Exploring Cultural Heritage Knowledge Graphs – Case Correspondence Networks in Grand Duchy of Finland 1809–1917

Henna Poikkimäki[1,*], Petri Leskinen[1,2] and Eero Hyvönen[1,2]

[1]Aalto University
[2]University of Helsinki

Abstract

This paper argues for using methods and tools of Network Analysis (NA) to study contents of knowledge graphs (KG) in Digital Humanities (DH) research. As a case study, social and correspondence networks in the Grand Duchy of Finland 1809–1917 are considered with a focus on prosopographical data about historical people and, in particular, their correspondences (epistolary data). Letters have been an important form of communication, and networks based on letter metadata, contents and related biographical information can be used for rebuilding and analyzing historical social networks and for studying the flow of ideas and information. In correspondence network analysis, ego-networks focusing on only one person and his correspondents are common due to the nature of letter collections. Combining letter collections and biographical data helps move from ego-centric network approach towards sociocentric networks, as the larger network starts to emerge when letter collections from many individuals are brought together, although analyses still suffer from missing data. In this paper, we present results of the Constellations of Correspondence (CoCo) project that so far has created a KG of over million letters exchanged during 1809–1917 in Finland, re-using data of prosopographical KGs of the same period of time.

Keywords

correspondence networks, network analysis, knowledge graphs, linked data, digital humanities,

1. Introduction

Network analysis is a set of techniques derived from network theory, which has evolved from computer science to demonstrate the power of social network influences (Menczer, Fortunato, and Davis 2020). Using network analysis in an application domain analysis can add another layer of methodological triangulation by providing a different way to read and interpret the same data.[1] A knowledge graph (KG) (Ji et al. 2022) is a semantic network to represent and operate on data, where nodes are different types of entities and directed edges represent relations between them. The fact that KGs and NA are based on graphs suggests that NA can be used for studying

Digital Humanities in Nordic and Baltic Countries, Tartu, Estonia, March 5–7, 2025
*Corresponding author.
henna.poikkimaki@aalto.fi (H. Poikkimäki); petri.leskinen@aalto.fi (P. Leskinen); eero.hyvonen@aalto.fi (E. Hyvönen)
0000-0003-3362-8438 (H. Poikkimäki); 0000-0003-2327-6942 (P. Leskinen); 0000-0003-1695-5840 (E. Hyvönen)

Digital Humanities in the Nordic and Baltic Countries Publications – ISSN: 2704-1441
[1]Science Direct quote: https://www.sciencedirect.com/topics/social-sciences/network-analysis

data in KGs in natural ways. For example, the evolution of communities in influence networks of intellectuals have been studied using linked open data (LOD) set YAGO[2] (Petz, Ghawi, and Pfeffer 2022). However, overall, in spite of this obvious potential, there seems to be little research on applying NA to KGs.

This paper investigates this potential in the domain of Social Network Analysis (SNA), an application field of NA, where social structures are investigated by using networks and graph theory (Otte and Rousseau 2002). As a case study, Cultural Heritage KGs are studied with a focus on social relations and communications between historical people in the Grand Duchy of Finland 1809–1917. The paper tackles the following research questions: 1) *How to extract social networks from KGs?* 2) *How apply methods of NA to correspondence networks?* 3) *What kind of novel historical insights can be obtained in this way?*

The data for this research comes from KGs available on the Finnish Semantic Web infrastructure (Hyvönen 2024, 2023), including the Linked Open Data Services underlying CoCos (Tuominen et al. 2022) on epistolary data. This system is interlinked with the KGs underlying AcademySampo (Leskinen, Rantala, and Hyvönen 2022; Leskinen and Hyvönen 2021) and BiographySampo (Hyvönen et al. 2019; Tamper et al. 2023) on biographical and prosopographical data about people who lived during the same period of time, enriching the information available about the actors in the network. Our results and prototype implementations presented give novel insights into the social and correspondence networks in the Grand Duchy of Finland and suggest that methods and tools of NA are indeed useful in DH research. However, data literacy is needed for interpreting the results.

