FERNÁNDEZ RIVA, GUSTAVO

Network Analysis of Medieval Manuscript Transmission. Basic Principles and Methods

Journal of Historical Network Research 3 (2019) 30-49.

Keywords

medieval studies, manuscripts, network analysis, shared manuscript transmission, Middle High German, digital humanities

Abstract

Manuscripts are one of the main sources for the study of medieval history and culture. Their features, production, circulation and transmission have been the subject of research from different disciplines and perspectives. This article will introduce an innovative way to investigate medieval manuscript transmission using network analysis. The computational study of networks has recently shown some great advancement, both as a visualization strategy and as a mathematical model to study complex phenomena, and can be very productively applied to medieval book history. Here we will focus on the theoretical and technical foundations to create a network of shared manuscript transmission. These

networks allow researchers to apply innovative exploratory visualization techniques and statistical methods. As a test sample, a network created to examine the shared manuscript transmission of texts written in German will be presented. The data for this research has been compiled from the online database *Handschriftencensus* and it has been processed and analyzed using *Python* and *Gephi*. The focus of the article are the theories, methods and strategies to create a network of shared manuscript transmission.

1 Introduction: Why a Network of Shared Manuscript Transmission*

Manuscripts are one of the most important sources to explore the medieval past. Scholars have studied their production, circulation and destruction; how they were read and used; how they interacted with oral discourse and with culture in general.[1] European medieval manuscripts are important as cultural heritage, as aesthetic objects, and as a material link to the Middle Ages. Even beyond their textual content, features like layout and script are valuable research areas that offer insight into history and culture.

Tens of thousands of medieval manuscripts are housed in various institutions across the world and their metadata stored in several digital databases. The availability of those databases could offer scholars the possibility to investigate medieval manuscripts from a quantitative and comparative perspective – what has seldom been done. Even before the overall digital availability of relevant databases, Neddermeyer[2] carried out a quantitative study in an attempt to better understand the transformation processes from

* **Acknowledgments:** Earlier versions of this paper were read at the *Digital Humanities Lunch*, organized by the Institute of Polish Language of the Polish Academy of Sciences in Krakow, on 15 March 2019 and at the conference *Visualisierung als Provokation? Netzwerkanalysen in Mediävistik und Altertumswissenschaften* held in Heidelberg on 3-4 April 2019.
I want to acknowledge my gratitude to Jürgen Wolf, Daniel Könitz, and Bernhard Runzheimer from *Handschriftencensus* for approving my use of their data, receiving me, answering my questions about the project and asking for my feedback on how to improve it. The work they do is incredible valuable for research in Medieval German Literature.
Corresponding author: Gustavo Fernández Riva. gustavo.fernandez.riva@uni-heidelberg.de.

1 Two valuable introductions to the field are: Raymond Clemens and Timothy Graham, *Introduction to Manuscript Studies* (Ithaca/London: Cornell University Press, 2007) and Erik Kwakkel, *Books Before Print* (Arc Humanities Press, 2018).

2 Uwe Neddermeyer, *Von der Handschrift zum gedruckten Buch. Schriftlichkeit und Leseinteresse im Mittelalter und in der frühen Neuzeit. Quantitative und qualitative Aspekte* (2 vols. Wiesbaden: Harrassowitz, 1998).

manuscripts to printed books. More recently, Buringh[3] used statistical methods to estimate how many manuscripts were produced in different places and times, how many currently survive, and other similar questions. These innovative and valuable researches, however, have not always been kindly received and remain exceptional cases.[4]

The present article introduces a quantitative model to investigate shared manuscript transmission using tools and method from the field of network analysis and takes advantage of the availability of digital manuscript catalogues. An approach similar to the one proposed here has been recently advanced by Zdenko Vozar,[5] who used an online manuscript database (*manuscriptorium.com*) to study a corpus of Hussite reform collection in order to demonstrate complex relations between authors and their "collecting". Using network analysis he was able to explore which authors appear in the same collections. This is, in its core, an implementation of the same method explained in this article, but applied to a smaller corpus and focusing on authors instead of texts. Other interesting projects are also currently working on how network analysis can be used to explore medieval manuscripts, although most of them are still very exploratory, focusing on very specific corpora, and with research questions that do not really include the analysis of shared manuscript transmission.[6]

The goal of this article is to set the theoretical foundations for the network analysis of shared manuscript transmission. This method enables scholars to ask, and propose answers to, certain important questions in the field of medieval manuscripts studies. However, before presenting the model it is important to

---

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

4 Some shortcomings in the methodology of Neddermeyer's study are a point of harsh critique in the reviews: Günther Görz and Ursula Rautenberg, „Medienwechsel Bibliometrisch“, *IASL Online* (2001), https://www.iaslonline.lmu.de/index.php?vorgang_id=2334; Hanns Peter Neuheuser, „Rezension zu Uwe Neddermeyer: Von Der Handschrift...'', *Bibliothek* 25, no. 1 (2001): 110–12. Buringh's research is more solid from the mathematical perspective, although replication and more localized studies would be welcomed in order to further test the results.

5 Zdenko Vozár, „Metadata for the Middle Ages: A Network Analysis of Manuscriptorium.com“ (Paper presented at the Historical Network Research Conference, Masaryk University, Brno, Czech Republic, 2018).

6 Petra Mutlová has proposed an analysis of a corpus of texts from the Czech Reformation in order to test the hypothesis that the exchange of ideas among the Hussite reformers was governed by the laws of complex networks: Petra Mutlová, „Networks of Ideas in the Czech Reformation,“ Presentation at the *Historical Network Research Conference 2018*. Masarykova univerzita: Brno, 2018. Evina Steinova is working on early medieval copies of the *Etymologiae* of Isidore of Seville (~ 400 mss.), using networks to analyze innovations shared by manuscripts, e.g. a specific rewriting of one chapter: Evina Steinova, „Innovating Knowledge: Isidore's Etymologiae in the Carolingian Period,“ in *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 2 (2019): 12-15.

clarify what shared manuscript transmission is and why it is a relevant research object.

