know a set of 100 texts. IoU would be low (30/100 = 0.3), whereas our computation will give a percentage of 100%. This is necessary due to the state of the documentation.

Figure 5 provides an overview of the calculated similarities between the different liturgical uses. As we compare 216 different uses, we have 45796 potential edges, if we include comparisons from A to B and B to A as well as comparisons with self. There is a total of 22791 (N × (N – 1) / 2) edges if reciprocal links are de-

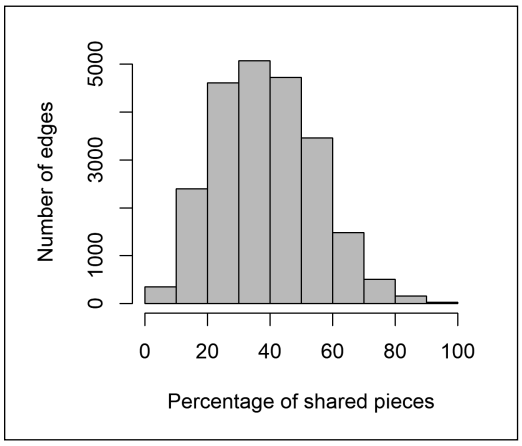

Fig. 5 Distribution of the similarity between 216 different uses. For example: in 200 cases of 1-to-1 comparison, the uses share 12% of their constituent pieces in the same positions.

clared once. Most uses have between 20% and 60% of the recorded texts in common in the same positions.

Given the repartition of percentages, we chose a cut-off point of 70 to retain a sufficient number of significant commonalities. Therefore, we have declared edges between nodes (liturgical uses) if the percentage of shared texts at the same positions is higher than 70%. Moreover, edges are weighted by the percentage. There are 691 edges linking 159 different nodes (institutions).[53]

We have introduced no other weighting parameters based on the length of the texts, their liturgical nature (psalms, hymns, antiphons, etc.), their performance in the collective liturgy (chanted or read), nor through scaling. We have also excluded the sixteenth-century use of Tours, which our records show as a mix of different sources that generates multiple links between otherwise separate clusters.

Figure 6 represents the textual network for the Hours of the Virgin in a similar way as Figure 3 represents the network for the Office of the Dead, with a Fruchterman-Reingold force directed algorithm. These two networks diverge and cannot be superposed. From a liturgical and historical standpoint, this is an important clue to the development of the liturgy at a local level. An office is introduced or reformed at a certain point in time and records, *volens nolens,* the connection of the institution at that moment. As both offices were adopted at different points in time, the connections are strikingly different.

For the Hours of the Virgin, an important first conclusion is that groupings appear on a geographical basis. For instance, a single group gathers almost all German dioceses. From a historical perspective, this correlates to K. Ottosen's demonstration for the Office of the Dead and shows how, beyond the desire of identifying manuscripts, recording the many pieces and elements of such an office may help in tracing regional coherences and probable shared liturgical developments. In this instance, the cluster of old Benedictine abbeys in a disconnected geographical setting in northern and eastern France is most likely a trace of very early liturgical connections when the small (votive) Office of the Virgin started to be recited on a daily basis. Likewise, a group covering Normandy and Brittany in France (Rouen, Sainte-Croix in Saint-Lô, Évreux, Bayeux, Avranches, Dol, Tréguier, Saint-Brieux), but also England with Sarum and York, bears witness to the known connections between Normandy and England.

53 There are 2177 edges between 185 nodes with a threshold at 60%. The overall graph is similar, but much denser, because of the numerous weak links.

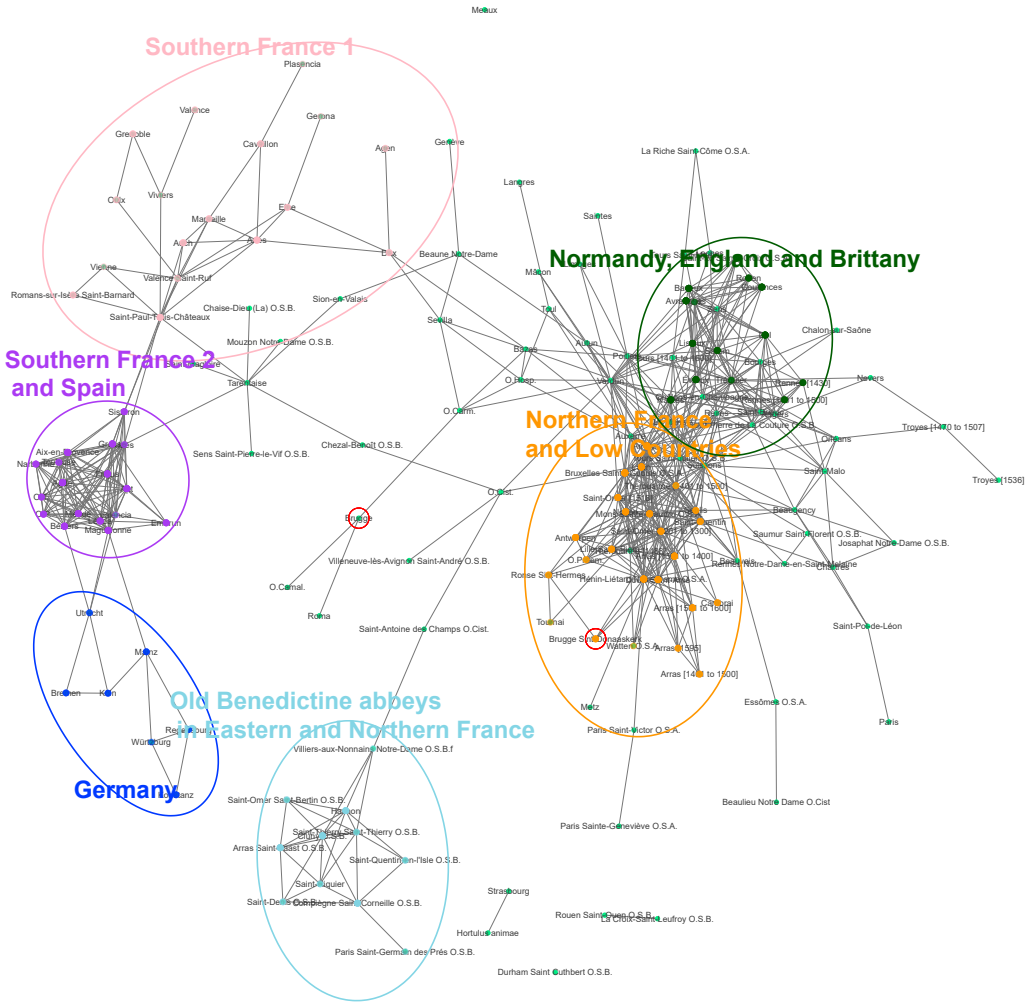

Fig. 6 Hours of the Virgin: network of liturgical uses. Network graph by the authors, using R libraries "network" and "ggnetwork", applying the Fruchterman-Reingold algorithm.

At a second level, splits and divisions are as important as the first regional clustering. Two groups cover southern France and Spain with some nearby cities (e.g., Narbonne and Montpellier or Marseille, Arles and Aix-en-Provence) in different clusters.

A special mention should be made to the liturgy of the Dominican order. In our statistics and network, it is 82% identical with Toulouse, Béziers and Maguelonne. As the order was founded in Toulouse in 1215, it may not come as a surprise. However, the liturgy of the order was completely redesigned under the direction of Humbert of Romans, fifth Master General of the Order, to accom-

modate the needs of an internationalized order and the demands from brothers of various origins and habits. A.-E. Urfels-Capot stresses that, in this operation, the Preachers have precisely introduced a major innovation in the votive office of the Virgin *in sabbato* with a complete set of lessons.[54] Even the "Dominican prototype" of 1256 (MS. Rome, Santa Sabina, XIV L 1) stresses that the set of lessons for the daily office is "not original": *"Istae lectiones, sive potius orationes, non originales sunt"*.[55] There was therefore no need for the Little Office to keep its text set close to the order's local origins.

V. Leroquais described the office based on two late books of hours (Hours of Frederick of Naples or of Aragon in MS. Paris, BnF, lat. 10532, and MS. Paris, BnF, Bibl. Arsenal, Ms-438), to which E. Drigsdahl added a printed book of hours and a comparison with the Hours of Jeanne d'Évreux.[56] These late sources may cast doubt on the context and actors using this set of texts and their faithfulness to the Dominican liturgy. However, we can confirm that the text given by V. Leroquais and E. Drigsdahl corresponds to the Dominican liturgy. For example, we have the same liturgy in the Dominican prototype, in which the daily office is described on f. 85^{v}–86^{r}, or again in the copy for the Master of the Order (MS. London, British Library, Add. MS 23935, on f. 139^{v}–140^{v}),[57] or in a random breviary (MS. Paris, BnF, Bibl. Arsenal, Ms-193, on f. 151^{v}–153^{v}).[58] P. Gleeson also stressed that the "Saint Dominic breviary", which furnishes a text linked to the founder of the order from before the Humbertian reform, is even closer to Toulouse and the

54 Urfels-Capot, 'Le sanctoral du lectionnaire de l'office', 319–20; Urfels-Capot, *Le sanctoral du lectionnaire de l'Office dominicain, 1254–1256,* 582–92.

55 Urfels-Capot, *Le sanctoral du lectionnaire de l'Office dominicain, 1254–1256,* 479.

56 Drigsdahl, 'Hore Beate Marie Virginis – Use of the Dominican Order'.

57 On this manuscript, cf. Huglo, 'Comparaison du "Prototype" du couvent Saint-Jacques de Paris avec l'exemplaire personnel du maître de l'Ordre des Prêcheurs (Londres, British Library, Add. ms 23935)'.

58 In all these manuscript sources, there is one difference from E. Drigsdahl's description. The antiphon *"Regem Virginis filium"* for the invitatory at Matins replaces the *"Ave Maria"* erroneously indicated by E. Drigsdahl, perhaps because in the Dominican liturgy, the verse *"Ave Maria"* is to be said before each hour. But the antiphon for the invitatory is also indicated as *"Regem Virginis filium"* in the books of hours of Jeanne d'Évreux (MS. New York, Metropolitan Museum, The Cloisters Collection, Acc. 54.1.2, f. 17^{r}), as well as in MS. Paris, BnF, lat. 10532, p. 108. The notes of V. Leroquais should be clarified for the responsory of the second lesson of Matins transcribed *"Beata es virgo Maria"* as in *"Beata es virgo Maria Dei genetrix quae credidisti"* (Cantus ID 6165), but it is *"Beata es Maria quae Dominum portasti creatorem"* (Cantus ID 6163). In the series of psalms at Lauds, the breviaries do not record Ps. 66 (*Deus miseratur*) before *Benedicite,* but it is present in the hours of Jeanne d'Évreux (f. 39^{v}) and Frederick of Naples (p. 124). Moreover, the breviaries and V. Leroquais and E. Drigsdahl do not explicitly indicate the psalms 149–150 after psalm 148 at Lauds, but they are attested by the Hours of Jeanne d'Évreux (f. 45^{v}–48^{r}) and Frederick of Naples (p. 128–129). At Vespers, V. Leroquais notes *"Ecce mater pulcre dilectionis"*; it stands for *"Ego mater pulchre dilectionis",* as correctly given by E. Drigsdahl.

Teutonic order.[59] This manuscript cannot, however, be considered as the source of the Dominican liturgy, as the lessons, responsories and versicles of the Office of the Dead do not correspond to any identified source.[60]

On top of this first observation, the main surprise in our study of the network of uses is the connection of German dioceses. Indeed, Constance, Regensburg, Würzburg, Cologne, Mainz, Bremen, but also Utrecht are collocated on the graph beyond the southern French groups and only have connections to this group. Our working hypothesis is that the German votive liturgy was developed (or perhaps reformed) at a relatively late stage in the thirteenth century based on examples from southern European regions. If our dating is correct, this probably happened through an imitation of Dominican or Teutonic offices. It would require an in-depth study to prove the connections, which is beyond the scope of this article. The question is nevertheless addressed below in section 4 for the specific case of Utrecht.

A final observation is that, apart from Utrecht in the German cluster, two other groupings concern the region of Flanders, Hainaut and the Low Countries. In this network, we have two representants for Bruges: one from Saint-Donatian, close to other regional uses, and a second, close to the isolated Roman use (both highlighted in red on Figure 6). The specificities of the textual networks of the liturgy in the southern Low Countries will now be addressed in part 4.

4. Network and hybridization: Definitions and examples from the Low Countries

In the liturgical networks devised above in section 3, several phenomena appear specific to the Low Countries. We will now turn to this region, which is the basis for our further analysis on hybrid texts and volumes.

The Low Countries were one of the most important centers for the production of books of hours in the Middle Ages. In our HORAE research project database,[61] we know of 520 books of hours that originate from this region. This production has already been the subject of numerous studies, mainly from an art historical perspective. For example, focusing on iconography and statistical correlations, D. Vanwijnsberghe evidenced seven variants on the Flemish Infancy cycle and

59 Gleeson, 'The Pre-Humbertian Liturgical Sources Revisited', 111.

60 Ottosen, *The Responsories and Versicles of the Latin Office of the Dead,* 241–42; Gleeson, 'The Pre-Humbertian Liturgical Sources Revisited', 110–11. The responsories correspond, to the extent the incomplete manuscript contains them, to the Bamberg minor series. The manuscript originated in southern France or Spain.

61 Stutzmann, 'HORAE – Hours: Recognition, Analysis, Editions'.

proved how different iconographical cycles were typical of a local production.[62] In an earlier study, he analyzed the artistic production with the unique set of texts provided by the Hours of the Virgin in a manuscript painted by Jean Tavernier (Brussels, KBR, IV 1290),[63] especially the short lesson (*capitulum*) *"Virgo Verbo concepit"* and the hymn *"Fit porta Christi"* at Compline.[64] He stressed that many Flemish books of hours present unidentified variants. We will address this point later.

This part will focus on the Hours of the Virgin and the hybridity of liturgical uses in manuscripts produced in this region.[65] First, we will further explore the links between the Dominican use and Utrecht and propose a new hypothesis on the constitution of the use of Utrecht. We will then explore a liturgical use called "Bruges" by V. Leroquais and prove that it comes from a book of hours for the use of Rome, and was obtained by the replacement of a whole section. In both cases, we will try to explain the process of hybridization.

4.1 The use of Utrecht, Windesheim and Geert Groote

Most books of hours of the northern Low Countries were written in Dutch from 1410 onwards, representing 90% of the extant books of hours from the region.[66] The linguistic features of translations, including the most prominent one by Geert Groote, and their relationship with Latin models, especially for the Hours of the Virgin, is a field of ongoing research. After the seminal publication by F. Gorissen in 1968 claiming that Geert Groote compiled the text that would become the cursus for the use of Utrecht and Windesheim,[67] a publication by J. Marrow called for a renewed scrutiny of the liturgical uses and variants in the Hours of the Virgin in the Low Countries,[68] and A. Korteweg proposed a clear distinction of the uses and an explanation for Geert Groote's text.[69] Analyzing the calendar and the litanies, as well as both the Hours of the Virgin and the Office of the Dead, she first observed that the uses of Utrecht and the Windesheim congregation are easily distinguished (the addition of Augustine as patron and of Meinulfus after 1430, changes in the order of the litanies).[70] She then repeated and exemplified

62 Vanwijnsberghe, 'The Cyclical Illustrations in the Little Hours of the Virgin in Pre-Eyckian Manuscripts'; Vanwijnsberghe, 'Le cycle de l'Enfance des petites heures de la Vierge dans les livres d'heures des Pays-Bas méridionaux'.
63 Vanwijnsberghe, 'L'apport du texte et des éléments de contenu'.
64 Ibid., 44.
65 The oral presentation of this paper included remarks on the Office of the Dead that would require a full demonstration. We reserve this for a future publication.
66 Korteweg, 'Books of Hours from the Northern Netherlands Reconsidered', 235.
67 Gorissen, 'Das Stundenbuch im rheinischen Niederland'.
68 Marrow, 'Notes on the Liturgical "Use" of the Hours of the Virgin in the Low Countries'.
69 Korteweg, 'Books of Hours from the Northern Netherlands Reconsidered'.
70 Ibid., 239–40.

K. Ottosen's findings on the differences for the Office of the Dead by adding several instances of manuscripts that contain both major and minor series.[71] With regards to the Hours of the Virgin, despite the lack of early witnesses, she introduced archival evidence to argue that a Utrecht use existed before Geert Groote's work – who acted mainly as a translator – and that it was integrated practically unchanged in the liturgy of Windesheim, only being regularized onto the Roman cursus.[72] Dutch translations were also copied based on different Latin models, for example transmitting the uses of Rome,[73] St. Gudula in Brussels, and Antwerp or at least a Brabantine institution,[74] or with the specific use for Malines (Mechelen).[75]

In our analysis of the liturgical uses of the Hours of the Virgin as a network (cf. Figure 6), we stressed that Utrecht was the use from the German cluster closest to the southern French cluster. Therefore, we ought to address the conclusions of F. Gorissen and A. Korteweg. F. Gorissen stated that Geert Groote wanted to compile a cursus avoiding Roman elements, because they were in use in the Franciscan liturgy.[76] He tries to connect the version by Geert Groote to the Bridgettines and, above this, to the Cistercians, and then comes to the Dominicans.[77] As the Bridgettines have a complex cursus with variations according to the days of the week, F. Gorissen points to the presence of texts not only on Thursday (*feria quinta*), but also on other days.[78] He concludes that the hours of Geert Groote were compiled from the offices of the Bridgettines and Dominicans.[79]

In Table 9, we compare the uses of Utrecht-Windesheim, Geert Groote and the Dominican order. The use of Utrecht-Windesheim and its variants are described according to A. Korteweg's description, with complements from V. Leroquais's and E. Drigsdahl's descriptions.[80] The use of the Dominican is described by combining the descriptions by V. Leroquais and E. Drigsdahl with our observations reported above.

---

71 Ibid., 241–46.
72 Ibid., 246–54.
73 Deschamps, 'Een zeldzaam afschrift van een middelnederlands getijdenboek, in de 14de eeuw in West-Vlaanderen ontstaan'; Marrow, 'Notes on the Liturgical "Use" of the Hours of the Virgin in the Low Countries', 285–86, 291.
74 Marrow, 'Notes on the Liturgical "Use" of the Hours of the Virgin in the Low Countries', 285–87, 291–93.
75 Marrow, 288, 294.
76 Gorissen, 'Das Stundenbuch im rheinischen Niederland', 90–91, 96.
77 Gorissen, 91–96.
78 Gorissen, 91–93.
79 Gorissen, 94, 96.
80 Leroquais, 'Paris, Bibl. nat. de France, NAL 3162. Répertoires bibliques, liturgiques et iconographiques du chanoine Victor Leroquais. VI Hymnes, antiennes, psaumes, leçons, répons et capitules des Heures de la Vierge.'; Drigsdahl, 'Hore Beate Marie Virginis – Use of Utrecht c. 1430'.

	Use of Utrecht	Geert Groote	O.Praed.
Matins			
Invitatory	In honore beatissimae Mariae virginis *Variant.* Ave Maria gratia plena	≡	≠ 'Regem Virginis' *Variant.* ≡ **ED*
Hymn	Quem terra pontus aethera *Variant. No hymn *VL *ED*	≠ ≡	≡ ≠
Antiphon I	Benedicta tu in mulieribus	≡	≡
Psalm I	Domine Dominus noster **VL*		≡
Psalm II	Caeli enarrant gloriam Dei **VL*		≡
Psalm III	Domini est terra **VL*		≡
Lesson I	Sancta Maria, uirgo uirginum	≡	≡
Responsory I	Sancta et immaculata virginitas **VL *ED*		≡
Lesson II	Sancta Maria, piarum piissima	≠	≡
Responsory II	Beata es Maria quae Dominum portasti **VL*		
Lesson III	Sancta Dei genitrix que digne meruisti	≠	≡
Responsory III	Felix namque es sacra virgo **VL*		≡
Lauds			
Antiphon I	O admirabile commercium	≡	≠ 'Post partum'
Psalm I	Dominus regnavit **VL *ED*		≡
Psalm II	Jubilate Deo **ED*		≡
Psalm III	Deus, Deus meus **ED*		≡
Psalm IV	Deus misereatur nostri**ED*		≡ **ED*
Psalm V	Benedicite omnia opera **ED*		≡
Psalm VI	Laudate Dominum de caelis **ED*		≡
Psalm VII	Cantate Domino canticum novum **ED*		≡
Psalm VIII	Laudate Dominum in sanctis ejus **ED*		≡
Short lesson	In omnibus requiem quesiui *Variant.* Ego quasi vitis **VL*	≠ ≡	≠ ≡
Hymn	O gloriosa domina excelsa	≡	≡
Benedictus Antiphon	Sub tuam protectionem confugimus *Variant.* O stella matutina *Variant.* Virgo piisima **VL*	≠ ≠ ≠ '[O] gloriosa' = O.P.	≠ ≠ ≠ 'O gloriosa' = G. Groote

	Use of Utrecht	Geert Groote	O.Praed.
Prime			
Hymn	Memento salutis auctor	≠	≡
	Variant. Rex Christe clementissime	≡	≠
	Variant. Veni creator spiritus	≠	≠
Antiphon	Quando natus es	≡	≠ 'Dignare me'
Psalm I	Ad Dominum cum tribularer **VL *ED*		≡
Psalm II	Levavi oculos meos **VL *ED*		≡
Psalm III	Laetatus sum in his **VL *ED*		≡
Short lesson	Ab initio et ante secula		≡
Responsory	Christe fili dei vivi **ED*		≠ 'Post partum'
Terce			
Hymn	*As Prime*	≡	≡
Antiphon	Rubum quem viderat	≡	≠ 'Gaude Maria virgo'
Psalm I	Ad te levavi oculos meos **VL *ED*		≡
Psalm II	Nisi quia Dominus erat in nobis **VL *ED*		≡
Psalm III	Qui confidunt in Domino **VL *ED*		≡
Short lesson	Et sic in Sion firmata sum	≡	≡
Responsory	Specie tua et pulchritudine tua **ED*		≠ 'Sancta Maria mater'
Sext			
Hymn	*As Prime*	≡	≡
Antiphon	Germinavit radix Jesse	≡	≠ 'In prole mater'
Psalm I	In convertendo **VL *ED*		≡
Psalm II	Nisi Dominus aedificaverit **VL *ED*		≡
Psalm III	Beati omnes qui timent **VL *ED*		≡
Short lesson	Et radicavi in populo honorificato	≡	≡
Responsory	Adjuvabit eam deus vultu suo **ED*		≠ 'Ora pro nobis'
None			
Hymn	*As Prime*	≡	≡
Antiphon	Ecce Maria genuit nobis salvatorem	≡	≠ 'Beata mater et innupta/intacta'
Psalm I	Saepe expugnaverunt me **VL *ED*		≡

	Use of Utrecht	Geert Groote	O.Praed.
Psalm II	De profundis clamavi **VL *ED*		≡
Psalm III	Domine, non est exaltatum **VL *ED*		≡
Short lesson	Quasi cedrus exaltata sum	≡	≡
Responsory	Diffusa est gratia in labiis tuis **ED*		≠ 'Elegit eam'
Vespers			
Antiphon I	Beata mater et innupta virgo	≡	≠ 'Sancta Dei genitrix'
Psalm I	Dixit Dominus Domino **VL *ED*		≡
Psalm II	Laudate, pueri, Dominum **VL *ED*		≡
Psalm III	Laetatus sum in his quae dicta **VL *ED*		≡
Psalm IV	Nisi Dominus aedificaverit **VL *ED*		≡
Psalm V	Lauda, Jerusalem, Dominum **VL *ED*		≡
Short lesson	Sicut cynamomum et balsamum *Variant.* In omnibus requiem **VL*	≠ ≡	≡ ≠
Hymn	Ave maris stella **VL *ED*		≡
Magnificat Antiphon	Sancta Maria succurre miseris **VL *ED*		≡
Compline			
Antiphon I	Cum jucunditate nativitatem	≡	≠ 'Virgo Maria non est'
Psalm I	Memento, Domine, David **VL *ED*		≡
Psalm II	Ecce quam bonum **VL *ED*		≡
Psalm III	Ecce nunc benedicite **VL *ED*		≡
Short lesson	Transite ad me omnes	≠ 'Sicut cynamomum'	≠ 'Ego mater pulchrae'
Hymn	Fit porta Christi pervia	≡	≠ 'Virgo singularis'
Nunc dimittis Antiphon	Glorificamus te Dei genetrix	≡	≠ 'Sub tuum praesidium'

Tab. 9 Comparison of the uses of Utrecht-Windesheim, Geert Groote and the Dominican order. Descriptions of Utrecht-Windesheim and Geert Groote uses are mainly from A. Korteweg, texts with asterisks are supplemented from other sources (*VL: Victor Leroquais; *ED: Erik Drigsdahl). For the Dominican cursus, see footnote 56. Signs ≠ and ≡ indicate the difference and identity with the use of Utrecht. Blank cells: missing information.

Agreeing with and expanding upon F. Gorissen's remarks, Table 9 shows that the uses of Geert Groote and Utrecht-Windesheim almost always agree with the Dominican use. Thanks to A. Korteweg, we can add that when the uses of Groote and Utrecht differ, at least one of them follows the Dominican version, while both disagree with the Dominican use for the series of antiphons from Lauds to Vespers, the responsories to short lessons (or *capitula*), and most of the cursus of Compline. In the following, we do not analyze the responsories to short lessons, because they are not always recorded or indicated in manuscripts.

While the uses of Cologne and the Cistercian order differ greatly in other respects, when it comes to the antiphons, Table 10 shows that the entire cycle of antiphons for the psalmody from Lauds to Vespers could have been taken from the Cistercian order, and the use of Cologne could plausibly explain the recorded variants. The reverse is true for the short lesson at Vespers (or, alternatively, the use of Cologne was based on the Cistercian use and kept possible variants).

As for the constitution of the cursus of Compline, it cannot be explained so easily. Cologne may have been the model for the hymn *"Fit porta Christi",* already mentioned by D. Vanwijnsberghe in the Tavernier hours (see above), and the antiphon for *Nunc dimittis.* Alas, we do not know its first antiphon. The short lesson *"Transite ad me omnes"* is unique to Utrecht at this position (it differs both from Geert Groote's and the Dominican use). Therefore there is no direct model. However, Cologne is one of only sixteen cursus where we know of this short lesson.

Only a renewed enquiry into the specific liturgy of Cologne could extend or nuance our observations. As indicated by A. Korteweg, following P. Séjourné,[81] Cologne adopted the Little Office of the Virgin in 1240. It shares some features of the Cistercian use and differs substantially from what the Dominican office was and remained after the Humbertian reform in 1256. We do not know when Utrecht adopted the recitation of the Little Office.

Our working hypothesis is as follows. The use of Utrecht was based on the Dominican use, and its antiphons for the psalmody from Lauds to Vespers were replaced by the Cistercian antiphons. The material of some short lessons and Compline was rearranged based on the use of Cologne, to which Utrecht was suffragan. This hypothesis is coherent with A. Korteweg's demonstration that the use of Utrecht is older than the version of Geert Groote, since the former is closer to the Dominican use (e.g., no inversion of lessons). In the lesson "*Transite ad me omnes",* unique to Utrecht, Geert Groote's use does not have the Dominican lesson either.