In Section 2, related works on applying NA to letter data are first outlined and characteristics and availability of epistolary datasets are discussed. After that, the CoCo KG used in this paper is described and a pipeline for constructing networks to be used in NA tools is presented (Section 3). Finally, we demonstrate the benefits of using the KG as a basis for NA by showing examples of applying NA to the CoCo KG (Section 4. In conclusion (Section 5), benefits, challenges, and future potential of the proposed approach are discussed.

2. Applying Network Analysis to Epistolary Data

In digital humanities, the NA has a wide range of applications from studying historical transport networks (Brookes and Huynh 2018) and transmission of manuscripts (Fernandez Riva 2019) to historical social network analysis (Erickson 1997). Here we focus on analyzing correspondence networks based on letters. Letters have been an important way to exchange intellectual, cultural and other information. Letters of some individuals have been preserved in public or private collections, often owned by cultural heritage institutions, such as archives, libraries, and museums. Analysing letter collections with NA methods can offer glimpses to historical social networks. Analyses of such networks can reveal central actors and communication patterns in the network, as well as what information was spread, among whom and when and how all of this correlates with known historical events (De Weerdt 2007).

In the case of correspondence networks, actors or nodes can be people, families, or organizations. These nodes are connected with edges that describe the letters sent between nodes. NA

[2] https://yago-knowledge.org/

can be divided into the ego-network approach, where the network is build around one actor called the ego, and whole network analysis that aims to give a full picture of social circles in question. The former has less severe data requirements and focuses on the quality of ties but can fail to acknowledge important ties between actors that are not the ego, while constructing the whole network includes these ties and can reveal phenomena that cannot be seen when focusing solely on individuals, but requires data about all actors in the network (Wetherell 1998; Morrissey 2015).

The simplest way to construct correspondence networks is use to letter metadata and form links between senders and recipients of the letters, and the number of letters between individuals can be assigned as edge weights. These networks allow for looking for central individuals or communication patterns. If the sending and/or receiving dates of the letters are present in the metadata, constructing dynamical networks or studying snapshots of correspondence networks during chosen time period can be done. In addition, contents of the letters and other biographical or archived information can be incorporated in the network, such as implications of relationships in the texts, topics of the letters, known relationships from biographies, mentions of people in question in newspapers, contents of their archived speeches and so on, if available (Edwards and Crossley 2009). Correspondence between different places, such as cities or villages can also be studied, if sending and receiving places of the letters are known (Riehle and Preiser-Kapeller 2020).

Like any digital analysis of historical data, NA suffers from fragmentation of historical records. Letters lost in history might take the attention away from otherwise interesting people, and it is very difficult to construct complete historical social networks. In addition to incompleteness of data, the differences in data quality as well as the data being stored into separate archives make analyzing the larger networks difficult. Missing letters might have been destroyed at some point, or people sending them or receiving them might not have been considered important enough for archiving or digitizing their letters. NA can fail to point out individuals belonging to some minority, but known to be important historically due to the latter reason (Moravec 2017). For example, Düring constructed social network between helpers and refugees in Berlin during the Holocaust and concluded that 20–30% of known influential actors are missing from the top 20% highest centrality scores, but while NA methods fail to recognize these individuals as central actors, centrality measures can still be used to narrow down potentially influential actors (Düring 2016). Moreover, as the amount of easily available digitized data grows, computer-based tools are useful in supporting traditional humanities tasks, such as close reading (Fiscarelli 2022).

The effects of the missing data in historical social networks can be minimized by focusing on ego-networks of individuals for whom there is a sufficient amount of information available (Riehle and Preiser-Kapeller 2020), or combining data from multiple sources to fill in the gaps in the whole network (Edwards and Crossley 2009). Depending on the research questions, the ego-network approach might be even preferred: if the life of one person is studied, ego-networks might be easier to construct. The network to be studied is then smaller, analyses require less memory and computational power, and there are less letters for close reading. For example, how the ego-networks of people living in the Grand Duchy of Finland evolve and affect the ego can be studied by concentrating on people, for which sufficient amounts of the data is available. By using linked data, where people and letters can be linked to external databases,

filling gaps caused by missing letters becomes easier as the biographical data is more easily available. Studying ego-networks before and after some major life events can also reveal how one's personal social circles change due to these events (Perry, Pescosolido, and Borgatti 2018). On the other hand, the downside of using ego-networks is that potentially interesting and important ties between other nodes might be left out.