The concept of 'transmission' highlights that, from this perspective, manuscripts are regarded primarily as the medium upon which texts are recorded and 'transmitted' to readers. This means, that many relevant aspects of the manuscripts (layout, materials, etc.) will be left aside, in order to focus on their role as a medium for storing and communicating a textual content. From a different perspective, 'transmission' also means considering manuscripts as a form of communication between the time of their creation and the present – to interrogate the set of objects available nowadays, and try to infer some features about their production, circulation and destruction. 'Shared' transmission, refers to the fact that codices (i.e. physical books) usually contained more than just one text. There were many ways in which compilations could be achieved. Peter Gumbert establishes a distinction between

> the kind where several texts are contained in an object which physically is a single, homogeneous book, and the kind where several physical units, which originally were separate, were joined within the covers of one binding.[7]

This distinction is not always easy to make and the catalogues and databases do not always include this information in a machine readable way. In both cases, however, we can assume some sort of agency which brought together textual objects following some (vague or concrete) criteria. Even if the reasons behind a particular compilation were relatively simple and practical (availability, chance, etc.), it is sensible to presuppose that the compilation could not be completely random, and respected at least some limitations and fundamental principles. To test the validity of this claim of non-randomness in the compilation process is one of the goals of the network analysis proposed in this article.

Compilations are enabled and encouraged by the technology of the codex itself. In the previously dominating format, the roll, navigating the text was much more difficult, and the amount of content that could be included in one object was comparatively small. The codex allows to navigate the text easier, so that, for example, tables of content become feasible. Also, the amount of sheets that can be bound together can vary greatly from a single quire to hundreds. Finally, combining different previously separated quires in a single codex is relatively easy. In short: for many technological and cultural reasons, compilation was a common practice in the production of medieval manuscripts. Also, as big bound books are sturdy and valuable, they are more likely to

[7] Johann Peter Gumbert, „One Book with Many Texts: The Latin Tradition", in *Codices Miscellanearum. Brussels Van Hulthem Colloquium* (Brussels: Archives et bibliothèques de Belgique, 1999), 28.

survive the pass of time than unbounded booklets. As a consequence of these facts, current catalogues show a great amount of large codices and compilations.

Shared manuscript transmission is a fruitful area of research and offers valuable insight into medieval texts and culture.[8] There are mainly two complimentary approaches, which need to be carried out in conjunction, even when only focusing on one of them: the text-centered approach and the manuscript-centered approach. The first consists in analyzing how one text changes depending on the manuscript context.[9] For example, some textual variants might be explained as a way of adapting the work to the principles guiding a particular compilation. However, sometimes the insertion in a different textual environment alone changes the interpretation of the work, because it is considered as part of a new context. The complementary approach is to focus on one or more particular miscellany manuscripts in their entirety.[10] This kind of research has a long tradition in medieval philology and tries to understand the principles behind a particular compilation and the reasons that guided the persons responsible for creating it. The question for the structure of a miscellany manuscript goes beyond finding a conscious intention behind the grouping of texts. Even if the compiler(s) did not have a clear goal, there must have been practical and cultural criteria that determined which texts were selected. Studying the shared manuscript transmission is a way of understanding the availability and the interpretation of certain works in a given cultural context.

Detailed studies of shared manuscript transmission, focusing on particular texts or manuscripts, are valuable and needed, but this article proposes a different way to approach the issue: considering a big data-set instead of individual examples. In this way, it is possible to ask questions that address the

---

8 Relevant volumes addressing the problem of miscellanies from multiple perspectives and languages are: Stephen Nichols and Siegfried Wenzel, eds., *The Whole Book. Cultural Perspectives on Medieval Miscellany* (Ann Arbor: University of Michigan Press, 1996); Keith Busby, ed., *Codex and Context. Reading Old French Verse Narrative in Manuscript* (Amsterdam/New York: Brill, 2002); Lucie Dolezalova and Kimberly Rivers, *Medieval Manuscript Miscellanies: Composition, Authorship, Use* (Krems: Medium Aevum Quotidianum, 2013).

9 Two excellent examples of this procedure are: Diana Müller, *Textgemeinschaften: der ‚Gregorius' Hartmanns von Aue in mittelalterlichen Sammelhandschriften* (Goethe Universität, 2011); Margit Dahm-Kruse, *Die Sammlung Als Kontext. Formen Der Retextualisierung Und Kontextualisierung Mittelhochdeutscher Versnovellen in Kleinepischen Sammelhandschriften Am Beispiel von Konrads von Würzburg 'Herzmaere'* (Tübingen: Narr Francke Attempto, 2018).

10 Miscellany usually refers to codices composed by several heterogeneous texts. For example, a manuscript of a long epic poem that includes in the last extra folios a short narrative text would not normally be considered a miscellany, although it certainly is a case of shared manuscript transmission. I will use the term 'miscellany' in this article rather loosely to mean a codex that compiles many texts.

overall structure of shared manuscript transmission. This would allow the particular research to be more accurate and meaningful, as they can be assessed against the general background.

The methodology presented in this article is based upon the realization that shared manuscript transmission can be modeled as a network, thus enabling the application of a series of mathematical measurements and clarifying visualizations. Some of the statistical methods presented in the following analysis do not rely necessarily on the network model, but are a complement that allows the model to be even more productive.