81 Korteweg, 'Books of Hours from the Northern Netherlands Reconsidered', 250.

	Use of Utrecht	O.Praed.	Selection of other uses
Matins			
Invitatory	In honore beatissimae Mariae virginis *Variant.* Ave Maria gratia plena	≠ *Variant* ≡	≡ Köln *Variant* ≡ Cist.
Lauds			
Antiphon I	O admirabile commercium	≠ 'Post partum'	≡ Köln, Cist.
Short lesson	In omnibus requiem quesiui *Variant.* Ego quasi vitis **VL*	≠ ≡	≡ Cist. *Variant* ≡ Köln
Benedictus Antiphon	Sub tuam protectionem confugimus *Variant.* O stella matutina *Variant.* Virgo piisima **VL*	≠ ≠ ≠ 'O gloriosa' = G. Groote	≡ Köln ≠ Cist. 'Beata dei genitrix'
Prime			
Hymn	Memento salutis auctor *Variant 1.* Rex Christe clementissime *Variant 2.* Veni creator spiritus	≡ ≠ ≠	≡ Cist. *Variant 1.* ≡ Köln
Antiphon	Quando natus es	≠ 'Dignare me'	≡ Köln, Cist.
Terce			
Antiphon	Rubum quem viderat	≠ 'Gaude Maria virgo'	≡ Köln, Cist.
Sext			
Antiphon	Germinavit radix Jesse	≠ 'In prole'	≡ Köln, Cist.
None			
Antiphon	Ecce Maria genuit nobis salvatorem	≠ 'Beata mater et innupta/intacta'	≡ Cist. ≠ Köln
Vespers			
Antiphon I	Beata mater et innupta virgo	≠ 'Sancta Dei genitrix'	≡ Cist. ≠ Köln 'Ecce tu pulchra'
Short lesson	Sicut cynamomum et balsamum *Variant.* In omnibus requiem **VL*	≡ ≠	≡ Köln *Variant.* ≡ Cist.
Compline			
Antiphon I	Cum jucunditate nativitatem	≠ 'Virgo Maria non est'	? Köln ≠ Cist.
Short lesson	Transite ad me omnes	≠ 'Ego mater pulchre'	≠ Köln 'Multae filae' ≠ Cist. 'Sicut cynamomum'
Hymn	Fit porta Christi pervia	≠ 'Virgo singularis'	≡ Köln ≠ Cist.
Nunc dimittis Antiphon	Glorificamus te dei genetrix	≠ 'Sub tuum praesidium'	≡ Köln ≠ Cist.

Tab. 10 Comparison between Utrecht, Cologne and the Cistercian order when the use of Utrecht differs from the Dominican use.

This new hypothesis of a Utrecht use based on the Dominican use with some systemic changes originates in the observation of the complete network of uses as provided in section 3.2.2. In it, not only Utrecht and Cologne, but also a series of German uses are close to the Dominican use. It is unlikely that all connections depend on Geert Groote's compilation and that none of these dioceses had its own use before Groote's work. On the contrary, from a liturgical standpoint and as seen for the different offices, it is much more likely that the Utrecht use emerged in imitation of a given cursus while considering the liturgical context. The last part of the hypothesis relies on the theory that texts are replaced in coherent patterns and in a meaningful process, as already stressed by K. Ottosen. Creating a use does not work by cherry-picking at random; the liturgical analysis requires a finer level of observation.

We will now apply the same method of analysis to the "Bruges" uses described by V. Leroquais and attempt to characterize the process of creation of a hybrid use.

4.2 The southern Low Countries and "Bruges" in the network

For books of hours in Latin produced in the Low Countries, Bruges was a leading site of production that developed in workshops from the end of the fourteenth century onwards.[82] Some of the production of books of hours in the Low Countries and Bruges was obviously destined for a local market. Beyond art historical evidence, textual clues for the origin in Bruges are the mentions of saints Adrian, Donatian, Walpurga, Gertrude, Aldegonde or Bavo of Ghent in the calendar, suffrages, and litanies.

The liturgical uses of the Hours of the Virgin in the region display a mix of regional specificities and links to more distant regions due to historical connections, as evidenced in our Figure 6. A rather large gathering groups together institutions and dioceses in northern France and Flanders, including Amiens, Arras (three uses attested in the fifteenth and sixteenth centuries, but not the one from the fourteenth century), Beauvais, Saint-Amé in Douai, Hénin-Liétard, Laon, Noyon, Saint-Quentin, Thérouanne (uses attested in the thirteenth century and in 1488), and Saint-Omer, and in modern-day Belgium, including Antwerp, Saint-Donatian in Bruges, St. Gudula in Brussels (O.S.A.), Sainte-Waudru in Mons (O.S.A.), Sint-Hermes in Ronse, Tournai, and Watten (O.S.A.). Also belonging to this cluster are Cambrai in modern-day France, but belonging to the Holy Roman Empire in the Middle Ages, and the Premonstratensian order, whose mother house lies in the same region, not far from Laon. Extending partially into

82 Bergen, 'De Meesters van Otto van Moerdrecht. Een onderzoek naar de stijl en iconografie van een groep miniaturisten, in relatie tot de productie van getijdenboeken in Brugge rond 1430'; Wijsman, *Luxury bound*.

the same region, there is a smaller clustered gathering of older Benedictine abbeys in eastern and northern France, with Hasnon, Saint-Thierry near Reims, Saint-Corneille in Compiègne, Saint-Quentin-en-l'Isle, Saint-Vaast in Arras, Saint-Bertin in Saint-Omer, Saint-Denis and Saint-Germain-des-Prés near Paris, Villiers-aux-Nonnains, and also Cluny in Burgundy. In addition, we have discussed how the use of Utrecht was connected to the uses in southern France and the Dominican liturgy.

Yet a large part of this production was destined for export. The texts, calendars, hours or litanies of such book of hours do not depend on the liturgical uses of the Low Countries, but are generally based on the target market with adaptations, especially for English uses. To discern such books of hours for the English market, N. Rogers analysed heraldic features, but also "Lincoln calendars" and rare saints, showing two prominent examples of how the devotion to specific saints may help to identify or presume English patrons,[83] but she does not go into a textual analysis of the offices. Another obvious indication is provided by books of hours with offices for the use of Sarum. The insertion of the widespread use of Rome may also be a clue to an intent to export, as it can be considered universal. 54% of the manuscripts produced in the cities of the Low Countries in our database show Hours of the Virgin for the use of Rome, with an additional 5% for the use of Sarum. The manuscripts sent to the English market most often have a calendar for Sarum or York, and specific texts like the "Fifteen O's" by Bridget of Sweden, the Commendation of Souls, the prayers to the Five Wounds of Christ, the Psalter of the Passion, or the Psalter of Saint Jerome.[84] S. van Bergen explores the "influence of customers", most noticeably in the choice of the suffrages or *memoriae* which are always "drawn from a fixed repertoire of twelve saints" in England,[85] and the "sliding scale" ranging from standard books of hours produced for the free market and available off the shelf, to customized and supplemented volumes, and to manuscripts copied to order.[86]

With this knowledge, we can now interrogate two uses in our datasets. There are two representatives for Bruges: one from Saint-Donatian, close to other regional uses, and a second one for an unspecific "Bruges", named so by V. Leroquais, the single closest use to the otherwise very isolated Roman use (both highlighted in red in Figure 6). Leroquais's source is the manuscript MS. Rouen, Bibliothèque patrimoniale Villon, 3024 (Leber 137), hereinafter MS. Rouen 3024.[87] The attribution to "Bruges" is indicated without an explanation, but the use is different

83 Rogers, 'Patrons and Purchasers'.
84 Bergen, 'The Production of Flemish Books of Hours for the English Market', 271; Bergeron-Foote and Samson, 'Horae à l'usage de Sarum: Montréal, McGill, LRCS, ms 98'.
85 Bergen, 'The Production of Flemish Books of Hours for the English Market', 275.
86 Ibid., 276.
87 MS. Paris, BnF, NAL 3162, f. 112r.

from the collegiate church of Saint-Donatian of Bruges.[88] This latter use is described by V. Leroquais, and after him by E. Drigsdahl, from a breviary printed in 1520.[89] Given the lack of manuscripts with this use, E. Drigsdahl suggested that this cursus was created by the editors of the breviary using unknown earlier sources: "No book of hours, produced in Bruges in the fifteenth century, has so far been found to contain this office! I suspect that the editors of the breviary took recourse to an old source in the archives, and revived a long-forgotten use".[90] We will discuss the date of Saint-Donatian's use later. For now, it is unclear if Leroquais shows a liturgical use for an unidentified institution in Bruges or a different state of Saint-Donatian's liturgy, although the answer is probably neither.

The manuscript was produced in Bruges around 1410. Its calendar contains several saints from the Low Countries, including the very informative Saint Donatian, and lacks the Office of the Dead – which is a feature present almost exclusively in Flemish manuscripts. This justifies the assumption by V. Leroquais. On folio 12v, in front of the Annunciation, a miniature shows a kneeling man accompanied by Saint George with a dragon, a spear, and a helmet with a coat of arms (red cross on silver) as his identifying attributes (Figure 7). It is therefore fair to assume that the product was intended for the English market. But the use is atypical, close to but different from that of Rome.

The Hours of the Virgin are largely for the use of Rome, including its most unusual pieces, but with one major exception in the Matins as is evidenced in Table 11. The Roman readings of Matins are replaced by others, which enjoy widespread use, both in the region of origin and in the target export market. The three readings ("*Sancta Maria virgo*", "*Sancta Maria piarum piissima*", "*Sancta Dei genitrix*") and their responsories ("*Sancta et immaculata*", "*Beata es Maria*", "*Felix namque*") are shared by numerous uses in Europe, not only in Saint-Donatian of Bruges and Saint-Peter of Lille, but also, for instance, in the Dominican order, the English use of Sarum and the Norman use of Rouen. Therefore, the modification of the Roman use which takes place in this manuscript is difficult to characterize, but it is as if the Roman nocturn had been switched as a whole against one that is customized, perhaps the local use of Saint-Donatian in Bruges.

However, a second difference appears in the blessings and is more specific. MS. Rouen 3024 presents the following three blessings before the three readings: *Alma virgo virginum, Precibus sue matris, Nos cum prole pia.* By contrast, in the use of Rome, where there is one blessing before the first reading (*Nos cum prole pia*), no blessing before the second reading, and the blessing *Per virginem*

88 Ibid., f. 183r.
89 *Breviarium ad usum insignis ecclesie sancti Donatiani Brugensis dyocesis Tornacensis, Pars hiemalis,* Parisius: Anthonius Bonnemere, 1520.
90 Drigsdahl, 'Use of Bruges St. Donatian 1520'.

Fig. 7 MS. Rouen, Bibliothèque patrimoniale Villon, 3024 (Leber 137), f. 12v–13r. Source: https://gallica.bnf.fr/ark:/12148/btv1b101019444/f32 and https://gallica.bnf.fr/ark:/12148/btv1b101019444/f33.

	Rome	**MS. Rouen 3024**	**Saint-Donatian Bruges**
Ad Matutinum	Ave Maria	Ave Maria	Ave Maria
Hymnus	Quem terra	Quem terra	(No hymn)
Ant. ps. 8	Benedicta tu	Benedicta tu	Benedicta tu
Lectio i. R. i.	In omnibus requiem R. Sancta et immaculata	Sancta Maria virgo R. Sancta et immaculata	Sancta Maria virgo R. Sancta et immaculata
Lectio ii. R. ii.	Et sic in Syon R. Beata es Maria	Sancta Maria piarum R. Beata es Maria	Sancta Maria piarum R. Beata es Maria
Lectio iii. R. iii.	Quasi cedrus R. Felix namque	Sancta Dei genitrix R. Felix namque	Sancta Dei genitrix R. Felix namque
Ad Laudes	Pss. 92, 99, 62, 66, B, 148, 150	Pss. 92, 99, 62, 66, B, 148, 150	Pss. 92, 99, 62, 66, B, 148
Antiphona	Assumpta est	Assumpta est	Sancta Dei genitrix

	Rome	MS. Rouen 3024	Saint-Donatian Bruges
Capitulum	Viderunt eam filie	Viderunt eam filie	Virgo verbo
Hymnus	O gloriosa domina	O gloriosa domina	O gloriosa domina
Antiphona	Beata Dei genitrix	Beata Dei genitrix	Beata Dei genitrix
Ad Primam	Pss. 53, 84, 116	Pss. 53, 84, 116	Pss. 1, 2, 5
Hymnus	Memento salutis	Memento salutis	Gloria tibi Domine
Antiphona	Assumpta est	Assumpta est	Sub tuam protectionem
Capitulum	Que est ista	Que est ista	Hec est virgo
Ad Tertiam	Pss. 119, 120, 121	Pss. 119, 120, 121	Pss. 119, 120, 121
Hymnus	Memento salutis	Memento salutis	Ave maris stella
Antiphona	Maria virgo assumpta est	Maria virgo assumpta est	Cum iocunditate
Capitulum	Et sic in Syon	Et sic in Syon	Paradisi porta
Ad Sextam	Pss. 122, 123, 124	Pss. 122, 123, 124	Pss. 122, 123, 124
Antiphona	In odorem	In odorem	Sancta Maria
Capitulum	Et radicavi	Et radicavi	Virgo Dei genitrix
Ad Nonam	Pss. 125, 126, 127	Pss. 125, 126, 127	Pss. 125, 126, 127
Antiphona	Pulchra es	Pulchra es	Alma virgo Maria
Capitulum	In plateis sicut	In plateis sicut	Per te Dei genitrix
Ad Vesperas	Pss. 109, 112, 121, 126, 147	Pss. 109, 112, 121, 126, 147	Pss. 109, 112, 121, 126, 147
Antiphona	Dum esset rex	Dum esset rex	Beata mater
Capitulum	Ab initio et ante	Ab initio et ante	Beata es Maria
Hymnus	Ave maris stella	Ave maris stella	Ave maris stella
Antiphona	Beata mater et innupta	Beata mater et innupta	Sancta Maria succurre
Ad Completorium	Pss. 128, 129, 130	Pss. 128, 129, 130	Pss. 12, 42, 128, 130
Antiphona	– (no antiphon)	– (no antiphon)	Post partum
Capitulum	Ego mater pulchre	Ego mater pulchre	Gaude Maria virgo
Hymnus	Memento	Memento	Fit porta Christi
Antiphona	Sub tuum presidium	Sub tuum presidium	Glorificamus te

Tab. 11 Comparison of the main pieces in the uses of Rome and Saint-Donatian of Bruges and in MS. Rouen 3024.

matrem before the third reading. According to our research, the first and third blessings of MS. Rouen 3024 only occur together in these positions in the Dominican and thus Teutonic and Carmelite orders, which, however, include the blessing *Sancta Dei genitrix* in the second position. The uses of the possible target markets of Rouen and Sarum only share the first blessing with MS. Rouen 3024. They include another blessing in the second position (*Oret voce/prece pia*) and have the blessing *Sancta Dei genitrix* as a third blessing. As for Saint-Donatian of Bruges, the only source distributes the blessings along the days of the week, with the following list from Sunday to Saturday: *Sancta Dei genitrix, Sancte Marie precibus, Alma virgo virginum, Precibus sue matris, In omni tribulatione, Sancte Marie intercessio, Ad gaudia civium.*

What was the rationale for this switch? The liturgical use of Bruges cannot be a complete source for MS. Rouen 3024, given the lack of *Nos cum prole pia.* The combination of blessings in MS. Rouen 3024 may indicate a special devotion or connection to the Dominican order. However, the second blessing of MS. Rouen 3024 *Precibus sue matris benedicat nos filius Dei patris* does not originate in the Dominican use, whose blessing *Sancta Dei genitrix* was also a valid option in the local context. Although known in other regions or abbeys like Saint-Germain in Paris or Geneva, albeit not in the second position, the blessing *Precibus sue* is used almost exclusively in the Low Countries. This blessing is found in the same position in Saint-Peter in Lille, and also in Bruges on Wednesday. It is as if the Dominican series was used, but changed because of the clashes of blessings and positions in more local uses.

Therefore, we believe that MS. Rouen 3024 does not represent an autonomous use, for which we would lack an institution to ascribe it to, but rather an example of a hybridization process during the production. Here the universal use of Rome is mixed with external elements, including the replacement of one section with texts of the same section for other liturgical uses, here including elements which are clearly local. The use is clearly hybrid.

The concept of "hybridity" is apt to capture the state of such manuscripts. It defines the mixing of elements borrowed from different uses and posits that we can identify two or more liturgical uses as a reference, which are well-formed and described, stable, and predates the mixed forms. Hybrid forms may (1) replace and switch sections; (2) complement missing parts; (3) cross several elements. In this study, we have only addressed the first type of hybridity.

	ROME	O.PRAED.	MS. ROUEN 3204	LILLE (S. Peter)	ROUEN/ SARUM
Ad Matutinum					
Jube Domne benedicere					
Benedictio i.	*Nos cum prole pia*	*Alma virgo virginum*	*Alma virgo virginum*	*Alma virgo virginum*	*Alma virgo virginum*
Lectio i.	In omnibus requiem	Sancta Maria virgo	Sancta Maria virgo	Sancta Maria virgo	Sancta Maria virgo
R. i.	Sancta et immaculata	Sancta et immaculata	Sancta et immaculata	Sancta et immaculata	Sancta et immaculata
Jube Domne benedicere					
Benedictio ii.	[No blessing]	*Sancta Dei genitrix*	*Precibus sue matris*	*Precibus sue matris*	*Oret voce/ prece pia*
Lectio ii.	Et sic in Syon	Sancta Maria piarum	Sancta Maria piarum	Sancta Maria piarum	Sancta Maria piarum
R. ii.	Beata es Maria	Beata es Maria	Beata es Maria	Beata es Maria	Beata es Maria
Jube Domne benedicere					
Benedictio iii.	*Per virginem matrem*	*Nos cum prole pia*	*Nos cum prole pia*	*Sancta Dei genitrix*	*Sancta Dei genitrix*
Lectio iii.	Quasi cedrus	Sancta Dei genitrix	Sancta Dei genitrix	Sancta Dei genitrix	Sancta Dei genitrix
R. iii.	Felix namque	Felix namque	Felix namque	Felix namque	Felix namque
Canticle: Te Deum laudamus					

Tab. 12 Hours of the Virgin: readings and blessings in MS. Rouen 3024 compared to Dominican and local uses.

5. Conclusion on textual networks: hybrid, local, intentional?

This contribution has demonstrated how the notions and techniques of network analysis can be applied to liturgical texts and shed new light on complex data, in which seriality plays a crucial part. Section 3.1 re-examined the dataset compiled by K. Ottosen on the Office of the Dead. We were able to demonstrate that some parts of his presentation were misleading and that a full view of the network allows us to formulate more satisfactory hypotheses, especially for Auxerre

and Poitiers. The use of network analysis in section 3.2 then helped us extend the notion of dynamics and historical evolution in liturgy that we saw for the Office of the Dead to the Hours of the Virgin, uncovering regional coherences, but also a link between the southern France, the Dominican order and Utrecht. In part 4, we dived deeper in the phenomenon of hybridity. With examples chosen at the regional level of the Low Countries, we have formulated new hypotheses to explain what we know of the uses of Utrecht and of Bruges, the latter represented by an apparently incoherent, hybrid text. Switching, inserting, and mixing were key operations.

Network analysis and statistics can offer proof of a regional diffusion and help to measure how local a text is, on its own or through its context. In fact, not only may some texts have a purely regional diffusion, but their position in the cursus may also be characteristic of the uses of a region. Texts such as the blessing *Precibus sue matris* could therefore serve as an identity token and be recognized as a piece of local liturgy, despite its use in distant institutions. However, we have no evidence that people were aware of the locality of this reading. As a consequence, we cannot infer intentionality for such a hybridization.

Local or regional uses can also present other characteristics or changes. Textual variants in the blessings – and probably in other pieces – will most likely be specific of regional uses.[91] Hybrid uses lead us to consider a complex process of circulation of liturgical texts. "Pure" uses are adapted and incorporated into local text collections, depending on the place of production and the target market. This process raises the question of the status of texts inscribed in a regional network.

Our findings question the value and normativity of liturgical uses in books of hours. In theory, they follow the prescriptions of a diocese or order in their integrity. They are not, however, invested with the same authority as the *ordo* of a diocese and can certainly be modified. However, we lack the understanding of the reasons for modifications. Are they due to special devotions, practices of workshops, or liturgical modes? If so, how is it that specific sections or specific types of texts, such as antiphons, are modified and not others? We can now describe groups of manuscripts as well as groups of uses, but we would like to understand better the actors involved, the workshops, scribes, purchasers, and the intention behind the choice of adding or removing texts from a pre-existing liturgical form.

---

91 For example, the blessing *"Oret mente pia pro nobis Virgo Maria"* is a rare variant of a common verse *"Oret voce pia pro nobis Virgo Maria."* To our knowledge, this variant only appears in books of hours produced in the southern Low Countries and (due to trade?) in England: in Bruges (MS. Paris, Centre culturel irlandais, E3), Ghent and Tournai (MS. Paris, Beaux-Arts, Masson 22), Hainaut (MS. Cologny, Fondation Martin Bodmer, Cod. Bodmer 180) and York (MS. Boulogne-sur-Mer 93).

Liturgical texts are part of the cultural life and evidence for local, regional and international influence, through the diffusion of practices and applied exegesis. We can study books of hours as networks of texts and of manuscripts, both through the compilation process at the level of the codices, and through the compilation process within each office. It is a network of manuscripts, of texts, and of norms. Manuscripts may contain offices for different uses, and the uses themselves can be analyzed as networks of texts, both in their conjunctions and in their inner constitution. In this context, phenomena of hybridity appear.

From a methodological point of view, we have added the notion of seriality to the core of the usual text networks in manuscripts. The correct thresholding to define what should be considered a link is, as always, a sensitive issue, as is the correct weighting of presence and position here.

We have also recognized hybrid texts and developed tools to analyze them. They are both surprising with regard to the normativity of liturgy, and unsurprising with regard to the copying mistakes and other customization processes at work in the books of hours. The process and intentions are now open to enquiry. The availability of models (or lack thereof) may play a role, even if the Roman use and its texts were known in the region (although perhaps not fully within the workshops). The orality and memory of performed liturgy likely also played a role, with an unconscious reminiscence or quotation practice through embodied liturgical knowledge, which would explain the specific changes to chanted pieces. Beyond that, the reception and use of these hybrid manuscripts is also entirely unknown. Nor do we know if readers could recognize and perceive these quotations, switches, insertions, and combinations. Without additional sources, it is difficult to assess the space for ignorance and mistakes, and conversely the level of knowledge and intentionality. Indeed, we would need external testimonies in order to study if the actors and readers were aware that their text sets were regional, if the quotations disclose a regional identity, and if they were aware of liturgical norms and wished to produce structured and coherent hybrids.

6. References

Baroffio, Giacomo. 'Testo e musica nei libri d'ore'. *Rivista Italiana di Musicologia* 46 (2011): 18–77.

Batiffol, Pierre. *Histoire du bréviaire romain.* Paris: A. Picard et fils, 1893. http://gallica.bnf.fr/ark:/12148/bpt6k81445x.

Bergen, Saskia van. 'The Production of Flemish Books of Hours for the English Market: Standardization and Workshop Practices'. In *Manuscripts in Transition. Recycling Manuscripts, Texts, and Images,* edited by Brigitte Dekeyzer and Jan van der Stock, 271–83. Corpus of Illuminated Manuscripts 15. Leuven: Peeters, 2005.

— 'De Meesters van Otto van Moerdrecht. Een onderzoek naar de stijl en iconografie van een groep miniaturisten, in relatie tot de productie van getijdenboeken in Brugge rond 1430'. PhD, University of Amsterdam, 2007.
Bergeron-Foote, Ariane, and G. Samson. 'Horae à l'usage de Sarum: Montréal, McGill, LRCS, ms 98'. In *Catalogue raisonné des livres d'heures conservés au Québec,* edited by Brenda Dunn-Lardeau, 279. Québec: Presses de l'Université du Québec, 2018.
Beyssac, Gabriel. 'Moyen Court', –1965. Manuscript notes in 7 volumes. Le Bouveret, Saint-Benoît de Port-Valais, Fonds Beyssac. Photocopy of an annotated copy available at Institut de Recherche et d'Histoire des Textes.
Bibliothèque nationale de France, and Louisa Torres. 'Ms-569. Livres d'heures, en latin, à l'usage de Gand'. BnF Archives et Manuscrits, 2019. https://archivesetmanuscrits.bnf.fr/ark:/12148/cc797710.
Chartraire, Eugène. 'Auxerre. IV. Liste des évêques d'Auxerre'. In *Dictionnaire d'histoire et de géographie ecclésiastiques,* V: 945–47. Paris: Letouzey et Ané, 1931.
Clark, Gregory T. 'Beyond Jacquemart Pilavaine, Simon Marmion, and the Master of Antoine Rolin: Book Painting in the Hainaut in the Penultimate Decade of the Fifteenth Century'. In *Manuscripts in Transition. Recycling Manuscripts, Texts, and Images,* edited by Brigitte Dekeyzer and Jan van der Stock, 391–98. Corpus of Illuminated Manuscripts 15. Leuven: Peeters, 2005.
— 'Beyond Saints: Variant Litany Readings and the Localization of Late Medieval Manuscript Books of Hours. The D'Orge Hours'. In *Books of hours reconsidered,* edited by Sandra Hindman and James H. Marrow, 213–33. Turnhout: Brepols, 2013.
— 'Table of Strictly Regional Feasts in Calendars Printed or Written for Southern Netherlandish and Contiguous Dioceses, Abbeys, and Colleges before 1559'. In Gregory T. Clark, *Made in Flanders: The Master of the Ghent Privileges and Manuscript Painting in the Southern Netherlands in the Time of Philip the Good,* 289–329. Ars nova: Studies in Late Medieval and Renaissance Northern Painting and Illumination. Turnhout: Brepols, 2000.
— 'Appendix III. Qu'est-ce qu'un calendrier parisien?' In Gregory T. Clark, *Art in a time of war. The Master of Morgan 453 and manuscript illumination in Paris during the English occupation (1419–1435),* 209–55. Toronto: Pontifical Institute of Mediaeval Studies, 2016.
— 'AmsterdamUBxxvC26 (228)'. Beyond Use: A digital database of variant readings in late medieval books of hours, 2020. https://beyonduse.sewanee.edu/details.php?mssby=Alphabetical&manuscript=AmsterdamUBxxvC26.

Daille, Béatrice, Amir Hazem, Christopher Kermorvant, Martin Maarand, Marie-Laurence Bonhomme, Dominique Stutzmann, Jacob Currie, and Christine Jacquin. 'Transcription automatique et segmentation thématique de livres d'heures manuscrits'. *Traitement Automatique des Langues* 60, no. 3 (2019): 13–36.

Deschamps, Jan. 'Een zeldzaam afschrift van een middelnederlands getijdenboek, in de 14de eeuw in West-Vlaanderen ontstaan'. *Archives et Bibliothèques de Belgique/Archief- en Bibliotheekwezen in België* 68 (1997): 215–23.

Drigsdahl, Erik. 'Hore Beate Marie Virginis – Use of Bruges St. Donatian 1520'. Late Medieval and Renaissance Illuminated Manuscripts – Books of Hours 1300–1530, CHD Center for Håndskriftstudier i Danmark, 2013. http://manuscripts.org.uk/chd.dk/use/hv_bruges.html.

— 'Hore Beate Marie Virginis – Use of the Dominican Order'. Late Medieval and Renaissance Illuminated Manuscripts – Books of Hours 1300–1530' CHD Center for Håndskriftstudier i Danmark, 2013. http://manuscripts.org.uk/chd.dk/use/hv_dominican.html.

— 'Hore Beate Marie Virginis – Use of Utrecht c. 1430'. Late Medieval and Renaissance Illuminated Manuscripts – Books of Hours 1300–1530, CHD Center for Håndskriftstudier i Danmark, 2013. http://manuscripts.org.uk/manuscripts/chd.dk/tutor/utrecht.html.

— Late Medieval and Renaissance Illuminated Manuscripts – Books of Hours 1300–1530. CHD Center for Håndskriftstudier i Danmark, 2013. http://manuscripts.org.uk/chd.dk/.

Gleeson, Philip. 'The Pre-Humbertian Liturgical Sources Revisited'. In *Aux origines de la liturgie dominicaine: le manuscrit Santa Sabina XIV L I,* edited by Leonard E. Boyle and Pierre-Marie Gy, 99–114. Documents, études et répertoires de l'Institut de Recherche et d'Histoire des Textes 67. Paris: CNRS, 2004.

Gorissen, Friedrich. 'Das Stundenbuch im rheinischen Niederland'. In *Studien zur klevischen Musik- und Liturgiegeschichte,* edited by Walter Gieseler, 63–109. Beiträge zur rheinischen Musikgeschichte 75. Cologne: Arno-Volk, 1968.