Another problem while constructing historically significant social networks is to recognize ties that were truly significant for the people in question: marriage or legal contracts between two actors clearly have an impact on their life, but letters sent between two actors might not imply significant relationship between them (Morrissey 2015). However, in the context of correspondence networks the significance of the relationship can be determined by close reading of the letters, or the significance of ties can be expressed with tie weight that, e.g., corresponds to the number of letters, assuming the more two individuals have sent each other means the more closely they are related to each other. Similarly, a tie weight can represent the strength of the relationship also in other types of networks. Many NA methods, such as centrality measures have been modified to weighted networks (Bellingeri et al. 2023).

Networks based on epistolary data have also been compared to contemporary communication networks: the prominent characteristics of contemporary communication networks are present in correspondence networks but there are also differences that might be caused by incomplete historical data or differences by the use of media and communication practises (Ureña-Carrion et al. 2022). In communication network analysis, communication roles for nodes can be detected based on the topological properties of the node. This has been done to study the evolution of communication roles of the Protestant Reformers based on correspondence networks during 1500–1565 (Roller 2022), to recognize anomalous behaviour in the Tudor letter networks (Ahnert and Ahnert 2019) or to study how the topology of the network affects the success of recommendation letters (Gerlach and Blumenthal 2023). If networks constructed from letters are considered as large-scale communication networks rather than historical social networks, the significance of ties, if the people were actually close in real life, is not necessarily a requirement and only letter metadata is needed for constructing the networks. For example, in mobile communication networks where the nodes are anonymous people and link strength corresponds to the number of minutes people have talked to each other, the focus is more on the communication patterns (Onnela et al. 2007).

Data sources of early Early Modern learned correspondences are proliferating rapidly, including, e.g., The Catalogus Epistularum Neerlandicarum[3], Early Modern Letters Online (EMLO)[4] (D. v. Miert 2008; Heuvel 2015; D. v. Miert 2016), Electronic Enlightenment[5], ePistolarium[6] (Ravenek, Heuvel, and Gerritsen 2017), Europeana[7] (Doerr et al. 2010; Freire et al. 2019), the Mapping the Republic of Letters project[8], Kalliope Catalogue[9], SKILLNET[10], and correspSearch[11].

[3] http://picarta.pica.nl/DB=3.23/
[4] http://emlo.bodleian.ox.ac.uk
[5] http://www.e-enlightenment.com
[6] http://ckcc.huygens.knaw.nl/epistolarium/
[7] http://www.europeana.eu
[8] http://republicofletters.stanford.edu
[9] http://kalliope.staatsbibliothek-berlin.de
[10] https://skillnet.nl
[11] https://correspsearch.net

The CKCC[12] corpus stands as a Dutch counterpart to the Republic of Letters, encompassing a substantial collection of approximately 20 000 correspondences (Heuvel 2015; D. v. Miert 2016; Hyvönen, Leskinen, and Tuominen 2023). The *correspSearch* dataset, compiled at the Berlin-Brandenburg Academy of Sciences and Humanities, encompasses approximately 150 000 letters that have undergone scholarly editing, featuring published summaries, transcriptions, and possibly commentaries (Dumont 2016). Visualizing the epistolary data is studied in *Mapping the Republic of Letters* project[13], in *Tudor Networks of Power*[14], and the *LetterSampo – Historical Letters on the Semantic Web*[15] system, a semantic portal and LOD service that aggregates Republic of Letters and correspSearch datasets, the forerunner of the CoCo project (Hyvönen, Leskinen, and Tuominen 2023; Leskinen et al. 2023). Bruneau et al. explore the application of Semantic Web Technologies to model the correspondences of French scientist *Henri Poincaré* and publish them on an online platform[16] (Bruneau et al. 2021). In the case of Finland, there are collections of correspondences by renowned cultural influencers, such as *Letters of Edelfelt*[17], *Elias Lönnrot Letters*[18], and *J. V. Snellman* (Eskelinen et al. 2001).