2 Sources: The Handschriftencensus Online Database

This article focuses on the underlying methods and theories to create a network of shared manuscript transmission, but the data from the online catalogue *Handschriftencensus* (HSC) is used as a test sample. This online database is the result of years of effort from a varied group of German scholars. It began around 2006 and only recently (since 2017) it has become a funded project at the *Mainzer Akademie der Wissenschaften und Literatur*. The project gathers information about medieval and early modern manuscripts containing Middle High German texts. Currently it records around 26,000 manuscripts and fragments between approximately 750 and 1520. From these, however, only about a quarter are high quality entries with a complete description. In some cases, the content of many folios is not identified, which imposes limits to the representativity of the analysis. The most famous and important works and manuscripts are described in detail, but more marginal or obscure texts might still be missing from the database. In spite of these limitations, HSC is by far the most complete and best organized database for the subject and is undergoing an active process of improvement that will yield very positive results in the coming years.

By reusing the HSC, the analysis inherits many of the criteria and biases of its underlying data model. This means, for example, that information regarding time and place of composition is not recorded in a standardized way that would make it machine readable. That information could be very useful to ask certain questions about manuscripts transmission, but they will need to wait until the data model is improved and updated.[11] Also, the criteria used in the HSC to define and incorporate materials is also inherited in the analysis. This includes the definition of manuscript, which is not as self-evident as it might seem at first, as quires that belonged together might have been taken apart at

[11] In a conversation with the team of the HSC in May 2019, they reported that an important restructuring of their website and data model is on the way, which will include standardization of and time and place of composition.

some point in the history of the object, or originally independent books might have been combined. In this regard, HSC does not offer an own definition of manuscript and tends to rely on previous decisions of archives and catalogues. There are also a huge number of written documents not included in the database: single leafs or small leaflets with public and private administrative documents. Excluding them is also a reasonable choice, as these documents are a different kind of object that would require a specific analysis. These kind of selections and exclusions are inevitable in any data model created and all databases will have certain issues and biases. The important thing is to be aware of them during the analysis and to assess if they could generate problems in the results.

Another important feature of the database to consider when evaluating the results is the way the literary works are identified. The foundation for defining a text is the *Verfasserlexikon*, the most important reference work in the field. For this reason, many different texts are condensed in one. For example, dozens of different works by the author Der Stricker are listed as *Stricker: Kleinere Reimpaardichtungen*. This editorial decision is perfectly valid for a reference work, but distorts the analysis proposed here. The resulting network will be profoundly affected by these editorial decisions.

Another issue to consider is that even the best database has underlying choices and biases created during the preservation of the objects. In other words, the currently available manuscripts do not represent the actual manuscript production during the medieval period and the loss rate is not randomly distributed. For example, big, luxurious books have a higher change of surviving. However, it is not impossible to draw conclusions about the general picture of medieval manuscript transmission from the status quo.[12] As will be shown, the network of shared manuscript transmission might even give us some insight into the effects of being part of miscellany manuscripts for the survival of a text.

3 Construction of the Network[13]

The basic principle of the network of shared manuscript transmission is that two texts that appear in the same manuscript context are connected. Within this general principle, there is room for some variations. In the first place, what is considered 'manuscript context' may change for different kinds of research. In the example for this article, I consider manuscript context as (generally) a

[12] Carmen Cardelle de Hartmann, „Überlieferungsprozesse: Sammeln – Auswählen – Kanonisieren. Eine Einführung", *Mittellateinisches Jahrbuch* 53, no. 1 (2018): 5.

[13] Images of the network in .svg format to explore in high detail with a web browser are stored in Zenodo: https://zenodo.org/record/2576509 | DOI: 10.5281/zenodo.2576509

physical codex – and more accurately, as one manuscript entry in the HSC database. However, for different kinds of analysis the manuscript context could be defined as a particular scriptorium, or a particular region at a particular time period.

Of course, these different criteria could be useful for specific research questions, but the most basic model is the one presented here. In this model, each text is a node. When two texts appear in the same physical codex they are connected with an edge. To represent if some manuscripts appear more than once together, a weight attribute for the edge is implemented. Figure 1 shows a sample part of the whole network to explain how it works:

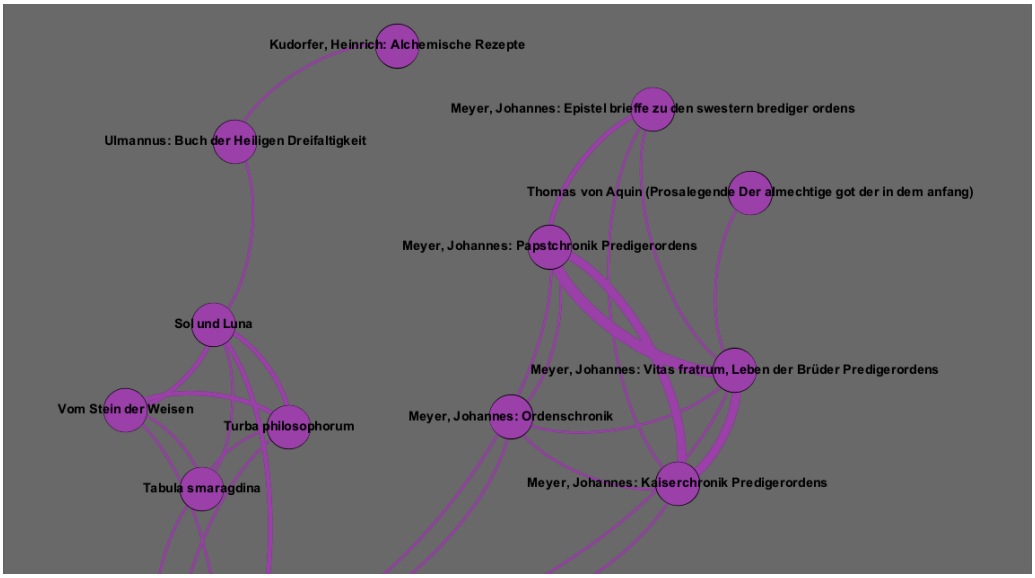

Figure 1. Excerpt of the network.[14]

In the example, one can see that *Alchemische Rezepte* by Heinrich Kudorfers appears in the same manuscript as Ulmannus' *Buch der heiligen Dreifaltigkeit*. The edge has a width of 1, which means that they only appear together in one manuscript. *Alchemische Rezepte* does not appear in the same manuscript context with any other work in the whole corpus, while the *Buch der heiligen Dreifaltigkeit* shares one manuscript with *Sol und Luna*. This work also shares manuscripts with *Vom Stein der Weisen*, *Turba philosophorum* and *Tabula smaragdina*. The many interconnections in this small cluster suggest that these four works must appear together in one manuscript (indeed: in Viena, Austrian National Library, Cod.