Hamburger, Jeffrey F. 'Another perspective: the book of hours in Germany'. In *Books of hours Reconsidered,* edited by Sandra Hindman and James H. Marrow, 65–95. Studies in Medieval and Renaissance Art History. London: Harvey Miller Publishers/Turnhout: Brepols, 2013.

Hazem, Amir, Beatrice Daille, Dominique Stutzmann, Christopher Kermorvant, and Louis Chevalier. 'Hierarchical Text Segmentation for Medieval Manuscripts'. In *Proceedings of the 28th International Conference on Computational Linguistics,* 6240–51. Barcelona: International Committee on Computational Linguistics, 2020. https://www.aclweb.org/anthology/2020.coling-main.549.

Heijden, Frieda van der. 'Or Ai Ge Trop Dormi: A Study of the Unfinished F-Pn Fr. 12786'. PhD, Royal Holloway, University of London, 2018. https://pure.royalholloway.ac.uk/portal/files/30965857/Or_Ai_Ge_Trop_Dormi.pdf.

Heikkilä, Tuomas, and Teemu Roos. 'Quantitative Methods for the Analysis of Medieval Calendars'. *Digital Scholarship in the Humanities* 33, no. 4 (1 December 2018): 766–87. https://doi.org/10.1093/llc/fqy007.

Huglo, Michel. 'Comparaison du "Prototype" du couvent Saint-Jacques de Paris avec l'exemplaire personnel du maître de l'Ordre des Prêcheurs (Londres, British Library, Add. ms 23935)'. In *Aux origines de la liturgie dominicaine: le manuscrit Santa Sabina XIV L I,* edited by Leonard E. Boyle and Pierre-Marie Gy, 197–214. Documents, études et répertoires de l'Institut de Recherche et d'Histoire des Textes 67. Paris: CNRS, 2004.

Hülsmann, Margriet. 'De wisselende samenstelling van de Utrechtse heiligenkalender: een onderzoek naar de taakverdeling bij het afschrijven en het decoreren van handschriften'. In *Boeken in de late middeleeuwen: verslag van de Groningse Codicologendagen 1992,* edited by Jos M. M. Hermans and Klaas van der Hoek, 53–64. Boekhistorische reeks 1. Groningen: E. Forsten, 1994.

— 'Text Variants in the Utrecht Calendar. A Help in Localizing Dutch Books of Hours'. In *Masters and Miniatures: Proceedings of the Congress on Medieval Manuscript Illumination in the Northern Netherlands (Utrecht, 10–13 December 1989),* edited by K. van der Horst and Johann-Christian Klamt. Studies and Facsimilies of Netherlandish Illuminated Manuscripts 3. Doornspijk: Davaco, 1991.

— 'Variation in Page Layout in Dutch Books of Hours as Evidence for Localization: Developing a Method for Qualitative and Quantitative Research'. In *Manuscripts in Transition. Recycling Manuscripts, Texts, and Images,* edited by Brigitte Dekeyzer and Jan van der Stock, 291–99. Corpus of Illuminated Manuscripts 15. Leuven: Peeters, 2005.

Julien, Octave. 'Délier, lire et relier'. *Hypotheses* 19 (2016): 211–24.

Korteweg, Anne S. 'Books of Hours from the Northern Netherlands Reconsidered: The Uses of Utrecht and Windesheim and Geert Grote's Role as a Translator'. In *Books of Hours Reconsidered,* edited by Sandra Hindman and James H. Marrow, 235–61. Studies in Medieval and Renaissance Art History. London: Harvey Miller Publishers/Turnhout: Brepols, 2013.

Kren, Thomas. 'Seven Illuminated Books of Hours Written by the Parisian Scribe Jean Dubreuil, c. 1475–1485'. In *Reading Texts and Images. Essays on Medieval and Renaissance Art and Patronage in Honour of Margaret M. Manion,* edited by Bernard J. Muir, 157–200. Exeter Medieval Texts and Studies. Exeter: University of Exeter Press, 2002.

Laporte, Maurice. *Aux sources de la vie cartusienne. Dl. 4/2: Sources des Consuetudines Cartusiae [15–41].* In Domo Cartusiae [Grande Chartreuse], 1967.

Lebeuf, Jean. *Mémoires concernant l'histoire civile et ecclésiastique d'Auxerre et de son diocèse.* Edited by Maximilien Quantin. Auxerre/Paris: Perriquet, Dumoulin, 1848.

— *Mémoires concernant l'histoire ecclésiastique et civile d'Auxerre, tome premier.* Paris: Durand, 1743.

Leroquais, Victor. *Les Livres d'heures manuscrits de la Bibliothèque nationale.* 2 vols. Paris, 1927.

— *Les sacramentaires et les missels manuscrits des bibliothèques publiques de France.* Paris, 1924.

— 'Paris, Bibl. nat. de France, NAL 3162. Répertoires bibliques, liturgiques et iconographiques du chanoine Victor Leroquais. VI Hymnes, antiennes, psaumes, leçons, répons et capitules des Heures de la Vierge.' Manuscript. https://gallica.bnf.fr/ark:/12148/btv1b10033474t.

Macks, Aaron. 'CoKL: Corpus Kalendarium'. CoKL: Corpus Kalendarium. http://www.cokldb.org/.

Madan, Falconer. 'Hours of the Virgin Mary (Tests for Localization)'. *Bodleian Quarterly Record* 3, no. 2 (1920): 40–44.

— 'The Localization of Manuscripts'. In *Essays in History, Presented to Reginald Lane Poole,* edited by H. W. C. Davis, 5–29. Oxford: Clarendon Press, 1927.

Marrow, James H. 'Notes on the Liturgical "Use" of the Hours of the Virgin in the Low Countries'. In *Manuscripten en miniaturen. Studies aangeboden aan Anne S. Korteweg bij haar afscheid van de Koninklijke Bibliotheek,* edited by Jos Biemans, Klaas van der Hoek, Kathryn Rudy, and Ed van der Vlist, 279–94. Bijdragen tot de geschiedenis van de Nederlandse boekhandel. Nieuwe reeks 8. Zutphen: Walburg Pers, 2007. https://www.academia.edu/38270677/Notes_on_the_Liturgical_Use_of_the_Hours_of_the_Virgin_in_the_Low_Countries.

Myking, Synnøve Midtbø. 'Les livres français dans la Norvège médiévale: connexions culturelles, religieuses et sociales'. *Pecia* 23 (2020): 163–213. https://doi.org/10.1484/J.PECIA.5.123974.

— 'Ter Doest, Lund, and the *Legendarium Flandrense:* Danish-Flemish Connections in the Late Twelfth Century'. *The Journal of Medieval Latin* 28 (2018): 115–40. https://doi.org/10.1484/J.JML.5.116029.

Ottosen, Knud. 'Responsories of the Latin Office of the Dead'. Cantus planus, 2008–2014. http://www-app.uni-regensburg.de/Fakultaeten/PKGG/Musikwissenschaft/Cantus/Ottosen/index.html.

— *The Responsories and Versicles of the Latin Office of the Dead.* Aarhus: Aarhus university press, 1993.

Plummer, John. '"Use" and "beyond Use"'. In *Time Sanctified: The Book of Hours in Medieval Art and Life,* edited by Roger S. Wieck, 149–52. New York: G. Braziller, 1988.

Porter, Dorothy. 'Books of Hours as Transformative Works'. *Dot Porter Digital* (blog), 18 November 2019. http://www.dotporterdigital.org/books-of-hours-as-transformative-works/.

Fernández Riva, Gustavo. 'Network Analysis of Medieval Manuscript Transmission'. *Journal of Historical Network Research* 3 (19 November 2019): 30–49. https://doi.org/10.25517/jhnr.v3i1.61.

Rogers, Nicholas John. 'Patrons and Purchasers: Evidence for the Original Owners of Books of Hours Produced in the Low Countries for the English Market'. In *'Als Ich Can': Liber Amicorum in Memory of Professor Dr. Maurits Smeyers,* edited by Bert Cardon, Jan van der Stock, Dominique Vanwijnsberghe, and Katharina Smeyers, 1165–81. Low Countries Series. Leuven: Peeters, 2002.

Rudy, Kathryn M. *Piety in Pieces: How Medieval Readers Customized Their Manuscripts.* Cambridge: Open Book Publishers, 2016. https://directory.doabooks.org/handle/20.500.12854/26848.

— *Rubrics, Images and Indulgences in Late Medieval Netherlandish Manuscripts.* Leiden: Brill, 2016.

Sot, Michel, ed. *Les gestes des évêques d'Auxerre, Tome III.* Les classiques de l'histoire de France au Moyen âge 47. Paris: Les Belles Lettres, 2009.

Stutzmann, Dominique. 'HORAE – Hours: Recognition, Analysis, Editions'. HORAE – Hours: Recognition, Analysis, Editions, 2020. https://heurist.huma-num.fr/heurist/?db=stutzmann_horae&website.

Urfels-Capot, Anne-Élisabeth. 'Le sanctoral du lectionnaire de l'office'. In *Aux origines de la liturgie dominicaine: le manuscrit Santa Sabina XIV L I,* edited by Leonard E. Boyle and Pierre-Marie Gy, 319–53. Documents, études et répertoires de l'Institut de Recherche et d'Histoire des Textes 67. Paris: CNRS, 2004.

— *Le sanctoral du lectionnaire de l'Office dominicain, 1254–1256: édition et étude d'après le ms Rome, Sainte-Sabine XIV L1 'Ecclesiasticum officium secundum ordinem fratrum praedicatorum'.* Mémoires et documents de l'École des Chartes 84. Paris: École des chartes, 2007.

Vanwijnsberghe, Dominique. 'L'apport du texte et ses éléments de contenu'. In *Les Heures Tavernier. KBR, Ms. IV 1290,* edited by Bernard Bousmanne, Pierre Cockshaw, Georges Colin, Frédérique Johan, Claude Sorgeloos, Céline Van Hoorebeeck, and Dominique Vanwijnsberghe, 41–50. Brussels, 2002.

— 'Le cycle de l'Enfance des petites heures de la Vierge dans les livres d'heures des Pays-Bas méridionaux'. In *Manuscripten en miniaturen: Studies aangeboden aan Anne S. Korteweg bij haar afscheid van de Koninklijke Bibliotheek,* edited by Jos Biemans, Klaas van der Hoek, Kathryn Rudy, and Ed van der Vlist, 355–66. Bijdragen tot de Geschiedenis van de Nederlandse Boekhandel. Nieuwe Reeks 8. Zutphen: Walburg Pers, 2007.

— 'The Cyclical Illustrations in the Little Hours of the Virgin in Pre-Eyckian Manuscripts'. In *Flanders in a European Perspective. Manuscript Illumination around 1400 in Flanders and Abroad, Proceedings of the International Colloquium, Leuven, 7–10 September 1993,* edited by Maurits Smeyers and Bert Cardon, 267–96. Corpus of Illuminated Manuscripts 8. Leuven, 1995.

Wieck, Roger S. *Time sanctified: the Book of Hours in medieval art and life.* New York: G. Braziller, 1988.
Wijsman, Hanno. *Luxury bound: illustrated manuscript production and noble and princely book ownership in the Burgundian Netherlands, 1400–1550*. Turnhout: Brepols, 2010.

The Journal of HISTORICAL NETWORK RESEARCH

INA SERIF

From Networks of Texts to Networks of Topics?

On the Classification of (Texts in) Compilations with a View towards Manuscript Transmission

Journal of Historical Network Research 9 (2023) 184–213

Keywords manuscripts, topic modeling, shared manuscript transmission, digital history, medieval studies, Jakob Twinger von Königshofen

Abstract As medieval manuscripts often consist of more than one text, the application of network analysis can show textual connections between codices and therefore shed light on the circulation of texts, of manuscripts, and thus of knowledge. However, a text-based analysis often faces difficulties resulting from insufficient manuscript descriptions and a lack of normalization of work titles. A broader view, which would compare not particular texts, but rather genres, areas of interest or fields of knowledge, may help to circumvent these problems; however, this broader approach must deal with problems regarding classification. Instead of finding connections between subjectively classified texts, one can make use of topic modeling as a means to computationally classify, and thus characterize, multiple-text manuscripts. On the basis of automatically detected topics, topic-based networks can be generated. The current potential of such an analysis was tested using a sample of codices that contain the late medieval chronicle of Jakob Twinger von Königshofen. Advancements in text recognition and normalization of non-standardized spelling could further enhance this method to investigate the connections between the codices of a specific corpus and develop a better understanding of the copying and transmission of premodern manuscripts.

C²DH LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

UNIVERSITY OF LUXEMBOURG

I. Introduction

Researching the textual transmission of a medieval work is often arduous, as it requires dealing with different textual manifestations.[1] A stemmatological approach tries to follow copying processes in order to connect textual witnesses to each other and classify them in relation to an (often imagined or constructed) original.[2] But such a text-based approach has its limits. Texts that could be classified as manifestations of a work are not always identified as such, for various reasons: existing manuscript descriptions may qualify it as unique, due to a lack of knowledge of the describer, and provide it with a new title (or no title at all); modifications during the copying process, like dialectal adaptations, may have altered a text so significantly that connections to related versions are blurred; the layout may also fail to signal the start of a new textual unit, making it hard to identify a certain segment of a manuscript as the manifestation of a specific work. Even if we had the full, searchable text for all handwritten codices available, a textual comparison would be extremely difficult due to the variance of medieval manuscript culture.[3] Considering that the majority of medieval manu-

Acknowledgements: Apart from the credit that goes to Evina Stein and Gustavo Riva for organizing the conference and planning this volume, special thanks go to Gustavo Riva for discussing my sometimes rather unstructured thoughts at different stages of writing this paper. I would also like to thank the two anonymous reviewers for their close reading of the paper and their very helpful comments.

Corresponding author: Ina Serif, ina.serif@unibas.ch

1 In the absence of a conclusive definition of the term 'work', I refer to it as a virtual denominator for a concrete textual representation, but without the intention to value the single text as 'better' or 'closer' to an imagined 'original' in the tradition of Karl Lachmann. For a discussion of the terms 'work' and 'text' and a convincing way out of the Lachmannian idealization of the work or 'Urtext', see Tjamke Snijders, "Work, Version, Text and Scriptum: High Medieval Manuscript Terminology in the Aftermath of the New Philology," *Digital Philology. A Journal of Medieval Cultures* 2, no. 2 (2013): 266–96. She modifies the term 'scriptum', coined by John Dagenais, making it a description of a "material unity of text, layout, and codicology" that can be related to others: "[...] it becomes possible (though not always necessary) to judge them [scripta, I. S.] as relatively similar to one another on a textual level (as variants or even attempted copies) or as more profoundly or characteristically different from one another on the textual plane (versions)." Ibid., 279–280, 285.

2 An attempt to connect manuscripts without necessarily linking them to a root (an original) after normalizing dialectal variance in the transmission of the *Parzival* can be found in Michael Stolz, "Linking the Variance. Unrooted Trees and Networks," in *The Evolution of Texts. Confronting Stemmatological and Genetical Methods. Proceedings of the International Workshop Held in Louvain-La Neuve on September 1–2, 2004,* ed. Caroline Macé, Linguistica Computazionale 24/25 (Pisa, 2006), 193–213.

3 See Bernard Cerquiglini, *Éloge de la variante. Histoire critique de la Philologie* (Paris, 1989); Stephen G. Nichols, "What is a Manuscript Culture? Technologies of the Manuscript Matrix," in *The Medieval Manuscript Book: Cultural Approaches,* ed. Michael Johnston and Michael Van Dussen, Cambridge Studies in Medieval Literature 94 (Cambridge, 2015), 34–59.

scripts contain more than one text,[4] one could try to balance this shortcoming by tracing shared transmissions. By following the copying processes of not just one work but several, or finding connections between 'scripta'[5] (and therefore connections between manuscripts through more than one text), one could for instance discover previously unknown textual testimonies for the researched work in question – this would mean more texts that are used for comparison, which could compensate for imperfect or incomplete manuscript descriptions.[6] Particularly short texts that are not clearly separated in a manuscript copy, but that occur together with other texts, could be traced more easily. However, this approach is also constrained, simply because of temporal capacities.[7] So, instead of relying on attributed work titles, a widening of the perspective on a codex and its contents could make up for inconsistencies or incompleteness. Such a broadened perspective would classify a codex, or rather its content, with additional properties, making it comparable with others on a non-textual level by adding additional metadata, such as the genre(s) of the containing texts. This would enable the exploration of another level of connection between manuscripts, and for the creation of networks between them based on this new metadata. In the following, different ways of codex classification and their advantages and shortcomings will be discussed, using the manuscript transmission of the late medieval German chronicle by Jakob Twinger von Königshofen. I will discuss human created classifiers as well as computational classifiers, each with a view towards the potentials of network analyses based on the specific classifiers. First, networks based on connections between texts will be examined, followed by an analysis of networks between genres. The application of topic modeling will then be proposed as a starting point for a topic-based network, discussing it as one option to detect manuscript networks based on themes/topics. First results and the extended applications of the method are discussed, which could lead to further insights into transmission processes of medieval miscellanies and, connected to these, of knowledge.

---

4 See for example Sarah Westphal-Wihl, *Textual Poetics of German Manuscripts, 1300–1500,* Studies in German Literature, Linguistics, and Culture (Columbia, SC, 1993); Michael Johnston and Michael Van Dussen, "Introduction: Manuscripts and Cultural History," in *The Medieval Manuscript Book. Cultural Approaches,* ed. Michael Johnston and Michael Van Dussen, Cambridge Studies in Medieval Literature 94 (Cambridge, 2015), 2–16; Nichols, "What Is a Manuscript Culture? Technologies of the Manuscript Matrix."

5 Following the terminology of Snijders, see note 1.

6 Differing titles for the same work at times makes identification difficult.

7 Tracing shared transmissions is potentially infinite: "Jede Mitüberlieferung einer Handschrift eröffnet eine eigene Textgeschichte, die wiederum häufig mit anderen Textgeschichten anderer Texte in dieser Handschrift verbunden sein kann." Freimut Löser, "Überlieferungsgeschichte(n) schreiben," in *Überlieferungsgeschichte transdisziplinär: Neue Perspektiven auf ein germanistisches Forschungsparadigma,* ed. Dorothea Klein, Horst Brunner, and Freimut Löser, Wissensliteratur im Mittelalter 52 (Wiesbaden, 2016), 15.

2. Networks of Texts

During my research on the German chronicle written by the Strasbourg cleric Jakob Twinger von Königshofen at the end of the fifteenth century,[8] I was confronted with a case of a particularly complex manuscript transmission.[9] Up to today, 128 manuscripts are known that contain the chronicle, wholly or in parts, and that were produced not only in Strasbourg, Twinger's home town, but as far away as Cologne, Augsburg, and Tyrol.[10] Around thirty of the manuscripts qualify as true, unedited copies, while in the large majority of the witnesses, the text differs in various ways:[11] abbreviated, augmented, updated, corrected, put in a different order, and more often than not combined with other texts, either with distinct boundaries marked by layout, headings, etc., or resulting in new compositions composed of several texts, where two or more were combined into new entities.[12]

The chronicle consists of six chapters, of which the last is an extensive index. While the first three chapters depict universal history,[13] chapters four and five narrate the past of the diocese and of the city of Strasbourg. The chronicle therefore covers many different interests and subjects: world history, the histories of secular and ecclesiastical rulers, as well as regional, diocesan and urban history. Furthermore, it contains a compressed knowledge tool in the form of the aforementioned index. Hence it is not very surprising that complete copies of the chronicle were often not necessary, or asked for, but that in some codices only the universal history was copied, for example – to serve as a basis for the insertion

8 Carl Hegel, ed., *Die Chroniken der oberrheinischen Städte. Straßburg 1,* Die Chroniken der deutschen Städte vom 14. bis ins 16. Jahrhundert 8 (Leipzig, 1870), 230–498; Carl Hegel, ed., *Die Chroniken der oberrheinischen Städte. Straßburg 2,* Die Chroniken der deutschen Städte vom 14. bis ins 16. Jahrhundert 9 (Leipzig, 1871).

9 See Ina Serif, *Geschichte aus der Stadt. Überlieferung und Aneignungsformen der deutschen Chronik Jakob Twingers von Königshofen,* Kulturtopographie des alemannischen Raums 11 (Berlin/Boston, 2020).

10 For an up-to-date list, see Ina Serif, "Der zerstreute Chronist. Zur Überlieferung der deutschsprachigen Chronik Jakob Twingers von Königshofen," *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte,* May 12, 2015, https://mittelalter.hypotheses.org/7063, last updated June 28, 2023. The entry in the medieval manuscript database *Handschriftencensus* currently records 115 entries, see https://handschriftencensus.de/werke/1906 [last accessed December 21, 2022].

11 Carl Hegel, who edited the chronicle in 1869/70, divided the transmission into three versions, A, B, and C, and based his edition on C, a version uniquely found in Twinger's autograph that burnt in 1870 in the Strasbourg library. For a discussion of Hegel's editorial decision regarding the prevalence of versions A and B in the existing manuscripts, see Serif, *Geschichte aus der Stadt,* 27–32.

12 Or 'scripta', following the terminology of Snijders, see note 1.

13 The first chapter spans from the Creation to Alexander the Great, the second and third give an account of the history of the Roman emperors, beginning with Caesar, and of the popes, starting with Peter.

of historiographical accounts of another town, substituting chapters four and five with local chronicles or annals.[14]

One way of tracing the transmission of this particular work is a text-based analysis. The co-occurrence of certain texts in several manuscripts hints towards intentional copying processes that reflect specific interests, not only of one individual scribe or commissioner. Detecting and tracing these occurrences can tell us more about reading interests and habits. Multiple occurrences of particular combinations can reveal connections that would go unseen if the content of a codex were not regarded as a whole, and it can tell us more about the migration of manuscripts. But apart from the pitfalls of working with manuscript descriptions mentioned above, a text-based analysis can face other difficulties. With respect to the transmission of the Twinger chronicle, the problem lies primarily in the sheer amount of testimonies. I built a database from existing manuscript descriptions, as well as from my own examinations, which contains the basic codicological information for all known textual witnesses – physical properties like writing surface, number of pages, and the dimensions of the codex, as well as its contents.[15] In the context of the transmission of the Twinger chronicle, nearly 500 different, distinct texts were identified within the 128 manuscripts, most of which appear only once in the corpus, with a few co-occurring in several codices. An attempt to analyse the patterns of textual transmission is rather unhelpful in this case, given the unique appearance of many texts (see Fig. 1).

In terms of concrete numbers, there are 437 single appearances of 489 texts, making up 89% of the total. This high percentage is to be expected and can be partially explained by the incompleteness of the data, due to the incoherence of the available manuscript descriptions mentioned above, and because not all codices could be analysed extensively to differentiate entries like "various prayers" or "several poems" into their constituent parts. If these could be split and potentially connected to other texts in the corpus, this would probably not change the general tendency, but rather point towards hitherto unknown connections in the manuscript transmission.

However, despite the impression of a lack of connections, some smaller clusters can be detected that share more than one text. The so-called *Konstanzer Jahrgeschichten* may serve as an example to illustrate potential insights, as

14 Good examples are the codices Freiburg im Breisgau, Universitätsbibliothek, Hs. 471 and Cologne, Historisches Archiv, Best. 7030 22. For a codicological overview, see https://handschriftencensus.de/13868 and https://handschriftencensus.de/12948, with further literature.

15 The data can be downloaded from Zenodo: https://doi.org/10.5281/zenodo.7469112 or via the GitHub repository: https://github.com/wissen-ist-acht/twinger_chronicle_mss.

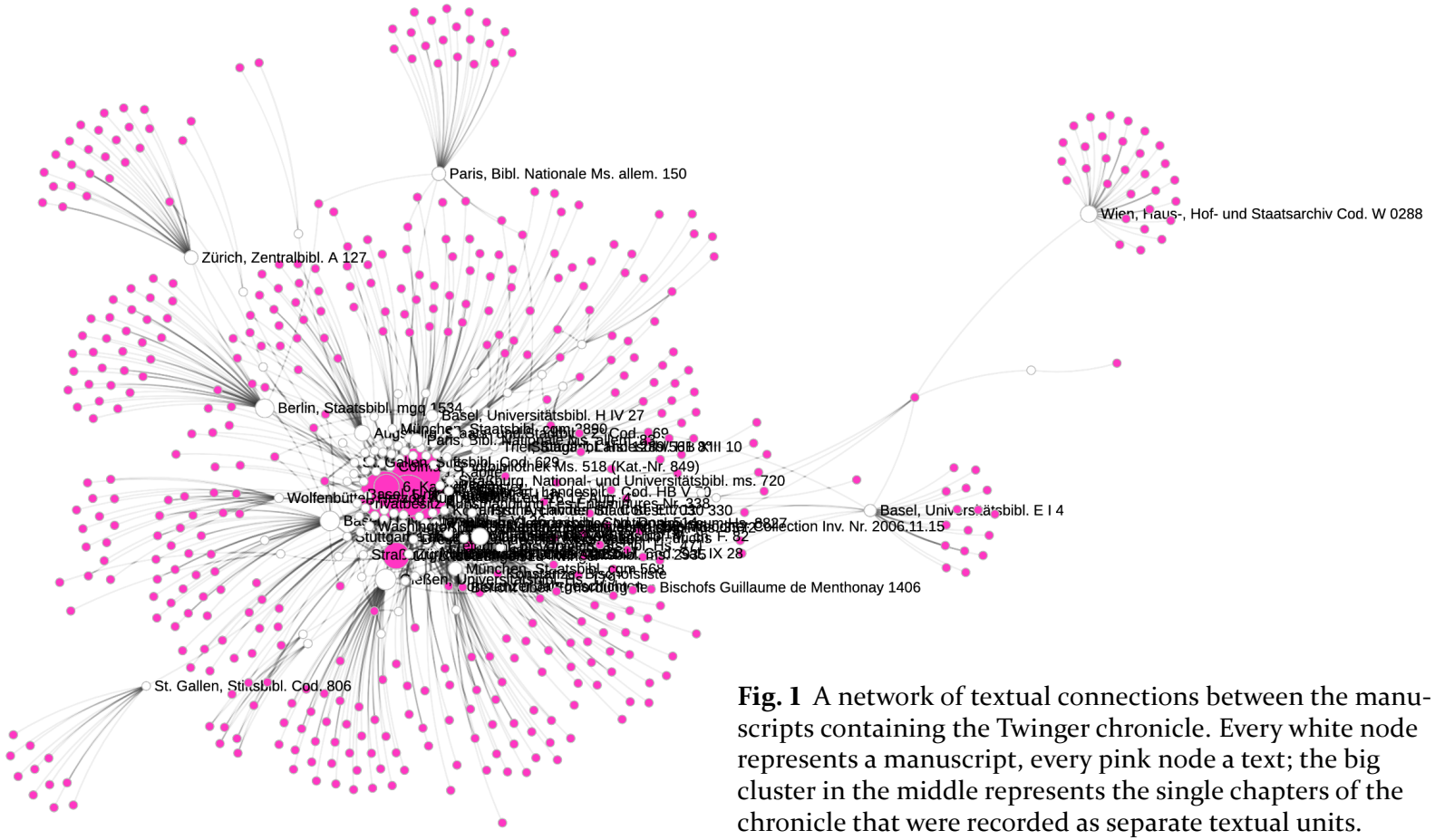

Fig. 1 A network of textual connections between the manuscripts containing the Twinger chronicle. Every white node represents a manuscript, every pink node a text; the big cluster in the middle represents the single chapters of the chronicle that were recorded as separate textual units.

eISSN: 2535-8863
DOI: 10.25517/jhnr.v9i1.136

well as the pitfalls of constructing a purely text-based network. The *Jahrgeschichten* are a short annalistic record in German consisting of notes, mainly about events in the city of Constance. The notes begin with the year 1256, when *bruoder Berchtold* preached in Constance for the very first time, and end, in most of the manuscripts, with 1388, reporting a huge fire in Constance and the neighbouring town of Stadelhofen. Twelve manuscripts that contain the *Jahrgeschichten* are known, eleven of which also contain the Twinger chronicle completely or in part.[16] A closer look at the manuscripts shows that they were not only produced in the city of Constance, where an interest for the *Jahrgeschichten* is rather self-evident, but also in Augsburg, some 200 kilometres northeast of it (see Fig. 2).[17]

Earlier research has tended to over-interpret the co-occurrence of this account with another historiographical text like the Twinger chronicle, leading to generalizations such as the following:

> Die in der Bischofsstadt Straßburg geschriebene Chonik hat in Konstanz ein breites Interesse gefunden. Es sind zehn [elf, I. S.] Handschriften bekannt, in denen an Twingers Text eigene Lokalnachrichten zur Konstanzer Geschichte angeschlossen wurden.[18]

16 Basel, UB, Cod. E VI 26; Dürnstein, Regularkanonikerstift, s.n. [now lost]; Freiburg im Breisgau, UB, Hs. 471; Gotha, FB, Cod. Chart. A 158; Heidelberg, UB, Cod. Sal. IX 28; Heidelberg, UB, Cpg 475; Karlsruhe, BLB, Cod. Don. 513; Munich, BSB, Cgm 567; Munich, BSB, Cgm 568; St. Gallen, Stiftsbibl., Cpg 630; Strasbourg, BNUS, ms. 5457. The codex Constance, Stadtarchiv, A I 1 only contains the *Jahrgeschichten,* without the passages from the Twinger chronicle. On the *Jahrgeschichten,* see Klaus Graf, "Die verschollene Twinger-Handschrift aus dem Regularkanonikerstift Dürnstein," *Archivalia,* March 27, 2013, https://archivalia.hypotheses.org/6941; Ina Serif, "Konstanzer Jahrgeschichten," in *Encyclopedia of the Medieval Chronicle, 2nd Online Edition,* ed. Graeme Dunphy and Christian Bratu (Leiden/Boston, 2016), http://dx.doi.org/10.1163/2213-2139_emc_SIM_001450. For the text of the single entries, see Franz Josef Mone, ed., *Quellensammlung der badischen Landesgeschichte 1* (Karlsruhe, 1848), 302–303; Franz Josef Mone, ed., *Quellensammlung der badischen Landesgeschichte 3* (Karlsruhe, 1863), 509; Gustav Scherrer, *Kleine Toggenburger Chroniken. Mit Beilagen und Erörterungen* (St. Gallen, 1874), 93–97.