Efforts like ours in the CoCo project[19] that aggregate letters from multiple collections to one repository using the same data model and vocabularies make constructing and studying the whole network easier, although the problems with missing letters and varying letter metadata quality still exist. CoCo data offers links to biographies that can be used when constructing historical social networks mainly based on letters, but lacks the contents of the letters. A semantic portal based on CoCo data and Sampo model[20] (Hyvönen 2023) makes also searching, browsing, and analyzing the data easier that can be of great help when interpreting the results, especially when the number of letters is large.

3. Data and Constructing Correspondence Networks

This section presents the data used in our case study and how networks were constructed from LOD. The data will be published openly in the same way as the Sampo systems in 2025 as the CoCo research project (2022–2025) ends.

3.1. CoCo Knowledge Graph

The data used in our study is a collection KG of epistolary metadata of the CoCo project regarding the period of the Grand Duchy of Finland (1809–1917). As the letter correspondences form the backbone in the dataset, it is further enriched by using external LOD publications, including

[12] CKCC is an acronym for *Circulation of Knowledge: A Web-based Humanities' Collaboratory on Correspondences and Learned Practices in the 17th-century Dutch Republic*
[13] Mapping the Republic of Letters: http://republicofletters.stanford.edu/
[14] Tudor Networks online: http://tudornetworks.net/
[15] LetterSampo project: https://seco.cs.aalto.fi/projects/rrl/
[16] The Correspondence of Henri Poincaré: http://henripoincare.fr/s/correspondance/page/accueil
[17] Albert Edelfelts brev, Svenska Literatursällskapet i Finland: https://edelfelt.fi/
[18] Elias Lönnrot Letters Online, Suomalaisen Kirjallisuuden Seura: http://lonnrot.finlit.fi/omeka/
[19] CoCo project: https://seco.cs.aalto.fi/projects/coco/
[20] Sampo series of LOD services and portals: https://seco.cs.aalto.fi/applications/sampo/

interlinked KGs in AcademySampo[21], BiographySampo[22], ParliamentSampo[23], Wikidata[24], and Getty ULAN[25]. The external datasets provide additional information like the places and times of birth and death, occupations, and interpersonal relations of the actors. Since the data is assembled from multiple data sources, the amount of available metadata varies a lot. Besides usually knowing the names of the sender and the recipient and the date of sending some data sources provide additional metadata like the places of sending or receiving, language used in the letter, or the letter content maybe further including references to other people. (Tuominen et al. 2022)

The project is ongoing and it currently contains over million letters from 12 distinct data sources, over 90 000 actors and metadata related to letters and actors. About 90% of the actors have type 'Person', 7% have type 'Group', 0.2% are of type 'Family' and the rest are 'Unknown'. Out of 'Person' actors, 9.8% are linked to AcademySampo, 6.7% to BiographySampo and 13.9% to Wikidata. In the dataset, 78% of the people have sent and/or received only one letter and 97% have sent and/or received under 10 letters. Out of people that have sent and/or received over 10 letters, 47% are linked to AcademySampo, 50% to BiographySampo and 64% to Wikidata, and they have more often other metadata, such as occupation or birth and death dates available derived from those external sources.

Major problem for temporal network analysis is the poor accuracy of sending dates of the letters. Most of the letters (70%) have a range of some years assigned as a sending date. For about 26% the exact year of sending is known and only about 1.7% of letters have the exact sending date available. Rest of the sending dates are missing or unclear. Missing metadata of letters and actors sets some limitations to what kind of analyses can be done.

3.2. Extracting Correspondence Networks from a Knowledge Graph

CoCo KG is based on domain specific data model, extended from that in use in the LetterSampo system (Hyvönen, Leskinen, and Tuominen 2023), based in the Resource Description Framework (RDF) (Decker, Mitra, and Melnik 2000), where data consists of subject-predicate-object triples. An RDF KG is a directed graph in which the subject and object correspond to nodes and the predicate is link that points from subject to the object.

We created networks from CoCo KG using the SPARQL query language and the Jupyter Notebook based Google Colabotory service (Colab)[26] which allows writing and running Python Code in the browser. For querying the data from the SPARQL-endpoint, we used SPARQLWrapper Python package[27] that can be used to query and process the data in Python code. Networks are constructed and analyzed using the NetworkX Python package (Hagberg, Schult, and Swart 2008). Our method and tool Sparql2GraphServer are described in more detail in (Leskinen, Hyvönen, and Tuominen 2021).