[14] Images of the network in .svg format to explore in high detail with a web browser are stored in Zenodo: https://zenodo.org/record/2576509, DOI: 10.5281/zenodo.2576509.

3001). As the width of the edges reflects the value of their weight attribute, it is possible to see that *Sol und Luna*, *Turba philosophorum* and *Vom Stein der Weisen* are attested together in more than one manuscript.

In this simple example, one can appreciate the most basic use of such a network as an exploratory tool. Understanding the shared manuscript transmission of this group of texts is easier with this network representation than navigating through the HSC database. This visualization allows scholars to understand how different texts relate to each other in a clear and simple way. This can also help discover connections that would be very hard to spot in a more traditional research environment. However, this exploratory use is just the tip of the iceberg of the possibilities of network analysis of manuscript transmission.[15]

To create the full network of shared manuscript transmission, the information from the website was stored in a *json* file with the following structure.[16]

> { 1: {'title': *Work Title*, 'manuscripts' : [*Ms1*, *Ms2*, *Ms3*] , 2 : {'title': *Work Title*, 'manuscripts' : [*Ms4*] , 3: {'title': *Work Title*, 'manuscripts' : [*Ms2*, *Ms4*, *Ms5*] , ...}

The HSC website does not provide, from the developers side, an API or other means of gathering the information in a structured way. However, the website allows webscraping with relative ease. The website includes a page with the index of all literary works in the database, which contains a link to the URL for each of them.[17] The first step to acquire the data is just to get a copy of all those URLs combined with the name of the texts. All these URL have a similar structure, ending in a number, which can be used as unique identifier for each text. Then a python program can iterate over that list, access each of the URLs, and retrieve the information about the manuscripts in which the work is attested. This is relatively easy as the manuscripts are listed in an *ol*-element with the attribute @class = 'signaturen'. The URL for manuscripts also have a unique identifier that can be used as id for each one of them.

Once the basic json file is created, a series of Python scripts were used to translate the data into the format required by the program chosen to create the network: *Gephi*.[18] Making a list of nodes is simple: the *script create-list-nodes.py* creates a CSV file with a row for each text in the original JSON file resulting from

---

15 In very convoluted parts of the network, a clear visualization can become difficult. In the future other programs and dynamic approaches, that only reveal selected part of the network should be used to improve what *Gephi* can offer.

16 This is true of the experiments that were used for this article, but I have recently replaced Json with XML to store the data.

17 http://www.handschriftencensus.de/werke.

18 Files and Python scripts used can be found on GitHub at the following URL: https://github.com/GusRiva/projects_data/tree/master/HSC

the webscraping, adding the title under the column *label* and the number of manuscripts as *rating*. Creating the list of edges is more complex. The scripts *create_list_manuscripts.py* and *create_list_manuscripts_with_works.py* reverse the original structure of the data: a new JSON file is created which lists the texts contained in each manuscript:

{*Ms1*: [*Txt1*, *Txt2*] , *Ms2* : [*Txt1, Txt3*] , *Ms3*: [*Txt2*, *Txt4*, *Txt5*] , ...}

Finally, the script *create_list_edges.py* writes the list of edges in a CSV file, each row indicating two nodes that need to be connected. For this, the Python package *itertools* is implemented in order to create all the size two combinations of the texts listed for each manuscript.

Gephi was the selected program to create the networks, because it yields very beautiful visualizations, has a multiplicity of algorithms to organize the layout of the network and offers the possibility to perform calculations of several features. Of course, other programs and tools could be used, like Neo4j, Python's *NetworkX* package or ORA-LITE. Those tools are more focused on mathematical analysis and less on visualisation and could be useful for future stages in the project, but, for now, the mathematical tools included in *Gephi* are adequate.

4 Analysis

There are many features of this created network that can be mathematically analyzed and offer important insight into the manuscripts transmission. In the following paragraphs, I will explain what these features are and why they are important, considering the network created with the HSC database as an example.

4.1 The Structure of Connected Components

In network analysis, a connected component is a subgraph in which any two vertices are connected to each other by paths. The number and size of connected components in a network of shared manuscript transmission reveals a great deal about the context in which those manuscripts were produced. A high number of connected components could mean that book and work tended to overlap, that most books contained one or few works in them and that miscellanies were not very common. It could also mean that the same groups of texts tended to be always transmitted together, that these groups were closed and well established and that there were strict limits to the possibilities of combining texts in the same codex. If, for the contrary, there are few or very big connected components, this means that compilation was common and there were no strict divisions due to,

for example, literary genre, that determined which works could be copied in the same book.