17 The origin of two of the codices in Strasbourg and Basel is explained by manuscript migration: the *Jahrgeschichten* were added at a later time, after the codices had left their place of origin – another fact which complicates transmission analyses. For three manuscripts, we do not have enough evidence (yet) for a precise localization.

18 "The chronicle that was produced in the episcopal city of Strasbourg arouse interest in Constance. Ten [eleven, I. S.] manuscripts are known in which own local news concerning the history of Constance were inserted after Twinger's work." See Eugen Hillenbrand, "Gallus Öhem, Geschichtsschreiber der Abtei Reichenau und des Bistums Konstanz," in *Geschichtsschreibung und Geschichtsbewußtsein im späten Mittelalter,* ed. Hans Patze, Vorträge und Forschungen 31 (Sigmaringen, 1987), 734.

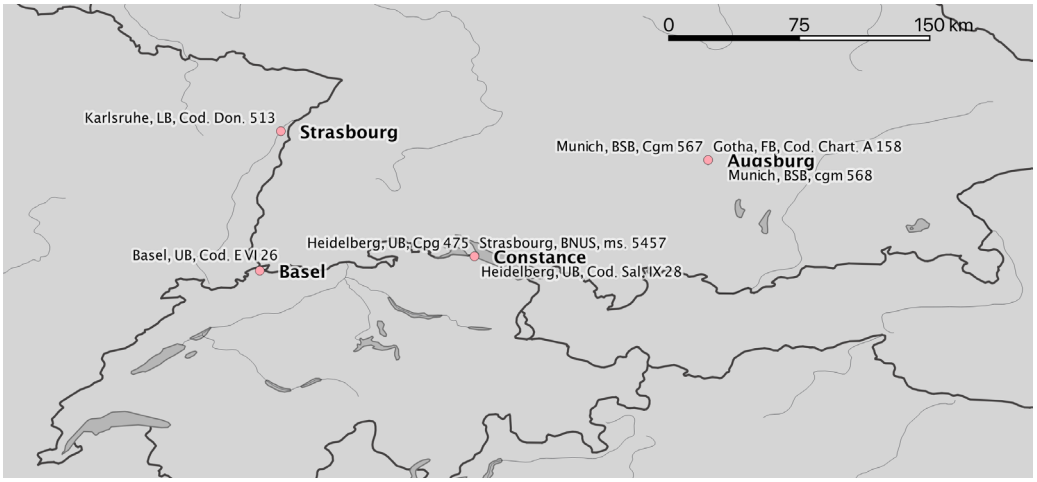

Fig. 2 Known places of production for eight of the manuscripts that contain the Twinger chronicle and the *Konstanzer Jahrgeschichten*.

When we look at the eleven codices in question, we see that they indeed share several texts – apart from the chapters of the Twinger chronicle (labelled here as *1. Kapitel, 2. Kapitel* and so forth) and the *Jahrgeschichten,* they include an account of the murder of the bishop of Lausanne, Guillaume of Menthonay, and a list of the bishops of Constance (Fig. 3).

Apart from the shared transmission, it is just as interesting to investigate which texts are *not* shared between the manuscripts, and to explore whether this allows for further assumptions or insights with respect to the compilation processes and manuscript migration. To be able to compare the different texts, I attributed tentative genres to them.[19] This reveals remarkable differences between some of the copies that are in need of explanation (Fig. 4).

While the codex Munich, BSB, Cgm 567 was probably in Hillenbrand's mind when he stated that the Twinger chronicle provokes interest in Constance and that local news concerning the history of the town were inserted after Twinger's

19 Genre attribution is always interpretive, and different levels of description are applied, such as structure or content. Several ontologies and references exist, all with their own advantages and disadvantages, e.g., the database *Geschichtsquellen des Mittelalters* that classifies every listed work, but without elaborating on the scheme, see https://www.geschichtsquellen.de/filter?filter=gattung. For my sample, I used 28 genres, without following a specific ontology: account, annals, chronicle, confession treatise, contract, didactic poem, directory, episcopal history, exemplum, family history, legend, letter, list, medical treatise, notes, novel, parody, poetry, proverb, reformatory account, regional history, royal legislation, treatise, universal history, urban history, vocabulary, and war history.

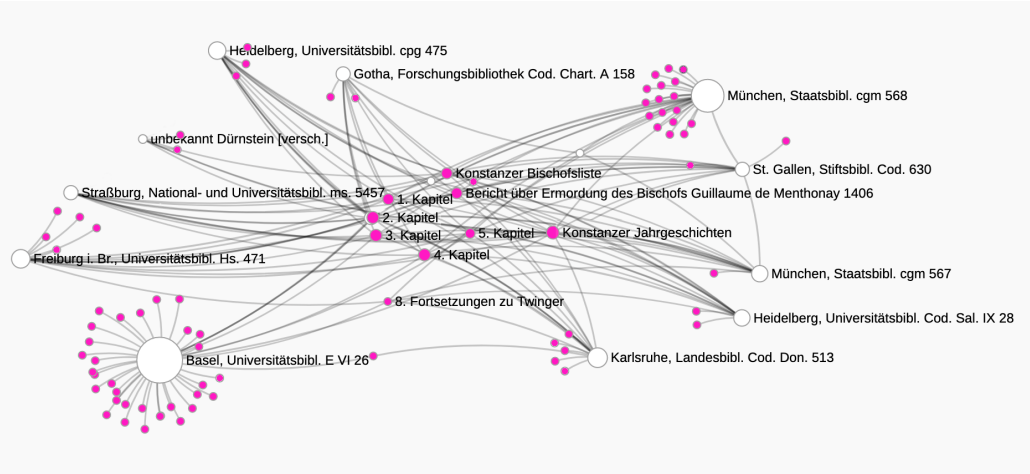

Fig. 3 Shared textual tradition in eleven manuscripts that contain, besides the Twinger chronicle, the *Konstanzer Jahrgeschichten,* a list of the bishops of Constance and an account of the murder of Guillaume of Menthonay, bishop of Lausanne, in 1406.

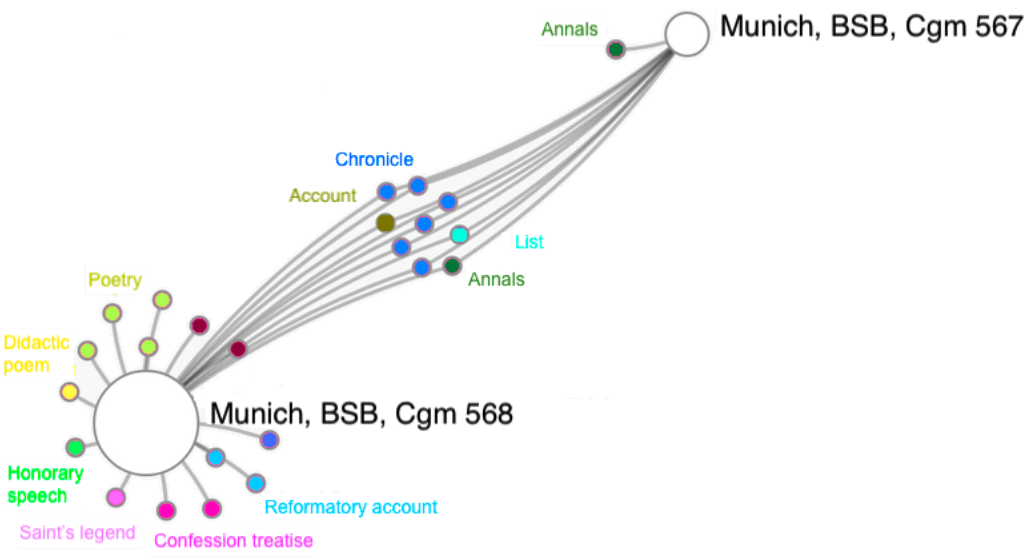

Fig. 4 Content of the codices Munich, BSB, Cgm 567 and Munich, BSB, Cgm 568, including genres.

work, the other compilation, Munich, BSB, Cgm 568, provides a different picture: apart from historiographical texts, we find poems, didactic poems, reformational accounts, confession treatises, and a legend of a saint.[20] A look at the actual texts reveals an interesting compilation in which universal and regional history was enhanced with religious and literary texts: after universal, regional and Strasbourg history, mediated by the Twinger chronicle, there follow two accounts about events of another city with the *Konstanzer Jahrgeschichten* and a list of the bishops of Constance. Afterwards, a list of the bishops of Augsburg is inserted, followed by the legend of St. Ulrich, one of the first bishops of Augsburg, leading to pieces with a religious focus, supplemented by prayers and treatises in Latin and German. The *Reformatio Sigismundi* and the *Reformatio Friderici* point towards the realm of literature, as do the poems of Thomas Prischuch and those of Jörg Zobel, added later. Thus, instead of a focus on the history of Constance, we find the opposite: the single texts mostly cover events in or persons from Augsburg – such as the list of bishops or the legend of the saint – or derive from Augsburg-based authors, like the poems of Thomas Prischuch. During the process of compilation, the original structure of the template,[21] namely the Twinger chronicle, the *Jahrgeschichten* and the list of bishops, was copied, but the addition of texts from quite different genres reduced the historiographical character of the collection. The broadening of the subjects also widened the potential readership, possibly resulting in an increased appeal of the production of compilations that were not purely historiographical in content. The question thus arises of whether a genre-based approach, and an attempt to construct networks from genres instead of texts, can help to better understand medieval manuscript production and make up for the insufficiencies of a text-based analysis.

3. Networks of Genres

Comparing manuscripts by looking at the genres of the texts they contain instead of the texts themselves can offer new insights, showing not only connections between manuscripts that would otherwise remain unseen, but also pointing towards interesting compilations from a programmatic point of view. There are two possible approaches to such a comparison: by attributing a genre to every text in a

20 For an overview of the codices, see https://handschriftencensus.de/9896 (Munich, BSB, Cgm 567) and https://handschriftencensus.de/6173 (Munich, BSB, Cgm 568).

21 For the identification of Munich, BSB, Cgm 567 as a template for Munich, BSB, Cgm 568, see Hegel, *Die Chroniken der oberrheinischen Städte. Straßburg 1,* 220. The main scribe is Johannes Erlinger, who may have produced the codex for himself; see Karin Schneider, "Berufs- und Amateurschreiber. Zum Laien-Schreibbetrieb im spätmittelalterlichen Augsburg," in *Literarisches Leben in Augsburg während des 15. Jahrhunderts,* ed. Johannes Janota and Werner Williams-Krapp, Studia Augustana 7 (Tübingen, 1995), 20–21. Apart from Erlinger's hand, there is a short addition by Konrad Bollstatter, a well-known Augsburg scribe, and the poems of Zobel were added at a later stage by an unknown hand.

manuscript, as has been shown for the Munich codices (see Fig. 4); or by classifying a manuscript as a unit, based on its textual content as a whole.

An attempt to apply the first approach on the entire sub-corpus of the eleven manuscripts that contain both the Twinger chronicle and the *Konstanzer Jahrgeschichten* shows some dominant genres (see Fig. 5).

If we compare this genre-based network with the text-based network above (Fig. 3), some compilations appear more coherent with regards to an underlying concept.[22] The codices Heidelberg, UB, cpg 475 (upper left), Freiburg im Breisgau, UB, Hs. 471 (on the left), and Karlsruhe, BLB, Cod. Don. 513 (lower right), for instance, seem to contain mainly historiographical texts, whereas Basel, UB, E VI 26 (lower left) and Munich, BSB, Cgm 568 (upper right) showcase a wider variety of genres. This kind of analysis seems to be possible for a small corpus, or miscellanies with few texts; it is not endlessly scalable, at least not so long as genre attribution is based on manual classification – not to mention the subjective nature of such a classification.

For an individual researcher, the second approach, classifying miscellanies as a whole, might be more feasible. The underlying hypothesis is that compilations were put together following some kind of concept, often combining works of the same genre. An exploratory analysis carried out by Gustavo Riva on the basis of 26,000 manuscripts containing Middle High German texts supports this assumption.[23] Working with the data of the *Handschriftencensus,* an inventory of the manuscript tradition of medieval German language texts, he constructed a network of shared manuscript transmission that shows clusters of texts that can be assigned to single genres. These clusters are of course fuzzy at the borders, but they show some broader tendencies, like the frequent combination of texts with similar genres in multiple-text manuscripts. This kind of analysis also shows that some texts are likely to fit into any kind of context, independent of the genre(s) of

22 We still lack a consistent terminology for manuscripts containing more than one text; while 'miscellany' is probably the least specific, terms like 'one-volume libraries' or 'multiple-text manuscripts' are in use, without clear definitions, and without referring to chronological aspects of the production, nor to structural or material characteristics. For recent reflections and studies see Michael Friedrich and Cosima Schwarke, eds., *One-Volume Libraries. Composite and Multiple-Text Manuscripts,* Studies in Manuscript Cultures 9 (Berlin; Boston, 2016); Marilena Maniaci, "Miscellaneous Reflections on the Complexity of Medieval Manuscripts," in *Collecting, Organizing and Transmitting Knowledge. Miscellanies in Late Medieval Europe,* ed. Sabrina Corbellini, Giovanna Murano, and Giacomo Signore, Bibliologia: Elementa ad Librorum Studia Pertinentia 49 (Turnhout, 2018), 11–22; Alessandro Bausi, Michael Friedrich, and Marilena Maniaci, eds., *The Emergence of Multiple-Text Manuscripts,* Studies in Manuscript Cultures 17 (Berlin/Boston, 2019).

23 Gustavo Fernández Riva, "Network Analysis of Medieval Manuscript Transmission. Basic Principles and Methods," *Journal of Historical Network Research* 3 (2019): 30–49.

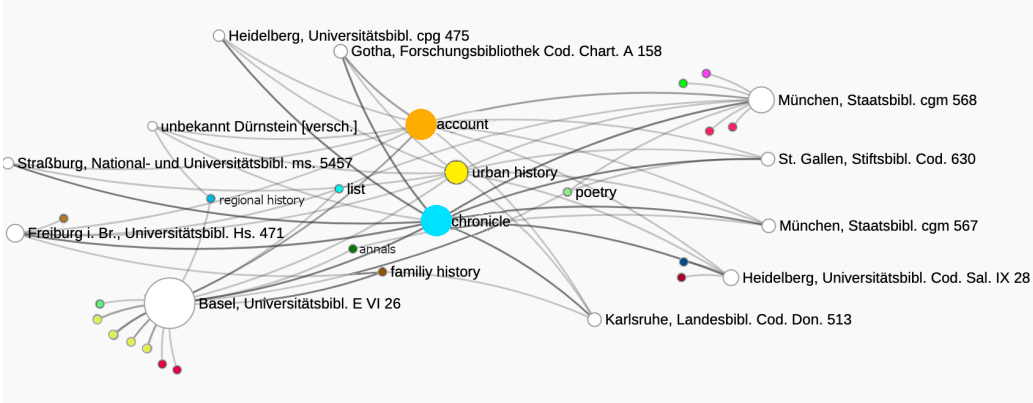

Fig. 5 Shared genre tradition in eleven manuscripts that contain, besides the Twinger chronicle (chronicle), the *Konstanzer Jahrgeschichten* (urban history), a list of the bishops of Constance (list), and an account of the murder of Guillaume of Menthonay, bishop of Lausanne, in 1406 (account).

the other texts.[24] But again, we encounter the question of classification: the attribution of a genre to a text is all but undisputed, making it simultaneously a productive yet hindering category.[25]

And while the classification of an (abstract) work or its textual manifestation is difficult enough, another difficulty arises with regards to the classification of miscellanies as a unit, a problem that has not yet been satisfyingly addressed.[26] More often than not, one main genre is chosen as a designator for the whole context, resulting in a "clash between genre study and codicology."[27] Miscellanies are more than the sum of their parts, but classifying them is very difficult – if not impossible[28] – and this is all the more true for compiled manuscripts that do not immediately appear homogeneous with respect to the (ascribed) genres of the

24 Ibid., 46.
25 Westphal-Wihl, *Textual Poetics of German Manuscripts, 1300–1500,* 8.
26 "But the vast majority of manuscripts have a miscellaneous character that defies the concept of genre as a principle of identity and separation." Ibid., 9.
27 Ibid. For an attempt to classify manuscripts by type to conduct a network analysis, see Octave Julien, "Délier, lire et relier. L'Utilisation de l'analyse réseau pour construire une typologie de recueils manuscrits de la fin du Moyen Âge," *Hypothèses* 19, no. 1 (2016): 211–24, doi:10.3917/hyp.151.0211. Julien groups the (mostly multi-text) manuscripts into ten categories: religion, moral literature, profane literature, history and politics, jurisprudence, practical texts, scientific texts, domestic and technical literature, and encyclopedias. He also refers to sub-categories, without further explanation.
28 "Sammelhandschriften sind mehr als die Summe der Einzelteile, aber sie in diesem integralen Sinn in den Blick zu nehmen, ist beinahe unmöglich." Jürgen Wolf, "Sammelhandschriften – mehr als die Summe ihrer Einzelteile," in *Überlieferungsgeschichte transdisziplinär: Neue Perspektiven auf ein germanistisches Forschungsparadigma,* ed.

containing texts or an (assumed) programmatic arrangement – as has been illustrated here for the codex Munich, BSB, Cgm 568.[29]

Thus, the questions arise: how should we deal with the codicological context of a work, and how can we make this context productive for analyses of production processes of reader's interests, of knowledge spreading, without perpetuating narrowing classifications, or replacing them with new constrictions? Is there something like a "unifying purpose"[30] of a compiled manuscript, and how could it be detected?[31]

4. Networks of Topics

As the manual classification of the content of compiled manuscripts poses several methodological problems, a computational approach may serve as an alternative. This can help to compensate for incomplete or inaccurate manuscript descriptions, in order to overcome the subjectiveness of classifications and allow scalability and therefore applicability for the different source corpora. The attempt to connect manuscripts on the basis of computationally calculated topics and its suitability for medieval miscellanies will be discussed in the following, again using codices from the Twinger chronicle as an example.

Dorothea Klein, Horst Brunner, and Freimut Löser, Wissensliteratur im Mittelalter 52 (Wiesbaden, 2016), 80.

29 "Das gilt besonders für die Textformationen, die sich nicht auf einen Nenner der Art 'enzyklopädische Sammlung', 'historiographische Kompilation', 'Liedersammlung', 'Bibelkompendium' oder 'Legendar' bringen lassen, und die als ein mehr oder weniger zufälliges Sammelsurium von Texten ohne bedeutungstragende Ordnungs- und Organisationsstruktur erscheinen. Es fehlen außerdem systematische Überlegungen zur Historisierung und zur Spezifik des Mediums 'Sammelhandschrift'. Statt auf generelle Funktionsweisen von Textzusammenstellungen wird der Fokus bislang auf einzelne Bücher und deren Rezipienten gerichtet, auf die Programmatik individueller Textsammlungen oder auf das historische Verständnis von Gattungen und Genres, das man aus den Textkombinationen meint ableiten zu können." See Diana Müller, *Textgemeinschaften. Der "Gregorius" Hartmanns von Aue in mittelalterlichen Sammelhandschriften* (Frankfurt a.M., 2013), 42, http://publikationen.ub.uni-frankfurt.de/frontdoor/index/index/docId/30069.

30 Stephen G. Nichols and Siegfried Wenzel, eds., "Introduction," in *The Whole Book: Cultural Perspectives on the Medieval Miscellany, Recentiores.* Later Latin Texts and Contexts (Michigan, 1996), 6.

31 The temporal aspect and the evolving character of manuscripts with regards to their content and structure, and therefore evolving "unifying purposes", are left out of this analysis; we still have to keep in mind that the "manuscript was in constant flux, always with the potential to be reshaped by its current owner." See Johnston and Van Dussen, "Introduction: Manuscripts and Cultural History," 5.

I chose topic modeling as a means to quantitatively approach texts.[32] Rather than merely counting word frequencies, like tf-idf,[33] the underlying idea behind topic modeling is that words that appear in the same context have the same, or a similar, meaning. Therefore, not only the frequency, but also the distribution of words in a document, or a corpus, is counted, using statistical methods. Depending on the distributions, topics are inferred, each of them consisting of a list of words that appear together in a statistically significant way. For my case study, I used the Dariah Topics Explorer,[34] software that is based on the statistical model Latent Dirichlet Allocation (LDA).[35] As this is a GUI tool, not all parameters that the model is based on can be changed, but it is a good starting point to determine rather quickly whether an analysis with topic modeling is a useful approach for a specific corpus. For any further analysis, I would recommend using programs that allow for complete control of all steps.[36]

---

32 For a useful overview of introductory texts, more technical articles and research projects that use topic modeling, see Scott B. Weingart, "Topic Modeling for Humanists: A Guided Tour," July 25, 2012, http://scottbot.net/topic-modeling-for-humanists-a-guided-tour/. Anne Purschwitz applied topic modeling to historical journals of the Enlightenment, attempting to discover (networks of) scholarly discourses, see Anne Purschwitz, "Netzwerke des Wissens – thematische und personelle Relationen innerhalb der Halleschen Zeitungen und Zeitschriften der Aufklärungsepoche (1688–1818)," *Journal of Historical Network Research* 2 (December 3, 2018): 109–42. For some critical remarks on the data basis of the construction of networks based on the result of topic modeling see Scott B. Weingart, "Topic Nets," November 10, 2012, http://scottbot.net/topic-nets/.

33 Term frequency–inverse document frequency. For a broader discussion of this statistical measure, see Stephen Robertson, "Understanding Inverse Document Frequency: On Theoretical Arguments for IDF," *Journal of Documentation* 60, no. 5 (January 1, 2004): 503–20, doi:10.1108/00220410410560582.

34 Available at https://github.com/DARIAH-DE/TopicsExplorer; Steffen Pielström, Severin Simmler, and Thorsten Vitt, "Topic Modeling with Interactive Visualizations in a GUI Tool," *Proceedings of the Digital Humanities Conference Utrecht 2019,* n.d., https://dev.clariah.nl/files/dh2019/boa/0637.html. For a short tutorial in German see Mareike Schuhmacher, "DARIAH Topics Explorer," *ForTEXT. Literatur digital erforschen,* accessed October 29, 2021, https://fortext.net/tools/tools/dariah-topics-explorer.

35 The model was introduced by David Blei, Andrew Ng and Michael I. Jordan for use in textual studies, see David M. Blei, Andrew Y. Ng, and Michael I. Jordan, "Latent Dirichlet Allocation," *The Journal of Machine Learning Research* 3 (March 1, 2003): 993–1022. David M. Blei, "Probabilistic Topic Models," *Communications of the ACM* 55, no. 4 (April 2012): 77–84. For a concise (and humanist approved) explanation of LDA and Gibbs sampling see Ted Underwood, "Topic modeling made just simple enough," July 4, 2012, https://tedunderwood.com/2012/04/07/topic-modeling-made-just-simple-enough/.

36 Gensim and Mallet are two programs that enable adjustment: Gensim is a Python library, while Mallet is based on Java (Dariah's Topics Explorer uses Mallet). A comparison of the results of the two programs, which use different sampling methods for different corpora, one of them the subset of Twinger manuscripts used here, can be found in Tobias Hodel, Dennis Möbus, and Ina Serif, "Von Inferenzen und Differenzen. Ein Vergleich von Topic-Modeling-Engines auf Grundlage historischer Korpora," in *Von Menschen und Maschinen. Mensch-Maschine-Interaktionen in digitalen Kulturen,* ed. Selin Gerlek et al. (Hagen 2022), 181–205. We are currently working on a follow-up paper that reflects in

The calculation of topics is based on the full text of a document, independent of text boundaries, how it has changed hands, or the different stages of production of each individual text. This allows for an analysis of the content of manuscripts without the need for prior manual inspection; this means that the method is potentially usable on very large corpora, given that their full text is provided in a machine-readable format.[37]

The corpus used for this proof of concept consists of seven multiple-text manuscripts, all containing the Twinger chronicle, three of which are part of the *Jahrgeschichten* corpus:[38] Dresden, UB, Mscr. F 98; Freiburg im Breisgau, UB, Hs. 471 (with *Jahrgeschichten*); Heidelberg, UB, Cpg 116; Heidelberg, UB, Cpg 475 (with *Jahrgeschichten*); Munich, BSB, Cgm 568 (with *Jahrgeschichten*); Stuttgart, LB, HB V 22; Wolfenbüttel, HAB, Cod. 16.17. I performed handwritten text recognition (HTR) on all seven manuscripts, using the software Transkribus and some of its generic, publicly available models.[39] I did not perform an elaborate post

---

detail on the methods and concepts used for topic modeling, such as the different steps of preprocessing – lower casing, removal of punctuation, chunking, etc. – and the adjustment of several parameters. Chunking in particular – the cutting of documents into equal parts – seems to compensate for the varying length of the medieval manuscripts.

37 Christof Schöch applied topic modeling to French Classic and Enlightenment literature and to texts of Arthur Conan Doyle respectively, attempting to model genre see Christof Schöch, "Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama," *Digital Humanities Quarterly* 11, no. 2 (2017); Christof Schöch, "Computational Genre Analysis," in *Digital Humanities for Literary Studies: Methods, Tools, and Practices,* ed. James O'Sullivan (College Station, TX, 2020), 219–31.

38 In a best-case scenario, the test corpus built for this trial would have consisted of the eleven manuscripts that share the Twinger chronicle and the *Konstanzer Jahrgeschichten,* because the comparison of the three different approaches discussed here would have been more exact. But some constraints led to a slightly different composition, mainly the lack of digital copies (Dürnstein, Regularkanonikerstift, s.n. [now lost]; Gotha, FB, Cod. Chart. A 158 [digitized very recently]; Munich, BSB, Cgm 567; St. Gallen, Stiftsbibl., Cpg 630; Strasbourg, BNUS, ms. 5457). Another reason was the rather unpromising results of handwritten text recognition (HTR) using existing models for certain manuscripts.