In the sample data from the HSC there is one huge connected component which makes up 76% of the nodes; a series of small connected components of 2 to 8 nodes (most with 2-4) which make up 6% of the total and many unconnected nodes which amount to 18%. This distribution can be seen in Figure 2. As this structure might be very typical for this kind of data and will be relevant for further analysis, it is useful to use specific names for each of these different 'zones' of the network and a geographical metaphor is highly appropriate. I propose to call the main connected component 'Continent', the small connected components 'Archipelagos' and the isolated nodes 'Islands'.[19]

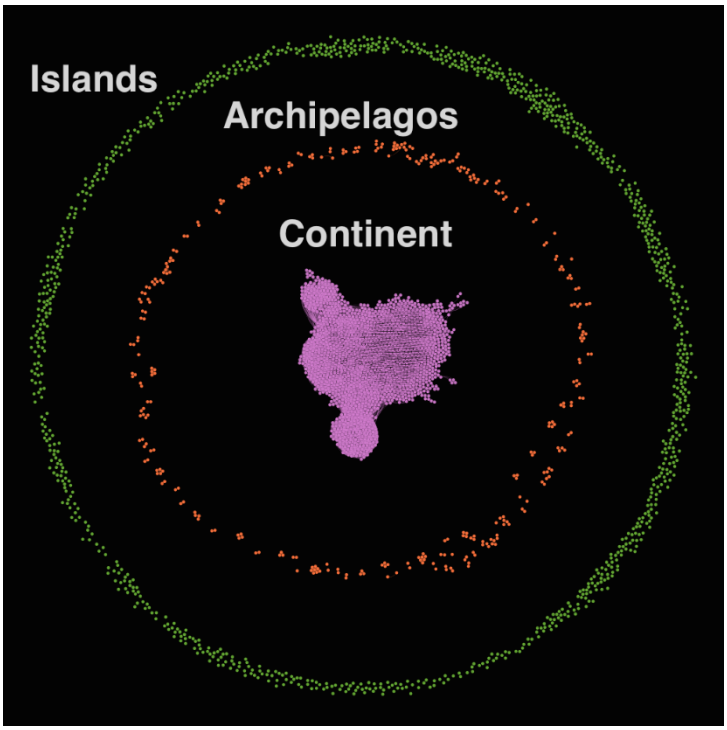

Figure 2. Full network with zones.

[19] The metaphor is not perfect, as archipelagos are actually collection of islands, while in this network islands can't be connected. However, I consider the names are very illustrative of the structure they try to explain if not taken the metaphor to the extreme.
The algorithm Force Atlas 2 contained in *Gephi* was used to structure the visualization. The islands and archipelagos were manually moved to their 'orbits' in the periphery of the image.

It is not hard to imagine that this structure might be repeated in other networks of shared manuscript transmission with similar features (French, Latin, Italian), however, it was not a priori to be expected. Considering that an important part of manuscripts in the database are fragments of books that might have originally contained more than one work, a tendency towards less connections in the network would be expected. However, there are so many miscellany manuscripts that they counteract the presence of fragments. The fact that there is only one big connected component including most works in the network suggests two things: 1- Compilation was the rule and not the exception. Most works appear at least some times in compilations. 2- There is no strict division of genres in the manuscripts (for example, religious and lay literature, prose and verse, etc.). There are at least some works that appear in different manuscript contexts and make such a big connected component possible.

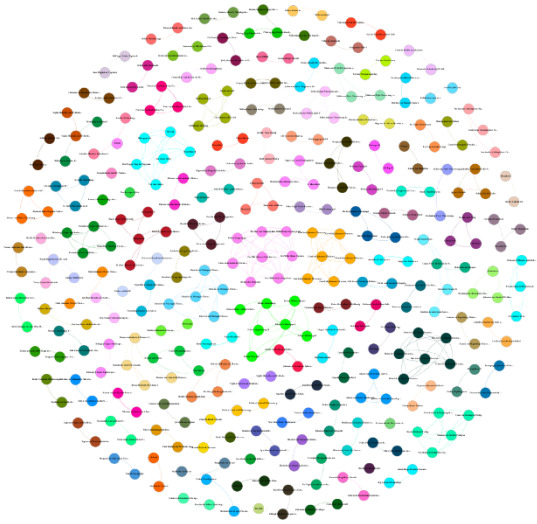

Figure 3. Archipelagos.

The most interesting analyses of such a network (for example, clustering and centrality, see below) focus exclusively on the continent, the biggest connected component, isolated from the rest and considered as a standalone network. However, a visualization of the archipelagos like the one offered in Figure 3, could be used for exploratory analysis.[20] The works in one of these small connected component must have, in most cases, a particular coherence and need to be understood as a particularly close group. In the sample data, 78% of all texts in this zone are also attested in only one manuscript. It could be

[20] The algorithm Fruchterman/Reingold contained in *Gephi* was used to structure the visualization.

argued that the analysis of any work included in one of these archipelagos needs to consider this shared manuscript transmission. As far as we know from the current evidence, the meaning of the individual texts in those groups depend on the other works that conformed their sole available context.

4.2 Degree

A node's degree is the number of links to other nodes. In a network of shared manuscript transmission, a high degree indicates a work which is most likely part of miscellany manuscripts. It would be expected that shorter works would tend to posses a higher degree, but this could not be tested on the dataset, as no information of length was included. Looking at the network there is, however, at least circumstantial evidence to assume that there is no real correlation between length and degree, as many important long works also have a high degree. In any case, degree is an important information to understand how a particular node fits within its manuscript context.

When considering degree, it is relevant to highlight the difference between 'islands' (degree = 0) and *unica* (works attested in only one manuscript). A text attested in many manuscripts, but always alone, is an island. A text attested in only one manuscript (*unicum*) but together with other works is not an island. However, trying to establish if there is some correlation between these two concepts in the data can be very insightful in order to understand the value of compilation for the survival of texts. In other words, if we find that most *unica* were actually attested alone, this means that works included in miscellany and other kinds of shared manuscripts have a higher chance of surviving. And indeed, in the dataset, 83% of islands are also *unica*, while only 40% of not islands are *unica* (and, as it's been mentioned before, 78% of archipelagos are *unica*). This interdependency of compilation and survival can be tested further considering if there is a correlation between degree and number of witnesses. When the Pearsons Correlation Coefficient is applied on the HSC data using degree and number of witnesses as variables, the results are r = 0.3935066925518164; p-value <0.001. [21] This suggests at least a weak correlation between the two variables, which would mean that works attested in compilations have a slightly higher chance of surviving. This is however a weak correlation, and it would be interesting to compare the results in other textual traditions.

[21] The Python function *pearsonr* from the S*ciPy* library was used to calculate this correlation coefficient. This function presupposes random chance as null-hypothesis (https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.pearsonr.html).

4.3 Clusters

One of the most fruitful consequences of modeling data as a network is the possibility to apply automatic community detection algorithms. These algorithms try to identify and differentiate clusters, i.e. groups of tightly interconnected nodes. These kind of clusters are relatively easy for humans to find in simple networks, but very complicated for complex networks. The community detection algorithms are also able to consider the weight of the edges, in order to give preference to the strong relationships over the weak ones. In this particular network, the nodes within the same cluster contain those texts that tend to be transmitted in the same manuscript context. In the network of the HSC data, it makes sense to isolate the continent to apply the community detection algorithm, as is shown in Figure 4. Each cluster detected is shown in a different color.[22]

Once clusters have been detected, it is worth asking if they can be explained; in other words, if there is a feature of the texts that explains why they are transmitted in the same manuscript contexts, or if it is just the result of chance. As the data has no labels for genre, nor place or time of composition, it is not possible to automatically compare if the clusters detected correlate with any of those attributes. However, a human inspection of the results suggests a strong overlap of communities and literary genres, as can be seen in Figure 5.[23] The correlation is not perfect as there are individual works or small groups of works that don't appear in the cluster expected according to their genre. A good example is the small group of *Minnereden* (late medieval texts in rhyming couplets, generally with little narrative content, about love) inside the purple community of mostly

[22] Depending of the variables used for the calculation the results may vary. In the example used here, standard parameters were used (Randomize/Use Weights/Resolution = 1.0). As a result, the modularity value is 0.654, and detects 15 communities. It would be possible to tweak the variables in order to detect more of less communities, but this number is appropriate for the analysis and offers a fair representation of the data.

[23] There are of course many different ways to classify the works written in Middle High German into genres. The objective here is not to propose a taxonomy of literary genres, but to showcase that the clusters in the network are not random and partially overlap with certain genres traditionally identified in literary history. A more detailed discussion on how medieval German literary genres are reflected in this particular network, as well as other specific issues that go beyond the scope of this article focused on methodology, will be carried out in a different paper.

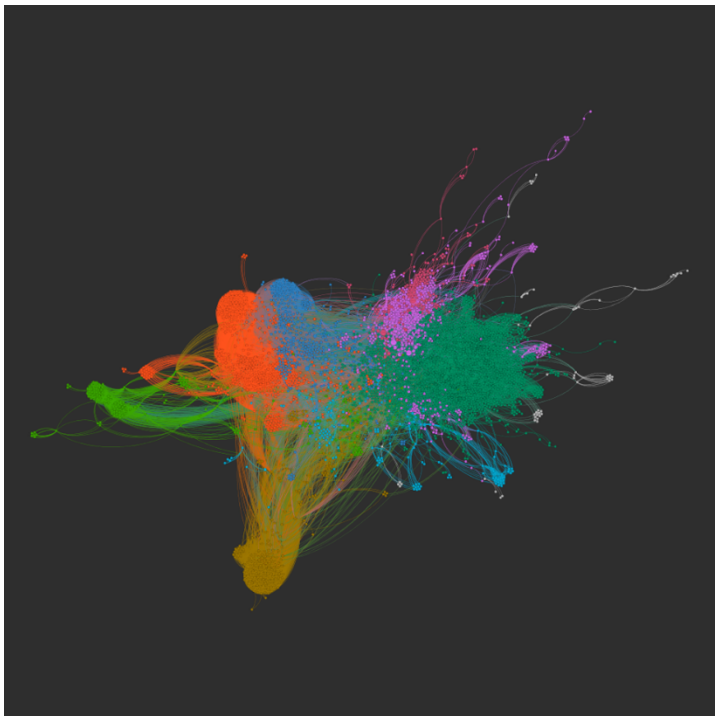

Figure 4. Clusters.

legal and sapiential literature, while most of the hundreds of *Minnereden* are either in the orange or light green clusters (see Figure 5). However, despite these exceptions, there is a clear tendency to group similar texts together. This suggest that the people responsible for creating medieval books applied some concept of genre when compiling different texts – a concept that corresponds, at least partially, with the modern classification of these works.

A virtue of the network model is that it allows to represent simultaneously that texts of the same genre tend to be transmitted together and that there is no strict division of genres. This means that at least some works can serve as nexus connecting the different clusters. The model enables the quantification of the connectivity and the clustering of the network and the identification of the works that perform as link between clusters, as will be shown in the next paragraph.