39 To perform the text recognition on the manuscripts, I received free credits from Transkribus, applying for their scholarship program (https://readcoop.eu/transkribus/scholarship) – thanks again! The models used were "German Kurrent XVI–XVIII M1", "Thun Missiven M3", "Medieval Scripts M2" and "Charter Scripts XIII–XV M4". (All of these are based on the CitLab HTR/HTR+ engine, which is no longer supported by Transkribus; while there are models based on the engine PyLaia that fit the different writings in the used manuscripts, the transcription output might differ from the output achieved with the other engine.) Playing around with the different models is worth the while, in particular with those covering longer time periods. The accuracy might suffer a bit, but especially for codices that consist of several texts by various hands, even some that are decades or centuries apart, in many cases there is no need to apply different models on different parts of the manuscript, so the time saved makes up for the lower accuracy.

processing on the resulting text, for two reasons.[40] Firstly, I would like to provide a proof of concept of this approach for larger corpora, and the necessity of lengthy HTR correction or other forms of data cleaning would preclude scalability.[41] Secondly, lemmatization or normalization runs the risk of altering the resulting topics in an unwanted manner: the peculiarities of premodern writing, without fixed spelling and dialectal variations, are already part of linguistic analyses to discover relationships between manuscripts; how this shows in a topic model will be discussed in a moment.[42]

To be able to evaluate the results on a slightly larger scale, the corpus was enlarged with some printed editions of several medieval texts, to see if the resulting topics differ from those with uncorrected texts. I also added the edition of the Twinger chronicle, to see how much the topics of the single manuscripts would be related to those of the edited text. Thus, in addition to the codices mentioned above, the printed text of the *Chronik der Eidgenossenschaft* by Petermann Etterlin, the *Oberrheinische Chronik,* the *Konzilschronik* by Ulrich Richental, the *Leben des heiligen Ulrich* by Albert von Augsburg, and the Twinger chronicle were also used for topic analysis, the latter being divided into two parts, the first containing chapters one, two and three, and the second with chapters four, five and six.[43] In none of the editions did any normalization take place, so the difference

40 Apart from stripping strings which contained information about the digitizing institution: many digitized images are marked with a copyright sentence, which mentions the holding library; this text was also recognized during recognition, but could be easily detected and deleted. Also, diacritics were dissolved: during text recognition, words with diacritics were often split into two strings, e.g., "brů" and "der"; dissolution resulted in one string, e.g., "bruoder".

41 For a meta study on the impact of OCR errors see Stephen Mutuvi et al., "Evaluating the Impact of OCR Errors on Topic Modeling," in *Maturity and Innovation in Digital Libraries,* ed. Milena Dobreva, Annika Hinze, and Maja Žumer, Lecture Notes in Computer Science (Cham, 2018), 3–14, doi:10.1007/978-3-030-04257-8_1. While there is a measurable impact of OCR errors on the output of topic model analyses, this impact is relatively small overall. Apparently, it doesn't affect the average coherence score between the models too much, Ibid., 12.

42 The data used for the analysis – the txt-files as well as the output of the topic modeling – are available at: https://github.com/wissen-ist-acht/tm_data.

43 Eugen Gruber, ed., *Petermann Etterlin. Kronica von der loblichen Eydtgnoschaft, Jr harkommen und sust seltzam stritten und geschichten,* Quellenwerk zur Entstehung der Schweizerischen Eidgenossenschaft 3, 3 (Aarau, 1965); Franz Karl Grieshaber, *Oberrheinische Chronik: Älteste bis jetzt bekannte, in deutscher Prosa* (Rastatt, 1850); Thomas Martin Buck, ed., *Chronik des Konstanzer Konzils 1414–1418 von Ulrich Richental: historisch-kritische Edition. Band 1: A-Version,* vol. XLIX, 1–3, Konstanzer Geschichts- und Rechtsquellen (Ostfildern, 2020); Karl-Ernst Geith, ed., *Albert von Augsburg: Das Leben des heiligen Ulrich,* Quellen und Forschungen zur Sprach- und Kulturgeschichte der germanischen Völker, n.F. 39 (163) (Berlin/New York, 1971); Hegel, *Die Chroniken der oberrheinischen Städte. Straßburg 1;* Hegel, *Die Chroniken der oberrheinischen Städte. Straßburg 2.* While the text of the *Konzilschronik* (Aulendorfer version), the *Kronica* and *Das Leben des heiligen Ulrich* were digitally available, I performed text

between the manuscripts lies mainly in the greater accuracy of the transcription. Some texts were already available as a text file, while for others I had to perform OCR on the digitized books. Regarding post processing, the same applies as for the manuscripts: I did not edit the text recognized, I merely removed the superfluous information derived from the digitization process.[44]

Usually, text analysis methods like topic modeling, where occurrences of tokens are counted, use a list of stop words, i.e., words that should be excluded from the analysis because they appear often, but do not carry much (semantic) meaning, yet possess syntactical or grammatical functions. Prebuilt lists exist, also for premodern languages; however, these still need to be adapted and enlarged upon, mainly because of different ways of spelling, like "und", "unde", "unnd", "vnnd", "unnt", etc. for "and". For my analysis, I used the stop word list for Middle High German, provided by the Classical Language Toolkit,[45] which I extended during analysis. Topic modeling is, like many other approaches to analysing texts, iterative, meaning that the findings of a first inspection can be used to improve the results – for example, the topics that were detected in a first round led to the exclusion of several words for the next round by adding them to the stop list (see Fig. 6).

After several rounds of modeling and exclusion of more stop words, increasing the amounts of topics from ten to twenty-five and the number of iterations of the model to 10,000, the results looked more nuanced (see Fig. 7).

Aside from a list of topics, the Topic Explorer also offers a document-topic-matrix that represents a network of topics. This shows the prevalence of a topic in a document using saturation: the lighter a field is, the less important is the topic, or the less common are the words of this specific topic within a document (see Fig. 8).

A closer look at this matrix provides us with three different kinds of results (Fig. 9): first, the fact that the fifth topic, "strosburg, stat, bischof" is more prevalent in the second part of the Twinger chronicle (namely chapters four and five) is not surprising from what we already know about the content, nor is the dominance of the twelfth topic "künig, bobest, rome" in the first part (chapters one to

---

recognition for the remaining texts. The results are slightly better for modern print, i.e., not Gothic type, but still very decent for the latter.

44 I also did not delete the critical apparatus. In all the editions used here, the quantity of the edited text was easily sufficient to outweigh these remainders.

45 This Python library performs natural language processing, especially for premodern languages. At the moment, it is available for nineteen languages. See cltk.org for the package and its documentation.

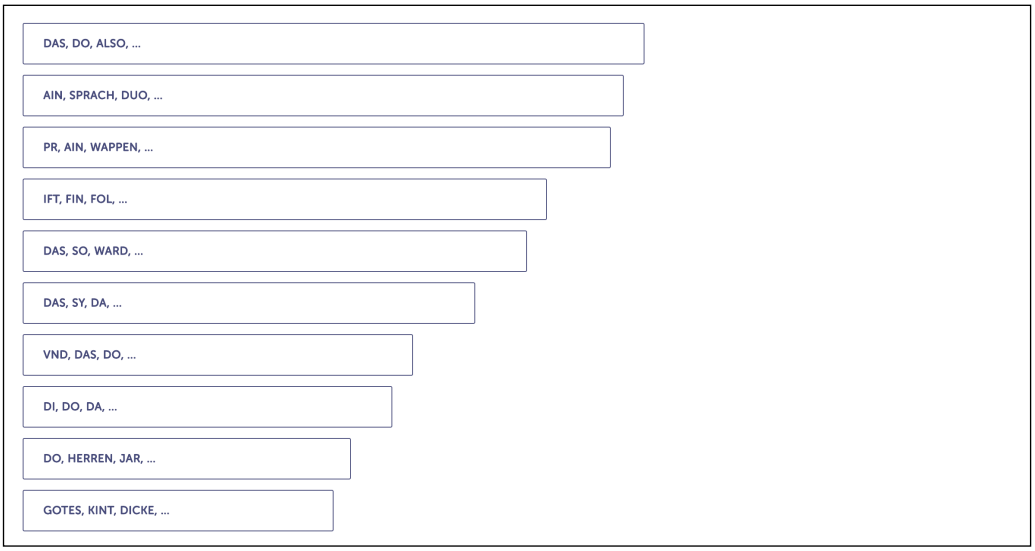

Fig. 6 List of ten topics, showing the three most common words for each topic (out of fifteen in total that are listed in the results). One side effect of omitting post processing is already visible, the incorrect recognition of "ſ" as "f" instead of "s".

Fig. 7 List of 25 topics, showing the three most common words for each topic. The skipped normalization and lemmatization are clearly visible, showing "kùng", "kùnig", "künig", "kung", "könig", "küning" as variations of today's "König" (king).

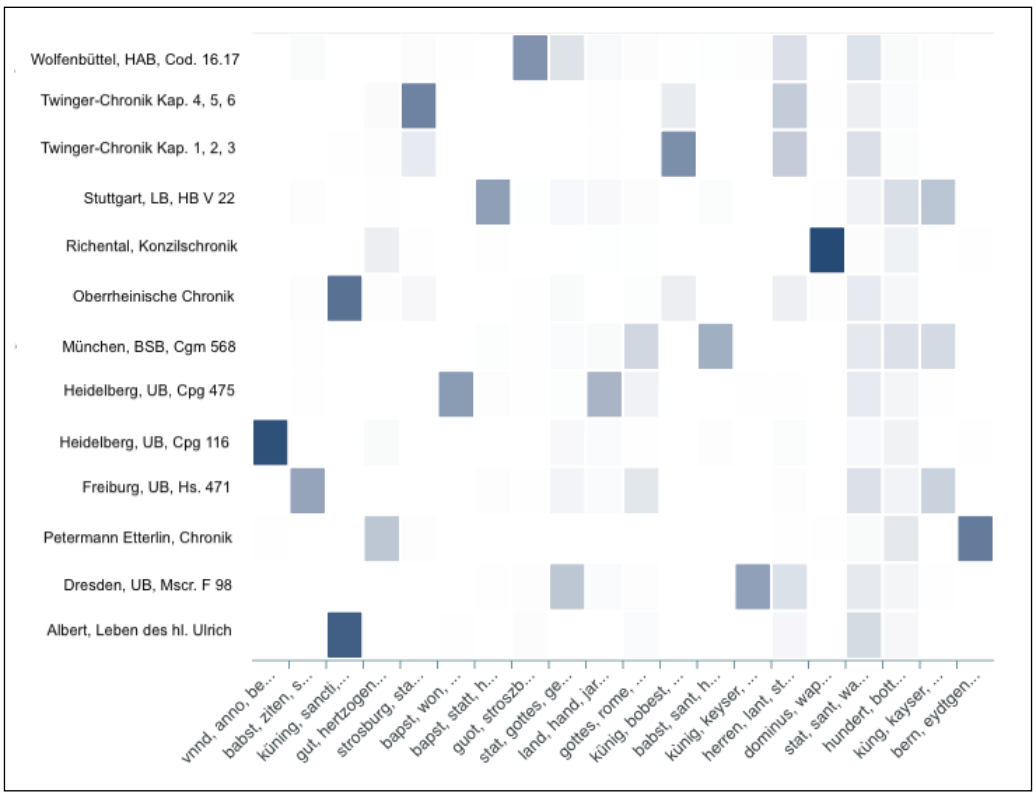

Fig. 8 Document-topic-matrix/topic network with weighed prevalence of each topic (13 documents, 20 topics).

three, green frames).[46] Here we can also see that topic modeling does not overly depend on lemmatization or "correct" spelling according to modern dictionaries. Second, we can confirm existing knowledge from a close reading of the manuscripts: the two codices Freiburg im Breisgau, UB, Hs. 471 and Munich, BSB, Cgm 568 share topic 11 with Heidelberg, UB, Cpg 475, and topic 19 with Stuttgart, LB, HB V 22 (orange frames). For the first group, we already know that the three miscellanies share several texts that deal with Constance, among them the *Konstanzer Jahrgeschichten.* In the second group, the Stuttgart codex does not contain the *Jahrgeschichten,* but two other texts that are concerned with the history of Constance: the *Konzilschronik* by Ulrich Richental and the *Konstanzer Chronik* by Gebhard Dacher. Here we get closer to the initial idea of identifying rela-

46 Unfortunately, there is no comprehensive visualization of the matrix with a list of words for the single topics. The overall trend should be visible, even through two or three words that are shown in the figure, and which I refer to in the text. A complete list of the words for each topic can be found in the repository: https://github.com/wissen-ist-acht/tm_data/.

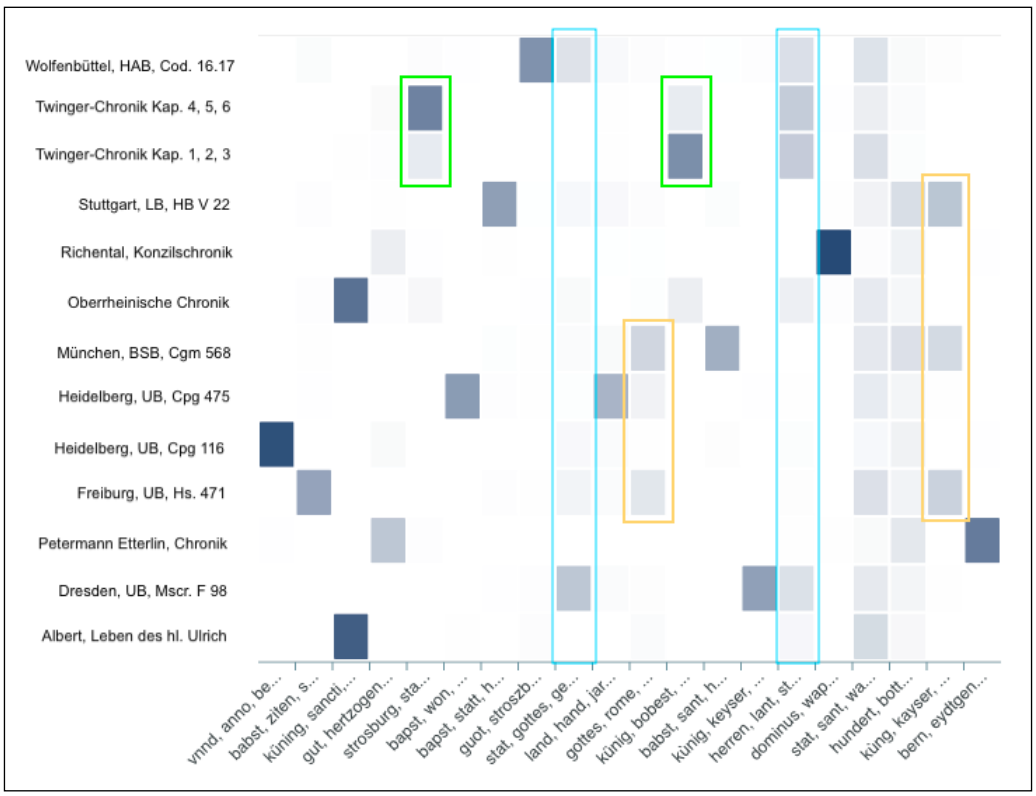

Fig. 9 Document-topic-matrix with weighed prevalence of a topic (13 documents, 20 topics), and special attention to topics five, nine, eleven, twelve, fifteen and nineteen.

tions between manuscripts through topics, without having to read them or rely on available descriptions. A third kind of finding is somewhat surprising: the two manuscripts Wolfenbüttel, HAB, Cod. 16.17 and Dresden, UB, Mscr. F 98 that share topics 9 and 15 do not have anything in common on a textual level – apart from the Twinger chronicle, of course (blue frames). The Dresden codex contains several texts that are concerned with the Burgundian War, whereas the Wolfenbüttel manuscript collects lyric, prayers, and cooking recipes.

The interpretability of such a matrix correlates with the number of documents. If we wanted to obtain a first impression of the relations between manuscripts in a much larger corpus, the visualization of the results as a network is helpful. However, for this small sample, we also get a nice impression of the connected codices (see Fig. 10).[47]

47 A CSV file of the document similarities as part of the Topics Explorer export was the basis for the network creation – many thanks go to Gustavo Riva for showing me how to

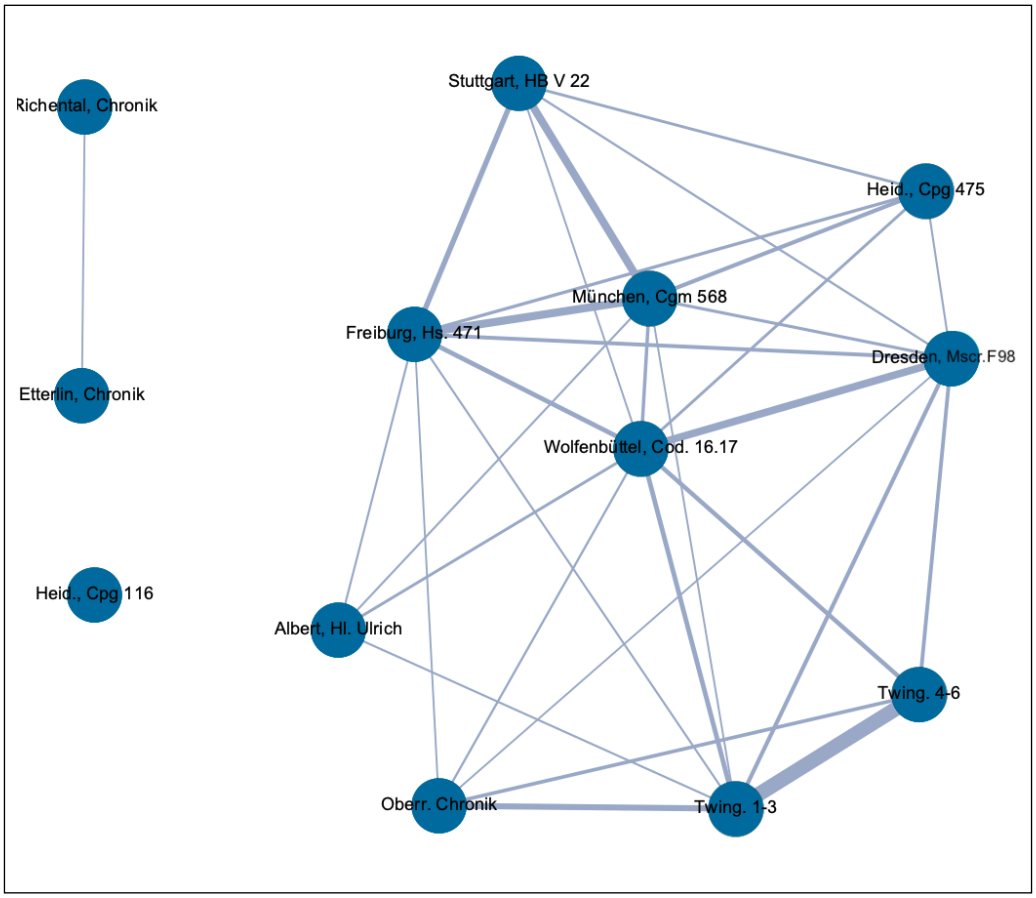

Fig. 10 Network of documents with weighed prevalence of a topic. Visualization created with Gephi.

As in the document-topic-matrix, we see stronger connections between the two parts of the Twinger chronicle (lower right), between the manuscripts with a focus on Constance (Freiburg im Breisgau, Stuttgart, Munich and Heidelberg, Cpg 475, at the top), as well as between the codices from Wolfenbüttel and Dresden (in the center). One result that is more visible in this kind of visualization is

do this! In this visualization, the edge weight ranges between 0.5 and 0.8. Higher and lower weights have been filtered out. Without this filter, a strong connection between the *Leben des heiligen Ulrich* and the *Oberrheinische Chronik* appears. If we consult the topics list, we see that they share only one topic, "küning, sas, sancti, ulrici, vita, herren". While "sancti", "ulrici" and "vita" only appear in the *Leben,* "küning", "sas" and "herren" only show up in the *Chronik.* The shared topic explains the visible connection, but apparently there is no textual/content based relation between the two documents. It seems that caution is advised whenever two documents are only connected through one single topic.

the isolation of the two edited chronicles of Richental and Etterlin as well as of the codex Heidelberg 116. In the document-topic-matrix, each of these has one very large topic and only very few shared with others. Codex Heidelberg, Cpg 116 is composed of chapter 6 of the Twinger chronicle, the *Weißenburger Chronik* by Eikhart Artzt, and the *Trotula* – texts that are unique in the corpus. That this results in them being outliers is very well demonstrated in the network.

So how can these findings be interpreted? The second result seemed promising, hinting towards the relation of manuscripts that contain the same text(s), or texts treating the same subject(s): they share the *Konstanzer Jahrgeschichten* or chronicles about the city of Constance. But this thematic focus is all but obvious from the actual words that make up the connecting topics: for the three manuscripts Freiburg im Breisgau, UB, Hs. 471, Munich, BSB, Cgm 568 and Heidelberg, UB, Cpg 475, the binding topic consists of "gottes, rome, geburt", and for Freiburg im Breisgau, UB, Hs. 471, Munich, BSB, Cgm 568 and Stuttgart, LB, HB V 22, the topic is made up of "kùng, kayser, volk", which does not in any way point towards the city of Constance. For the connecting topics of the codices Wolfenbüttel, HAB, Cod. 16.17 and Dresden, UB, Mscr. F 98, we can observe something similar: The topics "stat, gottes, geburt" and "herren, lant, starp" do not contain an easy-to-read hint at the connecting texts and/or topics in the particular manuscript. One explanation for the composition of the topics might be found in medieval writing practices: events are often dated referring to the birth of Christ, the ruling emperor, or the current pope, and as most of the manuscripts in this corpus contain historiography, there are many events that are contextualized with such a reference. One could exclude words like "gottes", "geburt", and all the different forms for king and pope, using the stop word list – but this would account for a bias that might already be too large, by eliminating words in order to get to the "real" meaningful terms – and eventually concepts.[48]

The topics generated here cannot serve as designators of manuscripts in a corpus with regards to their specific content. However, they do provide additional value in tracing the relationships between manuscripts and discovering networks: during copying processes, the linguistic peculiarities of the copied manuscript are often kept within direct adoptions of (parts of) the texts, leading to a fair consistency of spelling. Without normalization or lemmatization of the texts, the resulting topics actually point towards relations between manuscripts from a linguistic point of view. If we look again at the document-topic-matrix and compare the dialects that were assigned to the two codices Wolfenbüttel, HAB, Cod. 16.17 and Dresden, UB, Mscr. F 98, which to our knowledge do not share any texts nor treat similar subjects, this assumption can be confirmed (see Fig. 11).

48 For a discussion of the influence of stop word lists, see Hodel, Möbus, and Serif, "Von Inferenzen und Differenzen.

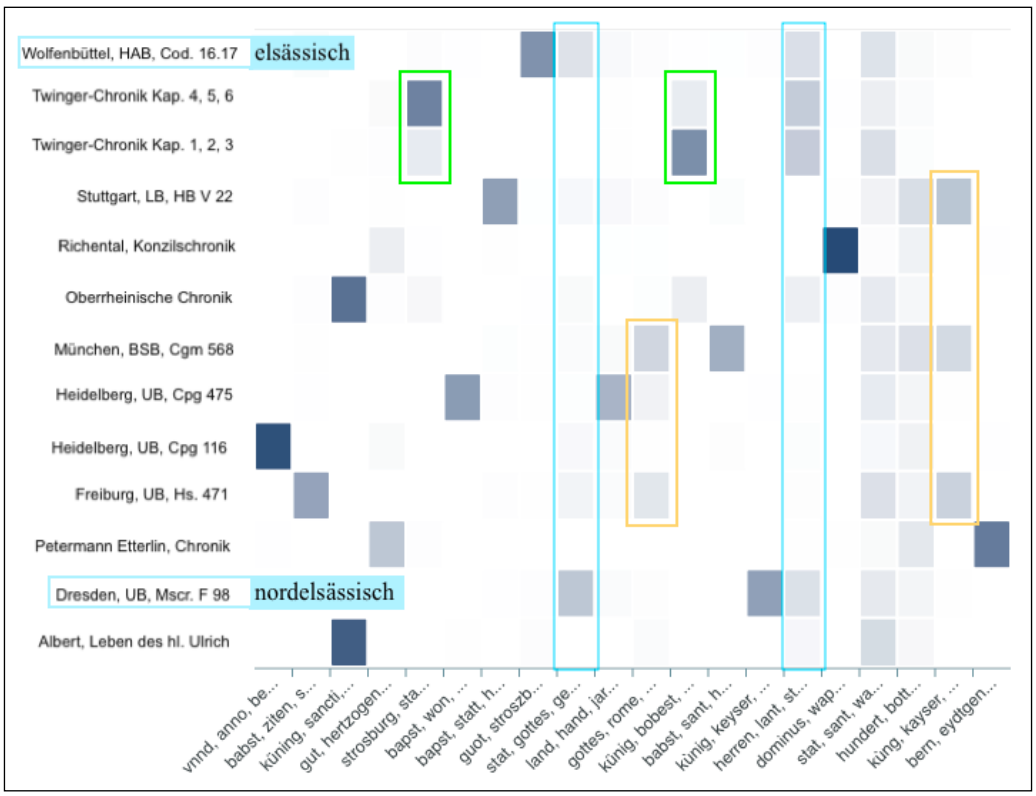

Fig. 11 Document-topic-matrix with weighed prevalence of a topic (13 documents, 20 topics), and writing dialects for the two related codices Wolfenbüttel, HAB, Cod. 16.17 and Dresden, UB, Mscr. F 98, as provided in the existing manuscript descriptions.

Both manuscripts are written in an Alsatian dialect, so the relation suggested through the topics very likely lies more in the writing of "lant" or "gottes" (instead of, e.g., "lande" or "gotz", spellings that are found in other codices in the corpus) than in the hitherto unknown shared transmission of (a) text(s). And while "elsässisch" and "nordelsässisch" are not too problematic if one considers the comparability of such metadata, things are more complicated for other codices (see Fig. 12).

While for Stuttgart, LB, HB V 22, we only know the place of production – as Constance is in an area with an Alemannian dialect – the label for the dialects of the other three codices do not seem too similar at first glance; however, they can in fact be put in close proximity (see Fig. 13).[49]

49 Based on Peter Wiesinger, "Die Einteilung der deutschen Dialekte." In *Dialektologie. Ein Handbuch zur deutschen und allgemeinen Dialektforschung,* 2nd half-vol., edited by

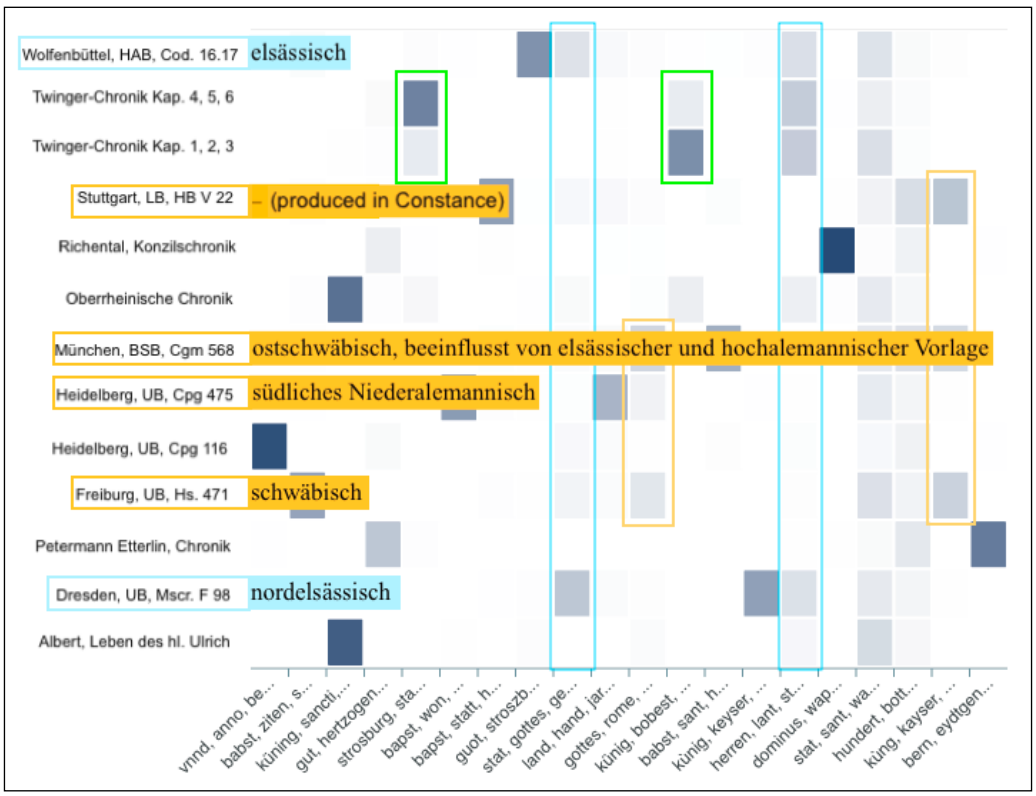

Fig. 12 Document-topic-matrix with weighed prevalence of a topic (13 documents, 20 topics), and writing dialects for the four related codices Freiburg im Breisgau, UB, Hs. 471, Heidelberg, UB, Cpg 475, Munich, BSB, Cgm 568 and Stuttgart, LB, HB V 22.