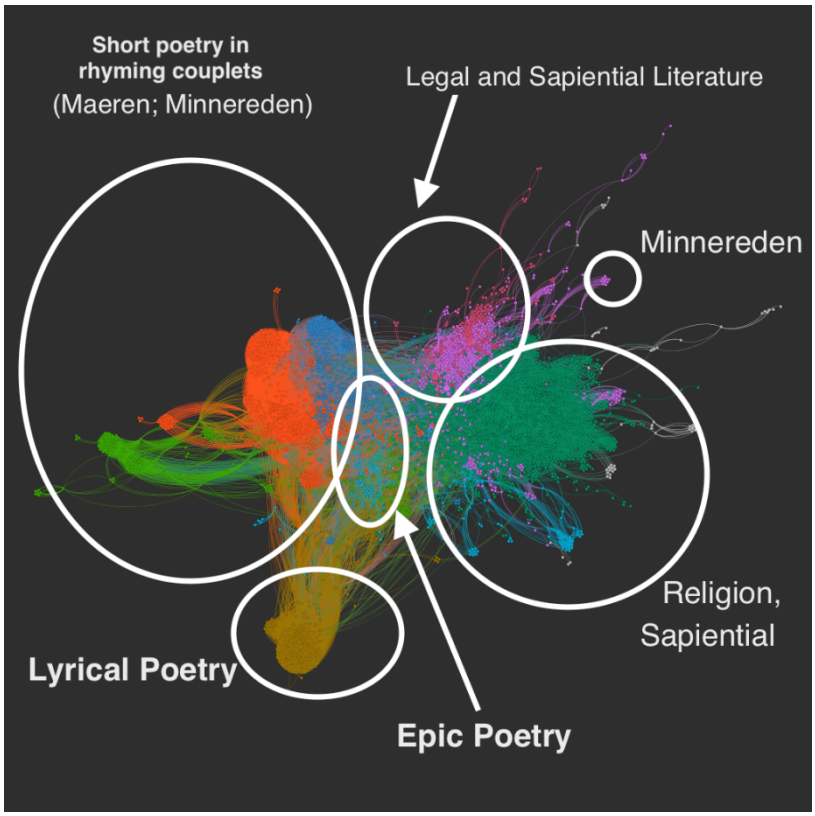

Figure 5. Clusters and genres.

4.4 Centrality

In the sample network, centrality measures only make sense when limited to the continent. There are many different measures of centrality that could bear valuable information, but it is necessary to understand how they might differ when modeling the phenomenon.[24] 'Eigenvector centrality' ranks the degree of a particular node and also considers the degree of the nodes to which it is connected. In this particular case, a high Eigenvector centrality would

[24] A general and informative discussion of network centrality can be found in: John Scott, *Social Network Analysis* (Los Angeles: SAGE, 2017), 95-106. There are many different concepts of centrality and different algorithms to calculate them. The two measurements implemented here, Eigenvector and betweenness, are common methods to calculate point centrality and represent two very different approaches. The first one focuses on a nodes' number and quality of connections; the second one, on its position in the graph. For a mathematical explanation see: Mark Newman, *Networks*. (Oxford: Oxford University Press, 2018), 159-176.

characterize works that fit particularly well in miscellany manuscripts. For this reason, Eigenvector centrality is a measurement of the compilation potential of a particular work and will tend to be high in short texts. 'Betweenness centrality' represents the frequency at which a node occurs on the shortest paths that connect every pair of points. In this sense, betweenness centrality is an excellent measure of what texts serve as nexus between different clusters of the network. Works with a high betweenness centrality are those which can easily be included in manuscripts with texts of different genres and help to create an interrelated network.

In the sample data the main results for both kind of centralities are similar (Figure 6 and Figure 7). In these Figures, nodes with a higher centrality value are bigger and red. Nodes with low centrality are smaller and yellow. The texts with highest Eigenvector and betweenness centrality are the mostly didactic poems of Freidank. These works are relatively short and fit any kind of manuscript context, regardless of genre, place and time of composition. In second place there is Cato, the german translation of the Disticha Catonis, also a didactic text, fundamental in the context of education during the Middle Ages. It also shares the feature of being able to be included in almost any manuscript context, as it is very famous and of a general didactic nature. It is also worth noting the two groups of texts with high Eigenvector centrality in the left part of Figure 6, which prove that high Eigenvector centrality means presence in big compilations. The upper group is mostly short narrative poetry in rhyming couplets (Versnovellen) which is usually transmitted in miscellany manuscripts. The lower group is lyrical poetry, that is mostly transmitted in song books.

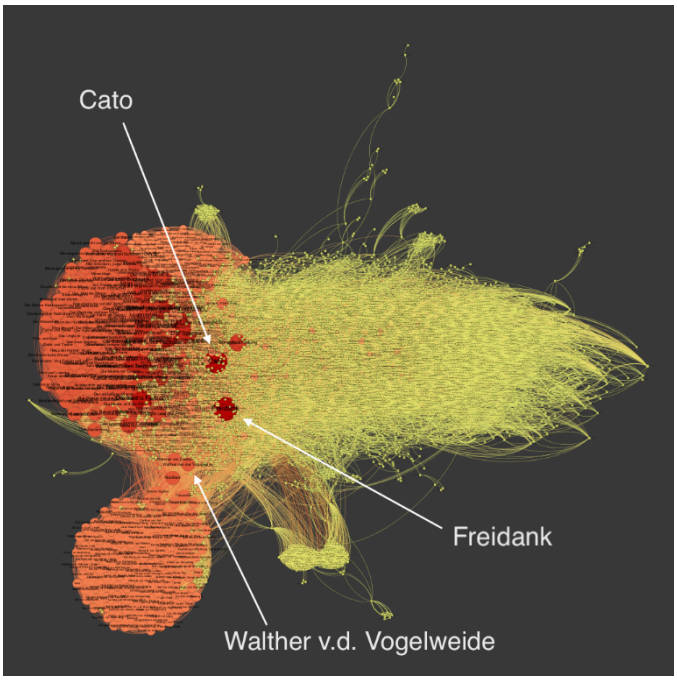

Figure 6. Eigenvector centrality.

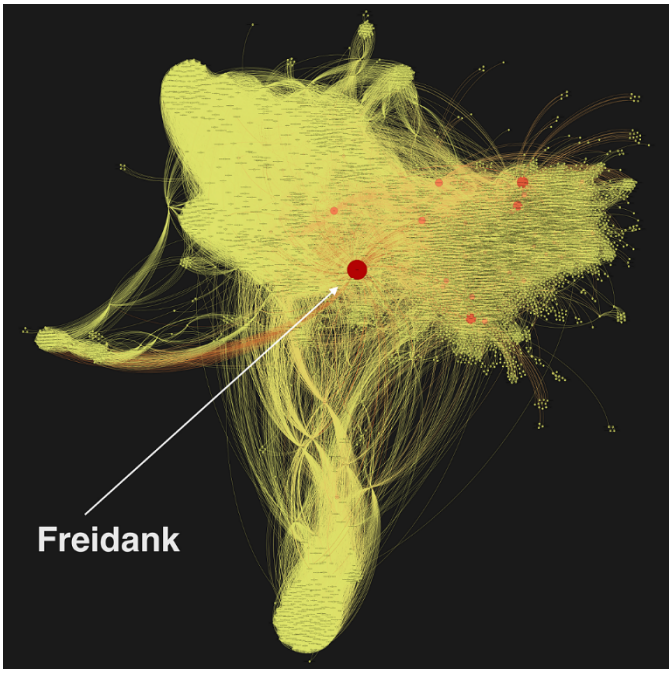

Figure 7. Betweenness centrality.

5 Conclusions

This paper displayed the basic principles and the advantages of modeling shared manuscript transmission as a network based on the particular example of medieval German manuscripts recorded in the HSC database. The model behind network theory permits not only a useful visualization for exploratory analysis, but also the implementation of different statistical tools and algorithms that shed light on the features of the underlying phenomena. Centrality measurements and community detection algorithms are the two main resources to assess how particular nodes or groups of nodes relate to the network as a whole. Other simple statistical methods (percentage of islands and archipelagos, percentage of unica) can complement the network analysis to create a better picture of the features of shared manuscript transmission.

In the future there are many possible ways to continue and improve this line of research. On the one hand, encoding metadata for genre as well as time and place of composition would allow to ask questions about clusters of nodes more efficiently. On the other hand, a comparative analysis of different manuscript traditions would be very fruitful. Discovering if manuscript transmissions in different languages, areas and times behave similarly would be very important to better understand the variation in the historical production and conservation of medieval manuscripts. Some measurement that don't reveal a lot applied on only one network (diameter, radius, average path) could be useful tools to compare different networks and assess their similarities and differences.

The use of network analysis is a promising strategy to investigate the shared manuscript transmission of medieval literature. Even considering the problems and biases in the available data, this model offers important insights into medieval textuality. This article is only a first attempt at implementation of the model. In the future, more detailed analysis of the particular network of German manuscripts are planned, as well as the creation of networks for textual transmission in other languages and for multilingual manuscripts. This perspective will hopefully complement other studies in the field and be a relevant tool to understand the practices and ideas that guided the creation of medieval manuscripts as well as their preservation and destruction.

6 Bibliography

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

Busby, Keith, ed. *Codex and Context. Reading Old French Verse Narrative in Manuscript*. Amsterdam/New York: Brill, 2002.

Clemens, Raymond, and Timothy Graham. *Introduction to Manuscript Studies*. Ithaca/London: Cornell University Press, 2007.

Dahm-Kruse, Margit. *Die Sammlung als Kontext. Formen der Retextualisierung und Kontextualisierung Mittelhochdeutscher Versnovellen in Kleinepischen Sammelhandschriften am Beispiel von Konrads von Würzburg 'Herzmaere'*. Tübingen: Narr Francke Attempto, 2018.

Dolezalova, Lucie, and Kimberly Rivers. *Medieval Manuscript Miscellanies: Composition, Authorship, Use*. Krems: Medium Aevum Quotidianum, 2013.

Görz, Günther, and Ursula Rautenberg. „Medienwechsel Bibliometrisch, Kurzrezension Zu Uwe Neddermeyer:..." *IASL Online*, 2001. https://www.iaslonline.lmu.de/index.php?vorgang_id=2334.

Gumbert, Johann Peter. "One Book with Many Texts: The Latin Tradition". In *Codices Miscellanearum. Brussels Van Hulthem Colloquium*, 27–36. Brussels: Archives et bibliothèques de Belgique, 1999.

Hartmann, Carmen Cardelle de. „Überlieferungsprozesse: Sammeln – Auswählen – Kanonisieren. Eine Einführung". *Mittellateinisches Jahrbuch* 53, no. 1 (2018): 1–10.

Kwakkel, Erik. *Books Before Print*. Arc Humanities Press, 2018.

Müller, Diana. *Textgemeinschaften : der Gregorius Hartmanns von Aue in mittelalterlichen Sammelhandschriften*. Goethe Universität, 2011. http://publikationen.ub.uni-frankfurt.de/opus4/frontdoor/index/index/year/2013/docId/30069.

Mutlová, Petra. "Networks of Ideas in the Czech Reformation". Presentation at the historical Network Research Conference 2018. Masarykova univerzita, Brno, 2018.

Neddermeyer, Uwe. *Von der Handschrift zum gedruckten Buch. Schriftlichkeit und Leseinteresse im Mittelalter und in der frühen Neuzeit. Quantitative und qualitative Aspekte*. 2 vols. Wiesbaden: Harrassowitz, 1998.

Neuheuser, Hanns Peter. „Rezension zu Uwe Neddermeyer: Von Der Handschrift..." *Bibliothek* 25, no. 1 (2001): 110–12.

Newman, Mark. *Networks*. Oxford: Oxford University Press, 2018.

Nichols, Stephen, and Siegfried Wenzel, eds. *The Whole Book. Cultural Perspectives on Medieval Miscellany*. Ann Arbor: University of Michigan Press, 1996.

Scott, John. *Social Network Analysis*. Los Angeles: SAGE, 2017.

Steinová, Evina. "Innovating Knowledge: Isidore's Etymologiae in the Carolingian Period". *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 2 (2019): 12-15.

Vozár, Zdenko. "Metadata for the Middle Ages: A Network Analysis of Manuscriptorium.com". Paper presented at the Historical Network Research Conference, Brno, 2018.