Of course, a proximity of writing dialects does not guarantee a proximity of the places of production – but it is a very good indicator for potentially related manuscripts, as is also suggested by the shared transmission of the *Konstanzer Jahrgeschichten,* the account of the murder of the bishop of Lausanne and a list of the bishops of Constance. In contrast to the information about the writing dialect in plain text, as is given in manuscript descriptions, without standardization or a universal classification system, the relation between the manuscripts through topics can be taken as additional and non-interpretive metadata, which can also be generated for codices which do not have a highly detailed description.

Werner Besch et al. Berlin/New York, 1983, 831. Even though the classification reflects the dialectal areas at the beginning of the twentieth century, the map can serve as a first indication for the localization of manuscripts.

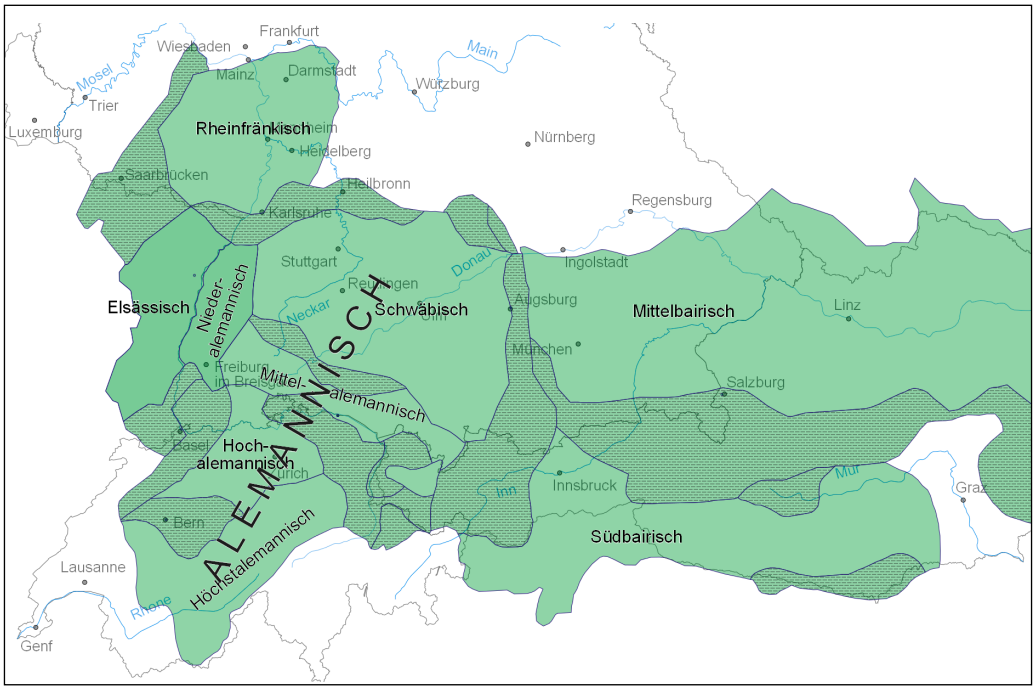

Fig. 13 Dialectal areas in southwest Germany, based on the classification by Peter Wiesinger. Own map created using SprachGIS (www.regionalsprache.de), 2023. The crosshatched areas were classified as transition zones that cannot be assigned to a specific dialect.

5. Conclusion

This article proposed the application of topic modeling to premodern manuscripts, in particular compilations, as an additional method for the classification of codices within a corpus. On the basis of this classification, the connections as well as the networks between multiple-text manuscripts can be illustrated, getting around some of the shortcomings of a purely text- or genre-based approach.

For the tracing of medieval manuscript transmission, network analysis based on single texts as connectors between codices has been successfully applied in both previous and current research. For certain corpora, this is a promising and fruitful approach that highlights the processes of copying and compiling, as is shown by the contributions in this issue. But it has its limits, among other things resulting from the imperfect metadata available for the majority of medieval manuscripts – one has to either rely on existing manuscript descriptions, or conduct a close reading of the documents in question, an undertaking which is not always feasible. To better understand the reading interests or intended functions of a specific compilation and its further circulation, a broader perspective, which

considers its complete content, could be insightful. Networks based on the genres of the distinct texts contained in a manuscript could be built and used for further exploration of a corpus. But genre-based classification always contains an interpretive element, and there is no universal ontology or taxonomy that can be used as a reference. To compensate for missing or inaccurate manuscript descriptions and to eliminate subjective or idiosyncratic genre attributions, topic modeling can be a viable method. Computationally calculated topics are inferred from the full text of a manuscript, independent of different stages of production or changing writing hands, thus taking into account the whole document.

Topic modeling calculates topics for a document in relation to the other documents in a specific corpus (which could also consist of the entirety of all existing manuscripts, of course). It therefore facilitates or enables comparison and can serve as an exploratory tool; material and paleographic analyses would need to follow, but they could benefit from pointers towards specific (groups of) manuscripts. And while some of the topics discussed in the examples above reflect dating conventions rather than revealing hidden content, it seems plausible that by curating distinctive lists of stop words and comparing their outcomes, topics could become more meaningful with respect to a thematic programme, and likely to the intention of the writer or their client – of course, attention has to be paid not to include another kind of interpretive bias through the excluded words. Developments and improvements in the field of text recognition and normalization and/or lemmatization for pre-modern languages would also add to the method, with the latter helping to focus more on content, but this seems to remain an unsolved problem for the moment.[50] A two-step approach of topic modelling, first with raw text, then with normalized text, could deliver both reliable results with regards to writing language as well as textual content of the manuscripts in a specific corpus.

At present, the apparent superficiality of the results – with regards to the "real" content – does not reduce the value added by the topics calculated. They provide useful indications towards relations between manuscripts without having to rely on mostly non-standardized metadata and without resulting in yet another tax-

50 The normalizer "Norma" is currently not being developed further, as the latest release dates from 2017; see https://www.linguistics.rub.de/comphist/resources/norma/index.html and https://github.com/comphist/norma. For a comparison of different approaches of normalization, including Norma, see Marcel Bollmann, "A Large-Scale Comparison of Historical Text Normalization Systems," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1 (Long and Short Papers)* (Minneapolis, MN, 2019), 3885–98, doi:10.18653/v1/N19-1389; Simon Flachs, Marcel Bollmann, and Anders Søgaard, "Historical Text Normalization with Delayed Rewards," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (Florence: Association for Computational Linguistics, 2019), 1614–19, doi:10.18653/v1/P19-1157.

onomy of subjective classifications. Experimenting with different forms of visualizations helps to interpret the results, while heatmaps and networks complement each other. In the worst case, the calculated connections merely point towards similarities between codices on a linguistic level, or to general textual similarities such as dating conventions. With the constant improvement of handwritten text recognition, by experimenting with stop words, and considering the wait for functioning normalizers, the results can certainly be improved, showing topic-based networks between manuscripts that help to better understand their production and transmission.

6. References

Bausi, Alessandro, Michael Friedrich, and Marilena Maniaci, eds. *The Emergence of Multiple-Text Manuscripts.* Studies in Manuscript Cultures 17. Berlin/Boston, 2019.

Blei, David M. "Probabilistic Topic Models." *Communications of the ACM* 55, no. 4 (April 2012): 77–84.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent Dirichlet Allocation." *The Journal of Machine Learning Research* 3 (March 1, 2003): 993–1022.

Bollmann, Marcel. "A Large-Scale Comparison of Historical Text Normalization Systems." In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1 (Long and Short Papers),* 3885–98. Minneapolis, MN, 2019. doi:10.18653/v1/N19-1389.

Buck, Thomas Martin, ed. *Chronik des Konstanzer Konzils 1414–1418 von Ulrich Richental: historisch-kritische Edition. Band 1: A-Version.* Vol. XLIX, 1–3. Konstanzer Geschichts- und Rechtsquellen. Ostfildern, 2020.

Cerquiglini, Bernard. *Éloge de la variante. Histoire critique de la Philologie.* Paris, 1989.

Fernández Riva, Gustavo. "Network Analysis of Medieval Manuscript Transmission. Basic Principles and Methods." *Journal of Historical Network Research* 3 (2019): 30–49.

Flachs, Simon, Marcel Bollmann, and Anders Søgaard. "Historical Text Normalization with Delayed Rewards." In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics,* 1614–19. Florence: Association for Computational Linguistics, 2019. doi:10.18653/v1/P19-1157.

Friedrich, Michael, and Cosima Schwarke, eds. *One-Volume Libraries. Composite and Multiple-Text Manuscripts.* Studies in Manuscript Cultures 9. Berlin/Boston, 2016.

Geith, Karl-Ernst, ed. *Albert von Augsburg: Das Leben des heiligen Ulrich.* Quellen und Forschungen zur Sprach- und Kulturgeschichte der germanischen Völker, n.F. 39 (163). Berlin/New York, 1971.

Graf, Klaus. "Die verschollene Twinger-Handschrift aus dem Regularkanonikerstift Dürnstein." *Archivalia,* March 27, 2013. https://archivalia.hypotheses.org/6941.

Grieshaber, Franz Karl. *Oberrheinische Chronik: Älteste bis jetzt bekannte, in deutscher Prosa.* Rastatt, 1850.

Gruber, Eugen, ed. *Petermann Etterlin. Kronica von der loblichen Eydtgnoschaft, Jr harkommen und sust seltzam stritten und geschichten.* Quellenwerk zur Entstehung der Schweizerischen Eidgenossenschaft 3, 3. Aarau, 1965.

Hegel, Carl, ed. *Die Chroniken der oberrheinischen Städte. Straßburg 1.* Die Chroniken der deutschen Städte vom 14. bis ins 16. Jahrhundert 8. Leipzig, 1870.

— ed. *Die Chroniken der oberrheinischen Städte. Straßburg 2.* Die Chroniken der deutschen Städte vom 14. bis ins 16. Jahrhundert 9. Leipzig, 1871.

Hillenbrand, Eugen. "Gallus Öhem, Geschichtsschreiber der Abtei Reichenau und des Bistums Konstanz." In *Geschichtsschreibung und Geschichtsbewußtsein im späten Mittelalter,* edited by Hans Patze, 727–55. Vorträge und Forschungen 31. Sigmaringen, 1987.

Hodel, Tobias, Dennis Möbus, and Ina Serif. "Von Inferenzen und Differenzen. Ein Vergleich von Topic-Modeling-Engines auf Grundlage historischer Korpora." In *Von Menschen und Maschinen. Mensch-Maschine-Interaktionen in digitalen Kulturen,* edited by Selin Gerlek et al., 181–205. Hagen, 2022.

Johnston, Michael, and Michael Van Dussen. "Introduction: Manuscripts and Cultural History." In *The Medieval Manuscript Book. Cultural Approaches,* edited by Michael Johnston and Michael Van Dussen, 2–16. Cambridge Studies in Medieval Literature 94. Cambridge, 2015.

Julien, Octave. "Délier, lire et relier. L'Utilisation de l'analyse réseau pour construire une typologie de recueils manuscrits de la fin du Moyen Âge." *Hypothèses* 19, no. 1 (2016): 211–24. doi:10.3917/hyp.151.0211.

Löser, Freimut. "Überlieferungsgeschichte(n) schreiben." In *Überlieferungsgeschichte transdisziplinär: Neue Perspektiven auf ein germanistisches Forschungsparadigma,* edited by Dorothea Klein, Horst Brunner, and Freimut Löser, 1–19. Wissensliteratur im Mittelalter 52. Wiesbaden, 2016.

Maniaci, Marilena. "Miscellaneous Reflections on the Complexity of Medieval Manuscripts." In *Collecting, Organizing and Transmitting Knowledge. Miscellanies in Late Medieval Europe,* edited by Sabrina Corbellini, Giovanna Murano, and Giacomo Signore, 11–22. Bibliologia: Elementa ad Librorum Studia Pertinentia 49. Turnhout, 2018.

Mone, Franz Josef, ed. *Quellensammlung der badischen Landesgeschichte 1.* Karlsruhe, 1848.

— ed. *Quellensammlung der badischen Landesgeschichte 3.* Karlsruhe, 1863.

Müller, Diana. *Textgemeinschaften. Der "Gregorius" Hartmanns von Aue in mittelalterlichen Sammelhandschriften.* Frankfurt a.M., 2013. http://publikationen.ub.uni-frankfurt.de/frontdoor/index/index/docId/30069.

Mutuvi, Stephen, Antoine Doucet, Moses Odeo, and Adam Jatowt. "Evaluating the Impact of OCR Errors on Topic Modeling." In *Maturity and Innovation in Digital Libraries,* edited by Milena Dobreva, Annika Hinze, and Maja Žumer, 3–14. Lecture Notes in Computer Science. Cham, 2018. doi:10.1007/978-3-030-04257-8_1.

Nichols, Stephen G. "What is a Manuscript Culture? Technologies of the Manuscript Matrix." In *The Medieval Manuscript Book: Cultural Approaches,* edited by Michael Johnston and Michael Van Dussen, 34–59. Cambridge Studies in Medieval Literature 94. Cambridge, 2015.

Nichols, Stephen G., and Siegfried Wenzel, eds. "Introduction." In *The Whole Book: Cultural Perspectives on the Medieval Miscellany,* 1–6. Recentiores. Later Latin Texts and Contexts. Michigan, 1996.

Pielström, Steffen, Severin Simmler, and Thorsten Vitt. "Topic Modeling with Interactive Visualizations in a GUI Tool." *Proceedings of the Digital Humanities Conference Utrecht 2019,* n.d. https://dev.clariah.nl/files/dh2019/boa/0637.html.

Purschwitz, Anne. "Netzwerke des Wissens – thematische und personelle Relationen innerhalb der Halleschen Zeitungen und Zeitschriften der Aufklärungsepoche (1688–1818)." *Journal of Historical Network Research* 2 (December 3, 2018): 109–42.

Robertson, Stephen. "Understanding Inverse Document Frequency: On Theoretical Arguments for IDF." *Journal of Documentation* 60, no. 5 (January 1, 2004): 503–20. doi:10.1108/00220410410560582.

Scherrer, Gustav. *Kleine Toggenburger Chroniken. Mit Beilagen und Erörterungen.* St. Gallen, 1874.

Schneider, Karin. "Berufs- und Amateurschreiber. Zum Laien-Schreibbetrieb im spätmittelalterlichen Augsburg." In *Literarisches Leben in Augsburg während des 15. Jahrhunderts,* edited by Johannes Janota and Werner Williams-Krapp, 8–26. Studia Augustana 7. Tübingen, 1995.

Schöch, Christof. "Computational Genre Analysis." In *Digital Humanities for Literary Studies: Methods, Tools, and Practices,* edited by James O'Sullivan, 219–31. College Station, TX, 2020.

— "Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama." *Digital Humanities Quarterly* 11, no. 2 (2017): §1–53. https://doi.org/10.48550/arXiv.2103.13019.

Schuhmacher, Mareike. "DARIAH Topics Explorer." *ForTEXT. Literatur digital erforschen.* Accessed October 29, 2021. https://fortext.net/tools/tools/dariah-topics-explorer.

Serif, Ina. "Der zerstreute Chronist. Zur Überlieferung der deutschsprachigen Chronik Jakob Twingers von Königshofen." *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte,* May 12, 2015. https://mittelalter.hypotheses.org/7063, last updated June 28, 2023.

— *Geschichte aus der Stadt. Überlieferung und Aneignungsformen der deutschen Chronik Jakob Twingers von Königshofen.* Kulturtopographie des alemannischen Raums 11. Berlin/Boston, 2020.

— "Konstanzer Jahrgeschichten." In *Encyclopedia of the Medieval Chronicle, 2nd Online Edition,* edited by Graeme Dunphy and Christian Bratu. Leiden/Boston, 2016. http://dx.doi.org/10.1163/2213-2139_emc_SIM_001450.

Snijders, Tjamke. "Work, Version, Text and Scriptum: High Medieval Manuscript Terminology in the Aftermath of the New Philology." *Digital Philology. A Journal of Medieval Cultures* 2, no. 2 (2013): 266–96.

Stolz, Michael. "Linking the Variance. Unrooted Trees and Networks." In *The Evolution of Texts. Confronting Stemmatological and Genetical Methods. Proceedings of the International Workshop Held in Louvain-La Neuve on September 1–2, 2004,* edited by Caroline Macé, 193–213. Linguistica Computazionale 24/25. Pisa, 2006.

Underwood, Ted. "Topic modeling made just simple enough," July 4, 2012. https://tedunderwood.com/2012/04/07/topic-modeling-made-just-simple-enough/.

Weingart, Scott B. "Topic Modeling for Humanists: A Guided Tour," July 25, 2012. http://scottbot.net/topic-modeling-for-humanists-a-guided-tour/.

— "Topic Nets," November 10, 2012. http://scottbot.net/topic-nets/.

Wiesinger, Peter. "Die Einteilung der deutschen Dialekte." In *Dialektologie. Ein Handbuch zur deutschen und allgemeinen Dialektforschung,* 2nd half-vol., edited by Werner Besch et al. Berlin/New York, 1983, 807–900.

Westphal-Wihl, Sarah. *Textual Poetics of German Manuscripts, 1300–1500*. Studies in German Literature, Linguistics, and Culture. Columbia, SC, 1993.

Wolf, Jürgen. "Sammelhandschriften – mehr als die Summe ihrer Einzelteile." In *Überlieferungsgeschichte transdisziplinär: Neue Perspektiven auf ein germanistisches Forschungsparadigma,* edited by Dorothea Klein, Horst Brunner, and Freimut Löser, 69–81. Wissensliteratur im Mittelalter 52. Wiesbaden, 2016.

The Journal of HISTORICAL NETWORK RESEARCH

CATHERINE EMERSON

Connecting Chronicles

Material and Social Bonds in the Circulation of *Chroniques Abrégées* in Fifteenth-Century Paris

Journal of Historical Network Research 9 (2023) 214–238

Keywords Gilles, Nicole (d. 1503), Fifteenth-century Paris, Historiography, Medieval Studies, Social Network Analysis

Abstract Using the methodology of Network Analysis, we can visualize the network of owners associated with manuscripts of fifteenth-century French histories and the material connections between the manuscripts themselves. This provides clues as to where to look for additional manuscripts in this corpus, which we know is incomplete. A clear distinction emerges between a group of male historians associated with the French court whose manuscripts were privately owned, and other texts in institutional collections which were subject to public consumption. The work of Nicole Gilles is examined as a case-study of the first sort of manuscript. Both sorts of manuscripts – whether associated with individuals or institutions, particularly monasteries and the royal court – are revealed as key points of exchange and contact. The study, based on the Archives Nationales de France, could be expanded to other archives to give a more complete picture of the way in which texts circulated. Certain actors, most notably Philip the Good, Duke of Burgundy, are revealed to have a similar pattern of manuscript ownership as institutions like the great abbey libraries of Paris.

C²DH LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

UNIVERSITY OF LUXEMBOURG

1. Introduction: Networks of People – Networks of texts

This paper examines a corpus of 26 French vernacular histories that circulated – and for the most part were written – in the fifteenth century. Substantial scholarship has been devoted to the issue of textual overlap between these histories, with scholars seeking to determine which texts are wholly distinct, which are separate redactions of the same text, and how the texts influenced each other.[1] Rather than addressing these textual questions, the current paper aims to shed some light on the personal and material connections that may have facilitated the transmission of texts. Network Analysis permits a visualization of the connections between people associated with manuscripts, allowing us to see how texts may have been transmitted and providing clues as to where lost or unattributed manuscripts might originally have been found. Network Analysis is used here primarily as a means to represent a large number of connections that cannot easily be shown by other means. The fifteenth and sixteenth centuries were not as bureaucratic as our age and so most people did not leave a trail of documents comparable in size to that which we create today. As a result, many people are hard to locate and have few documented connections to other people. This means, as we shall seee, that the network we are examining is not dense: that is, the likelihood that two nodes in the network will share an edge is low. However, by producing a visual representation of the network, we can see some of the connections between actors. This representation can be refined by grouping nodes based on modularity – a tool which allows us to identify communities of nodes within the network that share more edges with each other than with others. This permits us to identify actors who may have come into contact with lost or unidentified manuscripts containing texts in the same corpus. Since we know that such manuscripts did once exist, analysis of these clusters provides one set of clues as to where to look for them. Another sort of clue is provided by analysing the physical similarities between manuscript witnesses, which permits us to identify different patterns of ownership within the larger social network. The ego network of Nicole Gilles – that is, a subset of the larger network containing only the actors that are linked to this person – the author and owner of one of the manuscripts in question, provides a case study, showing where Gilles interacted with other people with documented connections to manuscripts containing texts in this corpus.

Acknowledgements: I would like to thank Dr Sarah Corrigan, who read an early draft of this paper.
Corresponding author: Catherine Emerson, University of Galway. catherine.emerson@universityofgalway.ie

1 See Guyot-Bachy and Moeglin, "Comment on été continuées les *Grandes chroniques de France*."

The corpus has been divided into separate texts according to groupings identified in the Jonas database.[2] Hosted by the French national *Institut de Recherche et d'Histoire des Textes,* this database is an exhaustive repertory of French medieval manuscripts. It presents the most recent scholarship on texts and on the manuscript witnesses to them, grouping texts by content rather than by title.[3] The corpus is listed in section 1.1 below, where departures from the groupings suggested by Jonas are signalled. It will be noted that many of the texts are referred to in at least some manuscripts as 'chroniques abrégées'. This designation can be regarded as a marker of genre rather than of text, given the diversity of texts that are given this title.

The physical properties of manuscripts have been determined largely with reference to the catalogues of the libraries housing them, as has information about the individuals associated with those manuscripts. I have investigated the social and legal connections between these individuals using the records of the French National Archives and in particular its *minutier central des notaires de Paris,* a vast register of legal documents signed in the city between the end of the fifteenth century and 2012.[4] This source provides documentation which is predictably rich in the case of individuals based in Paris. Men associated with the royal court are particularly prominent, due to the volume of legal relationships created by the business of government, such as homage for land. I have supplemented these records by consulting other documents in the French National Archives, such as the papers of the court itself. However, court records are less detailed for the period in question and contain fewer references to named individuals. Figure 1 below is a visualization of the resulting network, where nodes represent people or institutions and undirected edges represent the relationships between them. Since the court records are less detailed, most of the edges reflect connections documented in the *minutier.* This has the inevitable result of highlighting the legal relationships of Paris-based individuals. The prominence of the node representing Charles Duke of Berry, younger son of Charles VII of France, in figure 1 can be explained in this way. In 1462 and 1463, the year following the death of his father and the succession of his brother Louis XI, it was Charles – and not the king – who received homage for royal lands. As a result, there are a large number of records of legal relationships between Charles and other people. Charles was possibly the owner of a copy of Guillaume de Nangis's *Chronique abrégée des rois de France* (text 13 below: Brussels, Bibliothèque royale, 12246).[5] He was linked by family to

2 Leurquin, Anne-Françoise, Marie-Laure Savoye et al., "Jonas."

3 Guyot-Bachy, "La Chronique abrégée des rois de France et les Grandes chroniques de France," 209.

4 This resource contains over 100 million notarized documents, grouped into 122 reports. Documents from the fifteenth century are mainly found in MC/ET/XIX. This corpus is catalogued in Béchu, Greffe, and Pébay, *Minutier central des notaires de Paris.*

5 Provenance is established by means of a note 'C'est au seigneur de Berry', see Van den Gheyn et al., *Catalogue des manuscrits de la Bibliothèque royale,* 131.

his father and brother (who are both named as the subject of a number of the histories in the corpus), as well as to 141 other men and women, mostly because he received homage from them.

Examining the personal and legal connections between individuals gives some insight into how texts may have been transmitted. Combining this with information about the physical form taken by the manuscripts containing the texts allows us to see whether physically similar texts circulated in socially similar contexts. It is to be expected that physically similar copies may occupy a similar place in the lives of the people who commission, produce, and own them. Physically similar manuscripts may have been put to similar uses, stored in similar ways, and therefore also had similar fortunes as far as dissemination is concerned. In some cases, a similar material presentation may imply shared source material, for example where a passage is copied with its accompanying illustrations. This is not necessarily the case: physically similar books occupy the same niche in a material ecosystem. The content of those books may be an alternative text – that is, a different text fulfilling the same purpose – rather than an identical one.

This study of documented legal relationships and their intersection with the different material form in which texts circulated is complemented by a case study of Nicole Gilles. Gilles makes a good point of comparison, because we have a description, dating from 50 years after his death, that hints at how his work was transmitted, and because his life is richly documented in the French National Archives and elsewhere. A notary and secretary to the king, he was also a churchwarden in the parish of St Paul, and involved in a business partnership with the publisher Antoine Vérard.[6] The death of his wife Marie Turquam (1499) prompted an inventary of books owned by the household, including a number of unbound volumes that should have been returned to Vérard, and one of two surviving manuscripts which bear traces of Gilles's ownership.[7] By comparing what we know about Gilles to what we can conclude about people in similar positions in the network, we can make preliminary conclusions about how the texts may have been diffused. Looking at the material forms in which these texts survive tells us something about the sort of texts that were transmitted through different connections. This is thus a study of textual diffusion, rather than of reception: the evidence examined is traces of legal and familial relationships between individuals and information about the manuscripts they produced and owned, rather than evidence about how or whether the texts were read.[8] It is anticipated that in the future this work can be extended with reference to archives beyond the French National Archives and in the light of fresh provenance information.

---

6 See Scheurer, "Nicole Gilles and Antoine Vérard."

7 For information on these manuscripts, see Emerson, "Nicole Gilles and Literate Society," 56; Doucet, *Les Bibliothèques parisiennes,* 83–89.

8 For this distinction, see Brix, "Aux marges des manuscrits," 61.

1.1 Corpus

The corpus considered in this article consists of 26 vernacular histories which survive in manuscripts from the fifteenth century or the very early years of the sixteenth century. Many of the texts were written or compiled during this period, but earlier texts have been included because the survival of manuscript witnesses from this era confirms an ongoing interest in the text. Initially drawn up on the basis of texts identified by Kathleen Daly, the corpus has been extended and refined with reference to the Jonas database, which has in some cases revised scholarship concerning the textual tradition of individual manuscripts.[9] The texts considered are as follows:

1) *Chroniques abrégées*. A text apparently composed in Paris around 1330, it survives in 41 manuscript witnesses from the period, catalogued under titles ranging from *Les Chroniques abregées du commencement du Monde jusqu'au temps pape Jehan* (London, British Library, Harley 4001) to *Les hystoires et les croniques de Vincent abregiees, seconde rédaction, avec continuation jusqu'en 1347* (Paris, Bibliothèque nationale de France, français 1368);[10]
2) Nicole Gilles, *Annales et chroniques de France* (2 manuscripts);[11]
3) Jacques le Picart, *Chronique abrégée* (1 manuscript, catalogued as *Abrégé des chroniques de France, depuis l'origine des Français jusqu'à Charles VIII,* Troyes, Bibliothèque Municipale, MS 812);[12]
4) *Grandes chroniques de France* (83 manuscripts belonging to different redactions of the text);[13]
5) Louis Le Blanc, *Bref récit des rois de France* (5 manuscripts);[14]
6) Louis Le Blanc, *Mémorial des hauts faits des rois de France* (1 manuscript, Paris, Bibliothèque nationale de France, français 5869);[15]
7) Louis Le Blanc, *Sainte vie et les hauts faits de monseigneur saint Louis, roy de France* (1 manuscript, Paris, Bibliothèque nationale de France, français 5869);[16]
8) Louis Le Blanc, *Prétensions des rois d'Angleterre* (2 manuscripts);[17]
9) Jean Le Bègue, *Manuel* (1 manuscript, Paris, Bibliothèque nationale de France, latin 12815), not listed in Jonas but described by Daly;

9 See footnote 2.
10 Details of the text and its known witnesses can be found at: http://jonas.irht.cnrs.fr/oeuvre/3693. Where applicable, a comparable page will be cited for each text.
11 http://jonas.irht.cnrs.fr/oeuvre/2907.
12 http://jonas.irht.cnrs.fr/oeuvre/22360.
13 http://jonas.irht.cnrs.fr/oeuvre/3892.
14 http://jonas.irht.cnrs.fr/oeuvre/2319.
15 http://jonas.irht.cnrs.fr/oeuvre/2320.
16 http://jonas.irht.cnrs.fr/oeuvre/2321.
17 http://jonas.irht.cnrs.fr/oeuvre/22358.

10) Pierre Amer, *Manuel* (1 manuscript, Paris, Bibliothèque nationale de France, français 10988) (also described by Daly);
11) *Chronique de France 1403–1434* (9 manuscripts). A note in the Jonas database draws attention to the fact that different manuscripts present text with different end points;[18]
12) *Chronique abrégée des rois de France* (1 manuscript, Paris, Bibliothèque nationale de France, nouvelles acquisitions françaises 4811);[19]
13) *Chronique abrégée des rois de France jusqu'en 1382* (1 manuscript, Paris, Bibliothèque nationale de France, nouvelles acquisitions françaises 4951);[20]
14) Guillaume de Nangis, *Chronique abrégée des rois de France* (14 manuscripts);[21]
15) Guillaume de Nangis, *Chronique amplifiée des rois de France, dite de Guillaume de Nangis* (22 manuscripts);[22]
16) *Chronique abrégée et continuée dite de Baudouin d'Avesnes* (9 manuscripts);[23]
17) *Chronique universelle de la création à Charles VII* (8 manuscripts);[24]
18) *Chronique du règne de Charles VI,* previously attributed to Jean Juvenal des Ursins (2 manuscripts);[25]
19) Jean Chartier, *Chronique de Charles VII roi de France* (16 manuscripts);[26]
20) *Généalogie des rois de France* (64 manuscripts, including many in rolls);[27]
21) Jean le Tartier, *Chronique abrégée de 1095 à 1328* (4 manuscripts);[28]
22) *Chronique universelle de la création à Philippe IV* (13 manuscripts);[29]
23) Gilles Le Bouvier, *Chronique de Charles VII* (21 manuscripts);[30]
24) Guillaume de Nangis, *Chronique amplifiée,* continued with material from *Chronique* of Noël de Fribois. This redaction is presented in a single manuscript, Baltimore, Walters Art Gallery, W 00306, a manuscript that is also listed in Jonas as a witness to the two texts it combines (items 15 and 25 in this list);[31]
25) Noël de Fribois, *Chronique* (23 manuscripts);[32]
26) *Miroir historial abrégé de France,* considered by Daly and others to present striking similarities with the work of Noël de Fribois. It is attributed to him,

---

18 http://jonas.irht.cnrs.fr/oeuvre/23096.
19 http://jonas.irht.cnrs.fr/oeuvre/24080.
20 http://jonas.irht.cnrs.fr/oeuvre/22289.
21 http://jonas.irht.cnrs.fr/oeuvre/5295.
22 http://jonas.irht.cnrs.fr/oeuvre/5277.
23 http://jonas.irht.cnrs.fr/oeuvre/5326.
24 http://jonas.irht.cnrs.fr/oeuvre/10123.
25 http://jonas.irht.cnrs.fr/oeuvre/5748. For the attribution, see Lewis, "L'Histoire de Charles VI, attribuée à Jean Juvénal des Ursins."
26 http://jonas.irht.cnrs.fr/oeuvre/10112.
27 http://jonas.irht.cnrs.fr/oeuvre/4293.
28 http://jonas.irht.cnrs.fr/oeuvre/7094.
29 http://jonas.irht.cnrs.fr/oeuvre/5336.
30 http://jonas.irht.cnrs.fr/oeuvre/10924.
31 http://jonas.irht.cnrs.fr/manuscrit/73811.
32 http://jonas.irht.cnrs.fr/oeuvre/3294.

though the Jonas database considers this attribution erroneous. Jonas lists 5 manuscripts. A sixth, sold by Christie's on 25th May 2016 and listed in the catalogue of that sale as a *Miroir historial abrégé* with a possible attribution to Noël de Fribois, is now considered by the Jonas database as a witness to text 25 above.[33]

As this enumeration demonstrates, the definition of what constitutes a separate text is fluid, as texts often incorporate lengthy passages from other sources, and continuations using excerpts from other texts. This is particularly the case for the three final items on this list, but as we shall see, it is also true of the interrelation between 2 and 15, and between many of these texts and 4.

1.2 Methodology

Taking this corpus into consideration, I have noted all indications of a manuscript having been in contact with an actor – through production, sale or ownership etc. – in the fifteenth century or the first four decades of the sixteenth century. However, not all manuscripts can be linked to an individual in these early years. In this corpus of 26 texts, transmitted in 295 separate mauscripts, 64 occurrences of a text can be linked to 61 separate individuals.[34] There are also institutions, such as the Abbeys of Saint-Denis and that of Saint-Germain-des-Prés, whose libraries held copies of these manuscripts. In sum, just over 18% of manuscripts containing a text in this corpus can be associated with an individual or institution before 1540, with some actors being linked to more than one text. These actors form the nodes in a network shown in figure 1 below. Edges represent documented contact between these nodes as shown in the records of the French National Archives. The methods for constructing and visualizing this network, which result in figure 1, will be described more fully below.

Amongst the nodes we find patrons and owners such as Thomas Thwaytes, who ordered London, British Library, Royal 20.E.I–VI (a witness to text 4), and Henry VII of England, to whom the same manuscript was dedicated. We also find scribes and illustrators. These people are frequently anonymous craftsmen, whose names are not known. Nevertheless, in many cases there is evidence of a relationship between these anonymous figures and a particular patron. For instance, the Master of the Getty Froissart was one of the illustrators of London,

33 http://jonas.irht.cnrs.fr/oeuvre/5344. The catalogue of sale of this final manuscript can be found at: https://www.christies.com/lot/lot-noel-de-fribois-d1467-8-miroir-historial-abrege-5994385/? [accessed April 2022].

34 Some manuscripts, such as Brussels, Bibliothèque royale, 10233-10236, contain more than one work in the corpus (in this case 16 and 20). Where such a manuscript is linked to the same individual, this person is naturally linked to both texts. Conversely, some individuals, such as Philip the Good of Burgundy, owned several manuscripts containing more than one text in the corpus.

British Library, Cotton Augustus V (containing text 1), alongside the similarly anonymous Wavrin Master. The Master of the Getty Froissart is believed to have also worked for Louis de Gruuthuse, owner of another illustrated manuscript, Paris, Bibliothèque nationale de France, français 2691, containing text 19 and illustrations attributed to the Master of the Harley Froissart, the Master of the Commynes Froissart, and to Philippe de Mazerolles. Because the names of these artisans are not known, we cannot determine the full extent of their ego network. Nevertheless, it is important to include them because they provide links between bibliophiles and patrons, who in turn share edges with other people in the network. Such connections are best documented where a manuscript is illustrated. This is an instance of a phenomenon noted by Anotine Brix whereby illustrated manuscripts receive more attention, with the result that their provenance is better documented.[35] We will see below how illustrated manuscripts feature in the network in different ways, reflecting these different relationships between individuals and their manuscripts.

The network described in figure 1 is constructed using the concepts of primary and secondary relationships. A primary relationship is defined as a relationship between a person or institution who had direct contact with a manuscript and another person with whom they appear together in documentation, such as the *minutier*. The nature of the relationship can be purely legal (that between tenant and landlord, or client and patron), it can be familial (husband and wife, or father and son), or it can be corporate (colleagues or members of the same association). Secondary relationships are defined as those between actors already appearing in the network, but where neither party has a documented connection to a manuscript in the corpus. Looking at these secondary relationships increases the density of the network by revealing all the edges between its nodes, thus showing indirect links between two people associated with manuscripts. For instance, it allows us to see a relationship between Claude Goffier, whose inheritence was settled by Jacques de Beaune, and Goffier's friend Hans Breda, who accompanied Francis I on his Italian campaign. This in turn reveals an indirect connection between Jacques de Beaune and Francis I.

The sample is limited chronologically to the years 1440–1540. End points are formed by the careers of Francis I (d. 1547, formerly Francis of Angoulême – owner of Paris, Bibliothèque nationale de France, français 61 – a manuscript of text 20 – and of London, British Library, Harley 4878 – a manuscript containing text 1) and of Antoine, Cardinal Sanguin de Meudon (d. 1559, owner of St. Petersburg, National Library of Russia, Fr. F.v.IV.1, a manuscript of text 4). Data was processed in a spreadsheet, comprising 4801 rows, each documenting an edge between two nodes. There are 1153 nodes in the network, and 2582 separate edges. This means the graph has a density of only 0.003, even when secondary relationships are in-

35 Brix, "Aux marges des manuscrits," 62.

cluded. This is, as I have said, a low density, reflecting the relative absence of bureaucratic records for the period. However, even though there are comparatively few documented edges between nodes, we can still identify features of the network and draw some speculative conclusions about the connections between texts and the actors that came into contact with them during the period 1450–1550. In order to visualize this network, a graph was produced using Gephi graph visualization software. Edges have been coded in the initial spreadsheet to reflect whether one or both nodes represent an actor directly associated with a manuscript. This enables the visualization in figure 1, where nodes representing actors associated with manuscripts are coloured blue and those representing actors not directly associated with a manuscript are coloured yellow.

The weight of the edge between the two nodes shows the number of different connections between two individuals. The average degree of a node is 3.5, but there is a large variation. 45% of nodes have a degree of 1, meaning they share an edge with just one other actor, and a further 22% have a degree of two. Edges also have weights of up to 20. This number can reflect different documents which link two nodes, or different sorts of links mentioned in the same document. For instance, Nicole Gilles and the apocethary Guillaume Gaigny appear together in four documents issued between 1494 and 1500 because they were churchwardens in the same parish of Saint Paul, charged with maintaining the fabric of the church and receiving donations on its behalf. This results in an edge with a weighting of four, reflecting each of the legal relationships that linked the two men. Conversely, when Antoine, Cardinal Sanguin de Meudon rented a land and a title to his neice, Anne de Pisseleu, and her husband, Jean IV de Brosse, the deed is testimony to two sorts of link between each of the three people, one reflecting their family relationship and one their financial one.[36] This single deed then contributes a weighting of two to these edges, although the family relationship is only encoded once in the graph regardless of how many times it is mentioned in documents. The graph is undirected, partly because of the wide range of relationships encoded in the edges. If all the edges showed the same sort of relationship (for example if all the edges represented lease agreements between landlords and tenants), it would be possible to construct a directed graph, but this is not the case. Many of the relationships are between equals: for instance, Gilles and Gaigny acting to represent their parish, sometimes together with other men. Even where there is an imbalance, it is not always easy to determine which of the parties is favoured. For example, Louis and Pierre-Martin Affaitati appear on the graph because they lent money to Francis I in 1549. However, the relationship between creditor and debtor is not necessarily that of dominant and subordinate, particularly when the debtor is the monarch. As a consequence, the edges shown in figure 1 simply reflect the fact that two individuals knew each other and do not

36 Paris, Archives nationales de France, *Minutes de Michel de Felin,* MC/ET/III/13, 12 July 1537.

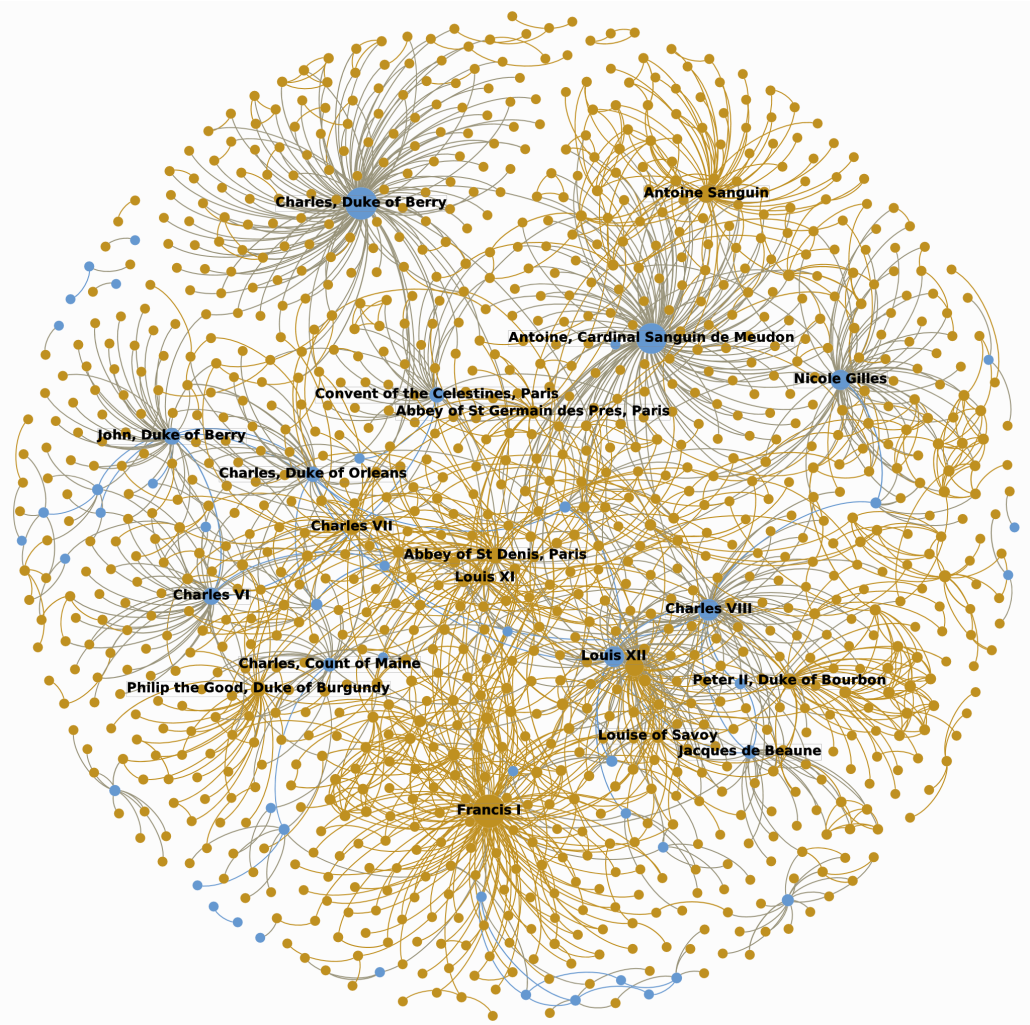

Fig. 1 Network of all known individuals associated with manuscripts (nodes in blue) and their links to other individuals not associated with manuscripts (nodes in yellow) based on documentary evidence in the French National Archives. Textual labels have been given only to nodes with a degree of 20 or more.

illustrate the quality of that relationship. Given that the purpose of the graph is to look at how texts may have been diffused through social connections, it is the fact of the connection which is of primary importance for the current study. A connection represents a possible pathway for diffusion, whether or not it is used for this purpose. Where there is no pathway, diffusion cannot happen.

In this figure, only nodes with a degree of 20 or more are labelled, while the colour has been assigned based on whether the individual is associated with a manuscript containing a text in the corpus (blue) or not (yellow). Given the meth-

odology used to construct the graph, which takes the manuscript as the starting point, it is not surprising that some of the most connected people in this graph are associated with manuscript copies. We also see a few large yellow nodes, representing well-connected individuals without a direct connection to a manuscript. These are men (and one woman, Louise of Savoy – Francis I's mother) with close family ties to people who collected books, and secondary relationships with the same people their relative interacted with. There are also a number of small blue nodes, representing early owners or producers of manuscripts who are not well connected in the network. The smaller blue nodes on the periphery are often the anonymous craftsmen whose presence in legal documents might become more apparent if their names were known. Others represent men such as Anthony, bastard of Burgundy, who were active outside Paris and therefore do not appear much in Parisian legal documents. If the network was expanded with reference to other archival material, such nodes might increase in prominence.

2. The Case of Nicole Gilles – Indications of how transmission might work

Before examining the distribution patterns of different types of texts through this network, it is illustrative to zoom in on one part of the network, Nicole Gilles, and examine how his case illuminates our study of transmission through the network. Gilles is a good example precisely because he was a Parisian, meaning that his ego network is likely to be more fully represented in the Paris records than that of men from outsite the city. As such, his case may illustrate what a network might have looked like for other, less-well represented, individuals. In his network, nodes can be clustered on the basis of shared edges, that is, subgroups of people who interacted with each other can be observed in Gilles's ego network. Some of these clusters are more connected to manuscripts in the corpus than others. Gilles also provides a good case study because we have additional evidence about his own connection to a manuscript in this corpus, allowing us to understand one way in which such manuscripts may have been compiled and used.

Gilles himself is associated not only with manuscripts in this corpus, but also with other manuscripts. The 1499 inventory of his books appears to be incomplete in that it includes Paris, Bibliothèque nationale de France, français 17088, a copy of the *Doctrinal des simples gens,* which he signed on the flyleaf, "C'est a N. Gilles's" but not Paris, Bibliothèque nationale de France, français 789, despite the fact that it bears Gilles's signature and the date on which he purchased it.[37] Also missing from the inventory is a manuscript, now Paris, Bibliothèque nationale de France, nouvelles acquisitions françaises 1417, generally consid-

37 Doucet, *Les Bibliothèques parisiennes,* 83–89. This manuscript is digitized at: https://gallica.bnf.fr/ark:/12148/btv1b107232000 [accessed April 2022].

ered Gilles's autograph copy of his French-language chronicle (number 2 in our corpus).[38] The text found in this manuscript, along with its presentation, give a fascinating insight into the way that a writer of a French-language chronicle approached his work in the late fifteenth century. The subject was a popular one at the time and the layout of Gilles's manuscript recalls that of contemporary printed books and presentation copies of manuscripts. Some sections have been marked in the top right-hand corner of the recto folio with indications of the reign being discussed, in a manner reminiscent of running heads.[39] Most paragraphs are marked with a title, and many also with a pilcrow, indicating that the writer was copying the conventions of formal textual transmission which facilitate orientation around a manuscript or a printed text. Whilst these features are reminscent of printed texts and presentation quality manuscripts, other features of Paris, Bibliothèque nationale de France, nouvelles acquisitions françaises 1417 demonstrate that it is a working copy: it contains numerous deletions, emendations, and insertions that indicate the author's revisions over a period of time, and led one of the text's early editors, Denis Sauvage, writing in 1549, to deplore the quality of the text that he was forced to engage with.

> [C]ertainement ceulx, à qui estoyent ces Exemplaires, ou Copies à la main, curieux d'avoir en un seul livre tout ce qui povoit avoir esté fait en tous pays (au moins en beaucoup) par chacune année, avoyent entremeslé, chacun à par soy, sur la marge de leurs livres, plusieurs choses, non accordantes à nostre principal: qui estoit aussi pour la pluspart accoustré de mesme les autres.
>
> [Certainly those people, the owners of these examples, or handwritten copies, anxious to have in one book everything that could have occurred in all countries (at least in many) year by year, had each individually added in, in the margin of their books, many things which did not fit in with our main topic: which was also, for the most part laid out like the others].[40]

Sauvage's testimony is interesting, because it suggests that not only did this editor use Gilles's autograph manuscript, but also that he was familiar with other codices laid out in a similar manner. He writes that he has based his edition on 'plusieurs vieux Exemplaires, et entre autres un, qui fait quelque foy d'estre de la main de l'Autheur' [many old examples and among others one, which has some claim to be by the hand of the author].[41] This comment implies that the practice of add-

38 This manuscript has been digitized at: https://gallica.bnf.fr/ark:/12148/btv1b9007155s [accessed April 2022].

39 This is particularly true of the earlier sections of the manuscript, for instance, Paris, Bibliothèque nationale de France, nouvelles acquisitions françaises 1417, fols. 2–9. Riche considers these early pages as representing authorial revision.

40 Sauvage, "A tous Lecteurs, dignes de ce nom."

41 Sauvage, "A tous Lecteurs, dignes de ce nom."

ing to and revising these manuscripts was common to many owners of copies of the text Sauvage was editing, not merely Gilles, its author. Gilles's chronicle does not appear to have been printed before 1525, two decades after the author's death. Earlier, supposedly lost, printed editions of the chronicle are cited by Jacques Le Long and subsequent scholars as having been published in 1492 and 1498, but Jacques Riche concluded that these were in fact editions of the French chronicle of Guillaume de Nangis (text 15 above), which presents similar material to Gilles's work, and is often published with a similar title. Sauvage's observations as to the state of Gilles's text in 1549 hint at a way that the text may have remained in circulation for so long after the author's death without a diffusion in print format. If Gilles shared his work with his associates during his lifetime, and if they engaged with their copies in the same way that the author did with his own, this would indicate a diffusion that predated printing. It would also imply the possibilty of other, lost or as yet unidentified, manuscripts of Gilles's work. Indeed, in 2013, Marie-Laure Savoye identified a manuscript, Vatican City, Biblioteca Apostolica Vaticana, Reg. Lat. 937, as a second witness to Gilles's text, which had previously been thought to be preserved only in Paris Bibliothèque nationale de France, nouvelles acquisitions françaises 1417.[42] The Vatican manuscript, catalogued as *Histoire depuis Charlemagne jusques a Philippe second, roy de France,* presents, as this title indicates, a less complete text than the Paris manuscript. It currently begins and ends mid-sentence, obscuring the full extent of the text originally copied. Savoye dates it to the first decade of the sixteenth century, making it slightly more recent than Gilles's autograph copy, and it contains none of the marks of early engagement – additions, erasures and emendations – that distinguish Paris, Bibliothèque nationale de France, nouvelles acquisitions françaises 1417. Its early provenance is also obscure; the only indication in this regard is that the manuscript was owned in the seventeenth century by Jean and Pierre Bourdelot, which tells us nothing more about how Gilles's text circulated in the years between the author's death and the first printing of his text.

2.1 Proliferation and Confusion of Vernacular Histories

The very fact that the Vatican manuscript was not recognized as a witness to Gilles's text for so long points to difficulties in separating the many similar historiographic works circulating in the period. We have seen how supposed early printed editions of Nicole Gilles are now thought to be editions by Jean Trepperel of Guillaume de Nangis's *Chronique* containing later continuations.[43] This inference is rendered more plausible by the fact that the first references to editions

42 See the notice in Jonas: http://jonas.irht.cnrs.fr/manuscrit/76109 [accessed June 2021].

43 For a discussion of Guillaume de Nangis's work and related texts, see Delisle "Mémoire sur les ouvrages de Guillaume de Nangis;" Guyot-Bachy "La Chronique abrégée des rois de France de Guillaume de Nangis;" Brix "Une réécriture méconnue des *Grandes Chroniques de France*."

of Gilles's work predating 1525 were made in 1719 by Jacques Le Long.[44] Le Long considers Gilles as a simple abbreviator of the *Grandes Chroniques de France,* for the early part of his work, and of Guillaume de Nangis for the later section of the work. Given these assumptions, Le Long may have been more likely to confuse the three works. There is no doubt that Gilles's chronicle is textually very close to the continuations of Nangis. However, the directionality of the relationship might be challenged given that the correspondence is particularly strong in those sections dealing with events that had occurred during Gilles's lifetime and after the death of Guillaume de Nangis in 1300.[45] It is known that Gilles owned a manuscript containing extracts from the Chronicle of Guillaume de Nangis, which he bequeathed in his will to one of his associates, Jean de Fontenay, who was possibly also the scribe who had produced the manuscript in the first place.[46] Fontenay, as we shall see below, had a number of personal and professional connections with Gilles, as well as with Gilles's business associate, the publisher and bookseller Antoine Vérard. Vérard in turn had access to manuscripts containing other texts that ressemble or overlap with Gilles's chronicle and that of Guillaume de Nangis. As the publisher of the first printed editions of the chronicles of Jean Froissart (1495) and Enguerrand de Monstrelet (1499), Antoine Vérard had a demonstrated interest in French vernacular historical writing. A manuscript, Paris, Bibliothèque nationale de France, français 32144, which was owned by Vérard, contains two texts in the corpus considered by this article, texts 19 and 23.

Moreover, as a royal secretary, Gilles had connections to a world that was heavily implicated in chronicle reading and writing. Kathleen Daly has pointed out that chronicle writing was a popular genre at the time, and was particularly composed of men who were royal secretaries: Noël de Fribois, Jacques le Picart, and Louis Le Blanc.[47] Daly shows how Fribois, who was a generation older than Gilles, le Picard and Le Blanc, took an approach to historiography which appears to have influenced that of his colleagues who succeeded him.

Noël de Fribois's chronicle survives in two separate recensions, listed as items 24 and 25 above. Another work, listed at item 26, is also often attributed to him. Once again, the reuse of material found in Guillaume de Nangis, especially in item 24, points to the fact that many of the texts in this corpus are very similar. Moreover, there are also overlaps between the subject matter or even the text contained in individual chronicles and that of the *Grandes Chroniques de France,*

44 Le Long, *Bibliothèque historique de la France,* 378, item no. 7433.
45 For an examination of such a passage in Gilles and continuations of Nangis, as well as in the work of Philippe de Commynes, see Emerson, "Nicole Gilles's Presentation of the Death of Louis XI."
46 This manuscript is now Bern, Burgerbibliothek, Cod. 70. This manuscript is discussed in Riche, "L'Historien Nicole Gilles," 36–37; 86–87.
47 Daly, "Mixing Business with Leisure."

which circulated widely in manuscripts produced during the fifteenth century. As this example reveals, the connections between the texts are often as complex as those between the men who produced and owned the manuscripts containing them. A sufficient amount of data exists to construct visualizations of a series of networks. Some visualizations demonstrate connections between actors associated with texts, while a parallel analysis sheds light on the connections between the texts themselves. Comparing the two allows us to see how different sorts of texts had different sorts of diffusion. The complexity of those connections means that visualization is the best way to make sense of the network.

2.2 Examining Nicole Gilles's Ego Network

Figure 2 shows a presentation of a section of the data shown in figure 1, relating only to those individuals connected to Nicole Gilles (i.e., his ego network). Every person in this graph has a primary relationship with Gilles, but their secondary relationships with each other have also been included. Thus, for instance, Jean and Pierre Turquam were both connected to Gilles, since they were his brothers-in-law, but also to each other, as brothers. These family connections were supplemented by legal relationships. Jean was nominated as guardian to Gilles's minor children and represented Pierre in legal cases. These groups have been coloured to distinguish them, using the modularity filter in Gephi. By applying the software's 'community detection' algorithm, 13 different groups were identified with a modularity resolution of 0.38. The community detection algorithm groups the network into distinct subgroups, setting the resolution at a level that produces a workable number of distinct groups. These can be mapped onto different areas of Gilles's life, but of course the partition does not reveal the full complexity, because each node is assigned to just one cluster – that with which it has the closest relations – and because some individuals encountered Gilles in more than one area of his activity.

One example of this is Jean de Fontenay, who appears in the top right-hand corner of the figure, coloured in dark blue. Fontenay was, like Gilles, a member of the royal court, described in the legal records in which he appears as a clerk, '[c]lerc ordinaire en la chambre des comptes'. However, he does not appear in the cluster of Gilles's colleagues, coloured green in the top left of the figure. Like Gilles, he was also a churchwarden in the parish of St Paul, but these men are distributed throughout the graph, some belonging to the cluster of light green colleagues, some coloured dark green, placed below them, and some coloured black at the bottom of the graph. This tripartite separation reflects three groups of men who had exercised this office in different years. Jean de Fontenay's node appears in a different part of the graph, reflecting his familial and financial ties to Gilles and his links to Gilles's publishing activity. He was married to Gilles's daughter, Jeanne, whose node appears next to that of her husband. As Jeanne's husband and Gilles's son-in-law, he was the recipient of a gift from Gilles of a house and garden in Pierrelaye. The couple sold this house to Gilles's business associ-

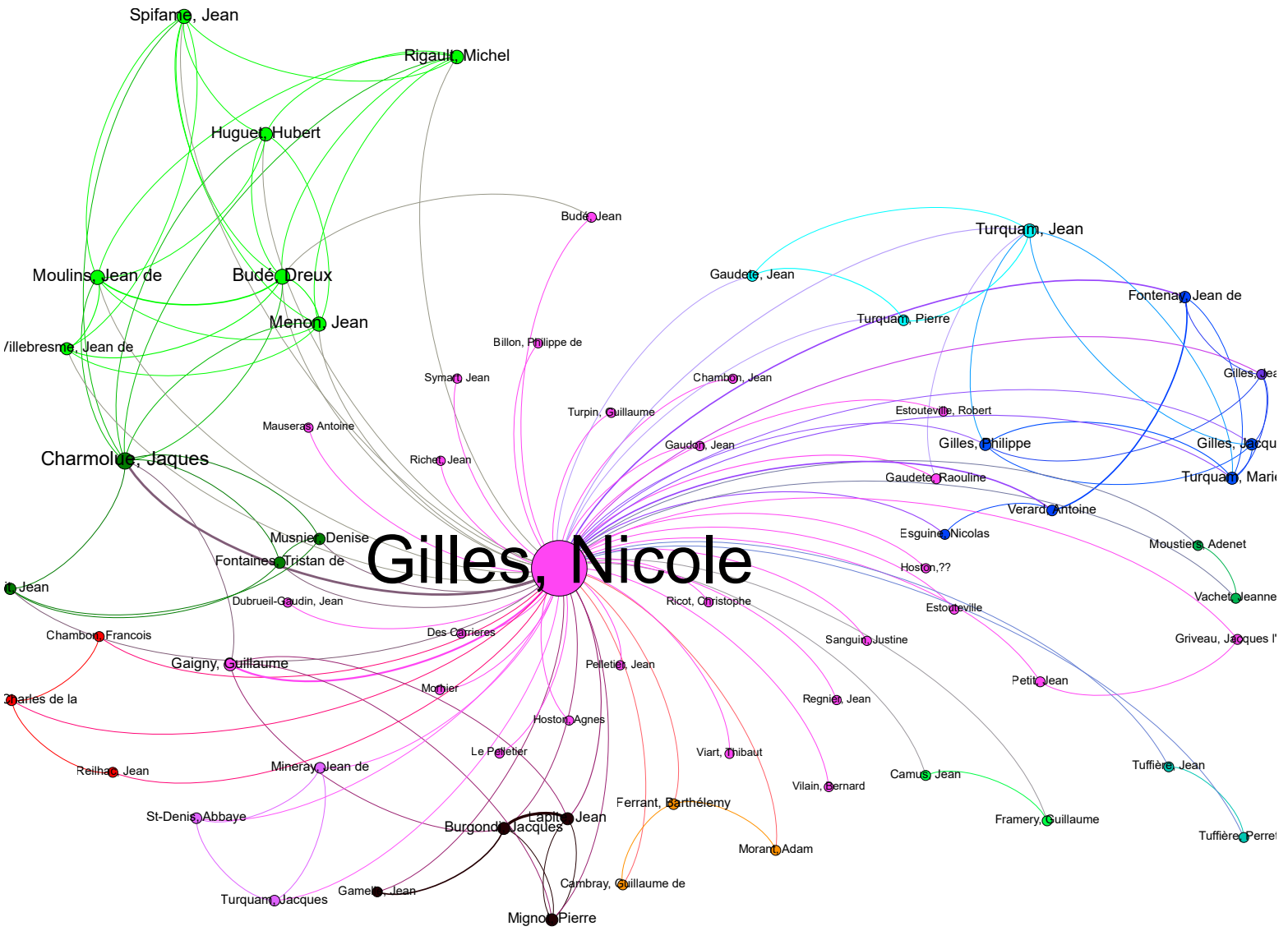

Fig. 2 Nicole Gilles at the centre of his ego network. Colours distinguish clusters of individuals linked to each other as well as to Gilles. Clustering has been performed automatically by graph modularity.

eISSN: 2535-8863
DOI: 10.25517/jhnr.v9i1.128

ate, Antoine Vérard, for two hundred *livres*.[48] Subsequent documents reveal that Vérard paid by transferring debts to Fontenay.[49] As we saw above, Gilles also bequeathed a manuscript of Guillaume de Nangis's chronicle to Fontenay, who may have been the manuscript's scribe. The common ownership of this manuscript is just one of a set of complex commercial, professional, personal and spiritual bonds that linked Fontenay and Gilles, and both men with Vérard.

As the example of Fontenay reveals, modularity analysis is best at situating an actor who only had one area of interaction with another actor. That is, it works well for identifying groups of friends or associates, but it is not as good at locating actors with ties to more than one of these groups.

If we were to overlay the data from figure 1 onto figure 2, we would see four nodes in Gilles's ego network highlighted in blue to indicate that they were associated with a manuscript in the years 1440–1540. These are Gilles, Vérard, and Fontenay, all of whom appear in the cluster of family and professional contacts that we have just noted, and the Abbey of Saint-Denis, coloured purple at the bottom left of the graph, linked to Gilles through a legal case in 1489, in which Gilles gave evidence on behalf of his nephew Jacques Turquam.[50] This appears in a different section from the other members of Gilles's family, because Jacques Turquam is not explicitly linked in any notarial document to any other inlaws of Gilles. However, his surname suggests a family connection, confirmed by his explicit designation as Gilles's nephew, suggesting that, at least for Gilles, his connections to the people and institutions who owned and produced books were limited to the relatively small portion of his associates who all interacted with his family.

2.3 Beyond Gilles's Immediate contacts

This conclusion is revealed as incomplete, however, if we bear in mind Daly's scholarship on chronicles written by royal secretaries. Even though Gilles, Le Picart, and Le Blanc were more or less contemporaries in the court of Charles VIII and Louis XII, Le Picart and Le Blanc do not appear in figure 2, because they are not mentioned in the same documents in the French National Archives that mention Nicole Gilles. Daly has made a convincing case for an institutional interest in historiography among the secretaries of the court, and this is true, even though it is unclear how much interaction individual postholders had with each other.

48 Paris, Archives nationales de France, *Minutes de Pierre I Pichon,* MC/ET/XIX/12, 10 January 1498.

49 Paris, Archives nationales de France, *Minutes de Pierre I Pichon,* MC/ET/XIX/12, 10 January 1498; MC/ET/XIX/13, 28 December 1498.

50 Paris, Archives nationales de France, *Minutes de Pierre I Pichon,* MC/ET/XIX/4, 4 September 1489.

The position of royal secretary was a limited, but broad category. Theoretically restricted to fifty-nine men, incumbents could resign their office in favour of a relative, which meant that two people could hold the same post simultaneously.[51] The fact that two men were both secretaries at the same time, therefore, might indicate a shared professional culture without implying direct personal contact. Moreover, the royal court has not produced the same detailed records of personal contact as those that appear in the notaries' records, meaning links between Gilles, Le Picart, and Le Blanc may have been lost, may not have been documented at the time, or may not have in fact existed. What is certain, though, is that these men who shared a professional background also produced very similar historical work, both in its content and in its physical form. As far as content is concerned, this can be explained by the fact that writers in this position all had access to the same sources available in the court, which they exploited in their chronicles. Alongside this, the writers shared a common education, training, and cultural background. They also shared a professional attitude to history which all of them seem to have seen as a private interest, given that they did not claim to be writing for a patron.

This similarity in professional attitude is reflected in a material similarity between manuscript witnessess of these authors' texts. Nicole Gilles's autograph manuscript is very similar to that of Jacques le Picard: both are laid out according to the conventions of formal manuscripts or printed books, with features such as pilcrows and running heads, and both contain a large number of additions and deletions, both in the text and in the margins. They are both manuscripts on paper of a similar size and length.[52] In terms of the material ecosystem of these *chroniques abrégées,* they belong to a class of texts – together with the *Manuel* of Pierre Amer (text 10 above) – which survive in very few manuscript witnesses (at most two), of which one is the author's autograph copy. Manuscripts of the *Miroir historial abrégé de France* (text 26) also share similar material properties in terms of the size of the average manuscript, and that all manuscript witnesses contain only this text, which is comparatively long, (over 200 pages), meaning that the four works can be grouped together in terms of the physical similarity of the surviving material witnesses. Gilles, Le Picard, and Fribois (the author to whom the *Miroir historial* was long attributed) share a professional background. This common background could explain the similarity in the presentation of their text. An examination of Gilles's ego network reveals associations with manuscripts in our corpus in particular parts of his social circle. This might suggest where we could look in the wider network for other people associated with manuscripts of ver-

51 Riche, "L'Historien Nicole Gilles," 47.

52 Jacques le Picart's text is preserved in Troyes, Bibliothèque municipale, MS 812, which is a manuscript on paper of 235 pages measuring 300 mm × 210 mm. Nicole Gilles's manuscript, Paris, Bibliothèque nationale de France, nouvelles acquisitions françaises 1417, is also a manuscript on paper. It measures 288 mm × 210 mm and has 242 pages.

nacular chronicles. If we were looking specifically for manuscripts more similar to Gilles's autograph copy, we might concentrate on nodes that might not share edges with Gilles's node but rather have a similar professional background. Comparing data of both sorts allows us to build a more complete picture of the connections which allowed texts to circulate.

3. Material, Textual, and Social Circulation

The 26 texts in our corpus can be grouped on the basis of the material similarities between the surviving manuscript witnesses. A distance matrix was established based on their features, including: the presence or absence of illustrations and of rubrication; the material support used (paper or velum or, in a few cases, both); the number of columns; the dimensions of the manuscript; the length of the manuscript; and the inclusion of other texts alongside the chronicle in our corpus. For each text, data was collected for each of the surviving manuscript witnesses containing the text, and the average of each of these metrics was compared to the average of the same metrics for manuscript witnesses of the other texts. One further metric has been used, which is the surviving number of manuscript witnesses to a particular text. This is, of course, related to the number of manuscripts originally produced, though not straightforwardly, as is well recognized.[53] It is also related to the care that has been devoted to a manuscript's preservation. A text that is preserved in a large number of witnesses was either more popular or has been better conserved than one that is not (or both). Although we do not know which of these alternatives is true of any given text, either one reflects significant data about the way in which the text circulated, and so I have considered it as one factor among many in composing the distance matrix. It should also be noted that some textual traditions are more uniform than others. 78% of the 64 manuscripts which contain text 20 are illustrated, but this means that nearly a quarter are not. In manuscripts of the same text, the average number of columns used is 2.5, but this number varies between 1 in Paris, Bibliothèque nationale de France, français 4990, and 5 in Manchester, Rylands Library, FR99; within a single manuscript, even the number of columns varies at different points. The use of averages to construct the distance matrix identifies overall trends in the material witnesses to a text, with larger numbers of manuscripts producing a more representative insight into the typical way in which a text is presented, because an anomalous manuscript has less of an impact on the averages.

Taking all these reservations into account, a number of groups can be established, based on the extent to which any textual tradition differs from the average

53 For a new attempt to resolve this problem, see Kestemont et al., "Forgotten Books."

Goup	Texts	Material characteristics	Colour in figure 3
A	1, 15, 19, 23	Relatively large manuscripts, half of which are presented in two columns and most of which contain another text or texts alongside one of the texts in this group.	Red
B	2, 3, 10, 26	Texts survive in a small number of copies, most witnesses contain only this text.	Pink
C	4	Long texts (over 320 pages), high proportion of manuscripts are illustrated and two thirds are on vellum.	Blue
D	8, 9, 18, 21	Short texts (less than 100 pages), generally circulating (though not in the case of text 8) in manuscripts containing other texts.	Light Blue
E	5, 6, 22, 25	Low number of surviving copies, mainly on paper. High proportion of illustrated manuscripts.	Yellow
F	12, 13	Similar in profile to group B but smaller format.	Not associated with nodes in this graph
G	7, 11, 14	Even shorter texts typically making up a smaller proportion of the manuscript it circulates in.	Magenta
H	16, 17	Ratio of height of book to width suggests thin manuscript.	Light green
I	20	A large number of rolls in this text (and this text alone in this corpus).	Turquoise
J	24	A paper manuscript but with colour decorations on the page and illustrations.	Gold

Tab. 1 French vernacular chronicles grouped by their material characteristics.

observed in the manuscripts considered as a whole. The clearest representation appears when the textual traditions listed in 1.1 above are separated into ten groups of manuscripts, as listed in this table.

The similarities in column three represent trends that can be observed across the surviving witnesses to a particular text, which have led to the texts being grouped together. Within these groups, individual manuscripts can diverge quite substantially from this norm. The colours referenced in column four refer to the colours used in figure 3 below. These colours have been added to reveal the pattern of ownership of manuscripts belonging to a particular group. Where an individual is associated with manuscripts containing texts from more than one group, the colours have been blended. F is not given a colour since no early owner of a surviving manuscript can be identified for this group.

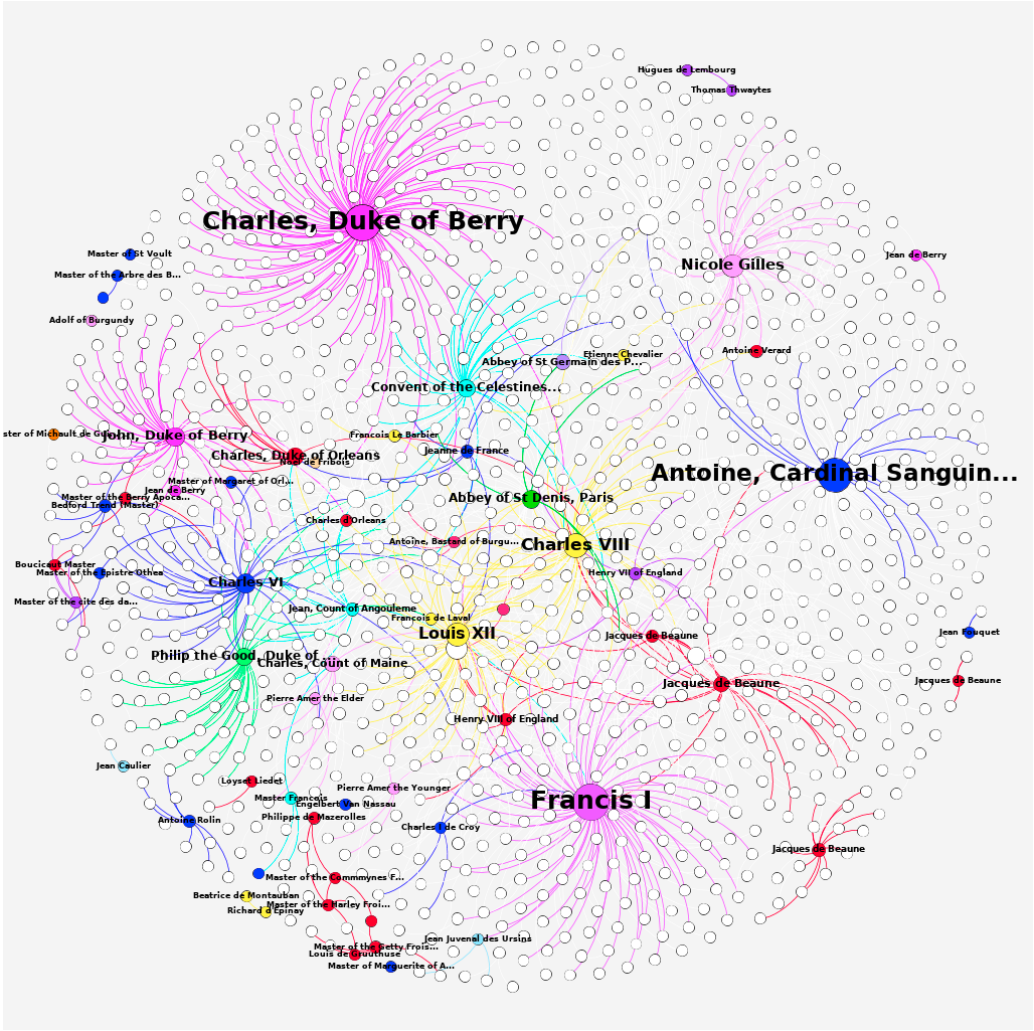

Fig. 3 Alternative visualization of the network shown in figure 1 with individuals associated with manuscripts coloured in accordance with the colours assigned in table 1.

Group C is the group which presents an aggregate of features that diverge least from the typical features of manuscripts in this corpus. This group consists only of manuscripts of text 4, the *Grandes Chroniques de France*. Although this group only contains one text, it is closely linked to all the other groups but one, reflecting material similarity to other manuscripts that parallels the fact that the text of the *Grandes Chroniques* is often reused or adapted in other chronicles, and also reuses and adapts their texts. The one tradition that is very different from the others and stands apart is group I, which contains only text 20, the *Généalogie des rois de France*. This text survives in 64 manuscripts, many of them extensive rolls comprising several columns of parallel text histories of different

regions.[54] Although this history displays a considerable amount of textual overlap with other texts in the corpus, the manuscripts in which it is transmitted are consistently physically dissimilar from the other manuscripts considered here. It also has a distinctly different pattern of early ownership. The catalogues of the libraries holding the 64 manuscript witnesses to this text identify only 6 individuals associated with these manuscripts in the period 1440–1540. This represents just 9% of the total number of manuscript witnesses to this text. By contrast, Nicole Gilles's chronicle belonging to group B comprises 9 manuscripts, which can be associated, through information provided in equivalent catalogues, with 4 individuals before 1540, meaning that their early provenance can be identified in 45% of instances. This different pattern of ownership relates to the material differences between the manuscripts in the two groups and to the different social uses of the documents concerned. The texts in group B tend to be recorded in private ownership, often that of their author, while the large size and consdierable amount of decoration of the manuscripts within group I lend to public display, suggesting institutional ownership of the sort that leaves few material traces.[55]

Putting Social and Material Networks together

To investigate this further, a refinement of the graph shown in figure 1 has been produced, showing the same network but with different colours representing the different material groups into which the corpus has been divided (figure 3). Here, only nodes associated directly with manuscripts are labelled. Around the perimeter of the graph there are a number of nodes with a low degree: these are mainly craftsmen whose names are not known who worked on illustrated manuscripts in groups A, C, and J. Group A is the largest group, with 100 surviving witnesses from the period, and we see nodes reflecting an association with a manuscript or manuscripts in this group across the network. Many of these nodes represent scribes and illustrators. Some, for example that of Charles duke of Orléans, are relatively well connected, and one, that of Francis I, is very well connected (with the highest degree of any node in the graph), by virtue of the length of his life and his position at the head the French government. His node is blue/pink, however, reflecting the fact that he is associated both with manuscripts containing texts in group A and texts in group I. Unusually in this graph, this represents two separate manuscripts. Very few individuals are associated with more than a single manuscript, and those who are are often craftsmen with low-degree nodes, for the reasons discussed in section 1.2. Even nodes associated with more than one text are

54 For a discussion of the manuscript tradition of an associated text, the *Chronique Anonyme Universelle,* which appears 28 manuscripts, mostly alongside the *Généalogie des rois de France,* see Fagin Davis, *La Chronique Anonyme Universelle.*

55 Indeed, as Pearson points out, entry into an institutional library can often result in the loss of information about early provenance, see Pearson, *Provenance Research in Book History,* 4–5.

most frequently people associated with one manuscript containing two texts in the corpus. Hence, Antoine Vérard's manuscript, Paris, Bibliothèque nationale de France, français 23144 contains both text 19 and text 23 (from group A), while Henry VII's manuscript London, British Library Royal 20.E.I–VI contains text 4 (group C) and text 19 (group A). Aside from Francis I, only the Burgundian bibliophiles Philip the Good and Charles of Croy and the institutional libraries of the Abbeys of Saint-Denis and Saint-Germain-des-Prés are associated with more than one manuscript containing a text in our corpus. In such a small population, conclusions can only be tentative, but it is worth noting that Antoine Vérard's mansuscript contained the text of Bouvier's and Chartier's chronicle, while Henry VII's contained Chartier's chronicle alongside the text of the *Grandes chroniques* and Charles of Croy owned separate manuscripts, containing the *Grandes chroniques* and Bouvier's chronicle. It seems, then, that texts 19 and 23 were frequently circulated with other material, and particularly as a supplement to the *Grandes chroniques*. Further, we can see Philip the Good (associated with the most texts in this corpus – 5 instances in groups A, C, and H) has a pattern of engagement with texts in this corpus that most closely resembles the institutional libraries of Saint-Denis (associated with copies of texts in groups C and E) and Saint-Germain-des-Prés (associated with copies of texts in groups A, D, and I). If we expanded examination of the network beyond the records in the French National Archives, it is likely that Philip the Good's node would gain prominence, as his transnational territory means that many of his legal relationships were outside the influence of Paris. It is interesting that this preliminary study shows a pattern of association with vernacular chronicles that mirrors that of the large institutional libraries of the time, particularly given the way that modern scholarship has treated Philip's collection as the origin of public research libraries.[56]

4. Conclusion

Looking at the ego network of Nicole Gilles, one early owner of a manuscript of a vernacular chronicle, shows us the personal connections that linked such owners. The use of the documents in the French National Archives, together with information from Denis Sauvage and Gilles's autograph manuscript, allows us to obtain a particularly detailed picture of the network of this one individual. Zooming out to look at other people associated with manuscripts enables us to supplement this picture, revealing connections through people like Francis I and Charles, duke of Berry. Texts 19 and 23, for example, or those in group A, to which those texts belong, are likely to have circulated through such well-connected individuals. Texts like those in group B are, conversely, most likely to be found in con-

56 See, for instance, the way that the library is described by the project *Libraire des ducs de Bourgogne,* https://www.kbr.be/fr/projets/la-librairie-des-ducs-de-bourgogne/ [accessed April 2022].

nection with people with similar backgrounds to Gilles, though not necessarily those represented by a node sharing an edge with Gilles's. Inevitably, this picture is only partial, since I have investigated connections in only one archive. Even if many archives were to be consulted, the picture would remain partial, since the preservation of records is not uniform, even within individual archives. There will also be gaps if only manuscripts are considered, since at the end of the period printed books are clearly part of the textual tradition.[57] However, attempts to redress this by considering printed books are likely to meet with limited success, given the scarcity of provenance information regarding printed books in the period. As a result, we should recognize that the approach taken here can be expanded to give a more detailed picture of the textual transmission of vernacular chronicles in French, but it will never achieve completeness. Nevertheless, this dual approach – considering both the material transmisison of the text and the personal connections between the people associated with physical copies containing the texts – is fruitful, both because it gives confirmation of things that we perhaps intuited but now see in a different perspective, and because it opens up fresh insights and avenues for exploration.

5. References

Béchu, Claire, Florence Greffe, and Isabelle Pébay. *Minutier central des notaires de Paris, minutes du XVe siècle de l'étude XIX.* Paris: Archives nationales, 1993.

Brix, Antoine. "Une réécriture méconnue des *Grandes Chroniques de France*: signalement, tradition manuscrite, sources." *Bibliothèque de l'École des chartes* 171, no. 2 (2013): 383–406.

— "Aux marges des manuscrits. Éléments pour une étude de la réception des *Grandes Chroniques de France*." *Questes. Revue pluridisciplinaire d'études médiévales* 36 (2017): 59–83.

Daly, Kathleen. "Mixing Business with Leisure: Some French Royal Notaries and Secretaries and their Histories of France, c. 1459–1509." In *Power, Culture, and Religion in France c. 1350–c. 1550,* edited by Christopher Allmand, 99–115. Woodbridge: Boydell, 1989.

Delisle, Léopold. "Mémoire sur les ouvrages de Guillaume de Nangis." *Mémoires de l'Institut national de France* 27, no. 2 (1873): 287–372.

Doucet, Roger. *Les Bibliothèques parisiennes au XVI^e siècle.* Paris: Picard, 1956.

Emerson, Catherine. "Nicole Gilles and Literate Society." In *'Le Bel Épy qui foisonne': Collection and Translation in French Print Networks,* edited by Catherine Emerson, 53–70. Oxford: Peter Lang, 2019.

57 On the failure of book catalogues to distinguish between manuscripts and printed texts, see Doucet, "Les Bibliothèques parisiennes;" Emerson "Nicole Gilles and Literate Society."

— "Nicole Gilles's Presentation of the Death of Louis XI and the Collection of Symbols of Kingship." *The Medieval Chronicle* 14 (2021): 47–62.
Fagin Davis, Lisa. *La Chronique Anonyme Universelle. Reading and Writing History in Fifteenth-Century France.* London: Harvey Miller, 2014.
Guyot-Bachy, Isabelle. "La Chronique abrégée des rois de France de Guillaume de Nangis: trois étapes de l'histoire d'un texte." In *Religion et mentalités au Moyen Âge,* edited by Sophie Cassagnes-Brouquet, Amaury Chauou, Daniel Pichot, and Lionel Rousselot, 39–46. Rennes: Presses universitaires de Rennes, 2003. https://doi.org/10.4000/books.pur.19784.
— "La Chronique abrégée des rois de France et les Grandes chroniques de France: concurrence ou complémentarité dans la construction d'une culture historique en France à la fin du Moyen âge?" *The Medieval Chronicle* 8 (2013): 205–32.
Guyot-Bachy, Isabelle, and Jean-Marie Moeglin. "Comment ont été continuées les *Grandes chroniques de France* dans la première moitié du XIVe siècle." *Bibliothèque de l'École des Chartes* 163, no. 2 (2005): 385–433.
Kestemont, Mike, Folgert Karsdorp, Elisabeth de Bruijn, Matthew Driscoll, Katarzyna A. Kapitan, and Pádraig Ó Macháin. "Forgotten Books: The Application of Unseen Species Models to the Survival of Culture." *Science* 375 (2022): 765–69.
Le Long, Jacques. *Bibliothèque historique de la France.* Paris: Gabriel Martin, 1719.
Leurquin, Anne-Françoise, Marie-Laure Savoye, et al., "Jonas, Répertoire des textes et des manuscrits médiévaux d'oc et d'oïl," accessed 28 June 2023, https://jonas.irht.cnrs.fr/.
Lewis, Peter S. "L'Histoire de Charles VI attribuée à Jean Juvénal des Ursins: pour une édition nouvelle." *Comptes rendus des séances de l'Académie des Inscriptions et Belles-Lettres* 140, no. 2 (1996): 565–69.
Pearson, David. *Provenance Research in Book History.* 2nd ed. London: British Library, 1998.
Riche, Jacques. "L'Historien Nicole Gilles (14..–1503). Sa vie, son manuscrit original, les premières éditions de ses Annales et sa place dans l'historiographie française." Diplôme d'archiviste paléographe, École des Chartes, 1930. Archives nationales de France 128, no. 2, Pierrefitte-sur-Seine, AB/XXVIII/83.
Sauvage, Denis. "A tous Lecteurs, dignes de ce nom." Introduction to *Annales et Chroniques de France, depuis la destruction de Troye jusques au temps du Roy Loys XI jadis composées par feu Nicole Gilles,* edited by Denis Sauvage. Paris: Du Pré, 1549.
Scheurer, Rémy. "Nicole Gilles et Antoine Vérard." *Bibliothèque de l'École des chartes* 128, no. 2 (1970): 415–19.
Van den Gheyn, J., Eugène Bacha, and Emile Wagemans. *Catalogue des manuscrits de la Bibliothèque royale de Belgique.* vol. X, *Histoire d'Espagne, Histoire de France, Histoire d'Italie.* Renaix: J. Leherte Courtin, 1919.