[21] AcademySampo – Finnish Academic People 1640–1899: https://www.ldf.fi/dataset/yoma
[22] BiuographySampo: https://www.ldf.fi/dataset/nbf
[23] ParliamentSampo: https://www.ldf.fi/dataset/semparl
[24] Wikidata: https://www.wikidata.org/
[25] Union List of Artist Names: https://www.getty.edu/research/tools/vocabularies/ulan/
[26] Google Colab: https://colab.research.google.com/
[27] SPARQLWrapper: https://github.com/RDFLib/sparqlwrapper.git

We construct a directed network where nodes are senders and/or receivers of letters edges that point from sender node to the receiver node are weighted with the number of letters. First we query edges of the network, i.e. all letters and their senders and receivers. The sender-receiver pairs are then grouped, number of letters from sender to receiver counted and letters from sender to receiver listed. After that metadata such as labels or birth and death years are queried for nodes, and sending dates and places for letters if available. Data returned from queries is in JSON-format and is transformed into suitable data types. NetworkX supports different types of graphs, and allows storing optional node and edge metadata on the graph object. We add nodes and related metadata to the graph, and then edges and related metadata.

The ego-network can be easily obtained from the node network using methods implemented in NetworkX, first we get all the successors(actors who have received letters from ego node) and predecessors (actors who have sent letters to ego node) of the ego node and then we get subgraph from the original network containing those nodes. Alternatively, edges and nodes for ego-network can be queried separately. In the case of ego-network we keep all types of nodes, in the case of the "whole" we keep only nodes of type 'Person' who have exchanged letters with other 'Person' types.

4. Analyzing Correspondence Networks

This section offers examples on how NA can be used to study correspondence networks using CoCo data. We show the "whole" network approach that includes all the people in the data who have sent letters to other people or received letters from others, and an ego-network approach to the personal network of Elias Lönnrot. The analyses were done using a Google Colab notebook that will be published openly on the project website after data publication.

4.1. Correspondence Network in Grand Duchy of Finland

In order to study the correspondences, we constructed using SPARQL a directed network from the CoCo KG where nodes correspond to people that have sent and/or received at least 1 to other people and edge weights correspond to number of letters sent. This kind of analyses have not been feasible before as the original letter data has been distributed in different collections and has been available only as documents represented using Dublin Core-like metadata with literal un-linked metadata element values. In the CoCo KG, in contrast, persons and places, for example, are mutually aligned resources with URI identifiers making network construction for NA tools and visualizations possible on a global Finnish level.

Network contains 76 044 people and edges are based on 903 654 letters from 11 different data sources. 68% of the people in the network are men, 26% are women and for the rest the gender is unknown. 74 562 people belong to the largest weakly connected component and in addition 629 other weakly connected components, out of which the size of the second largest component is 33 and most of them are of size 2. Some of the smaller components are built around families. There are no letters sent or received between people belonging to different weakly connected components. There are 3681 people who have not sent any letters and 66 654 people who have not received any letters. Table 1 shows 10 pairs of the people between whom there are most letters in the data set. Many of the node pairs with a lot of letters between them

are family members. This indicates that correspondence between family members was possibly more frequent and/or those letters were more likely preserved.

# letters	name	name
2526	Thuneberg, Otto Ivar August	Thuneberg, Lilli
2466	Thuneberg, Otto Ivar August	Thuneberg, Axel
2235	Procopé (Reuter), Aline	Reuter, Anna Hildur Elisabeth
1884	Mattsson, Maria Elisabeth	Mattsson, Gustaf Otto (Guss)
1865	Roos (Backmansson), Emmy	Roos, Elias Alfred
1817	Achté, Emmy	Ackté, Aino
1655	Wrede, Carin Emilia Augusta	Wrede, Hedvig Gustava Matilda
1619	Edelfelt, Albert	Edelfelt, Alexandra
1532	Tham (Ramsay), Emmy Beata Catharina	Ramsay, Sofia
1518	Stenbäck (Hultman), Emmy Maria	Ramsay, Sofia

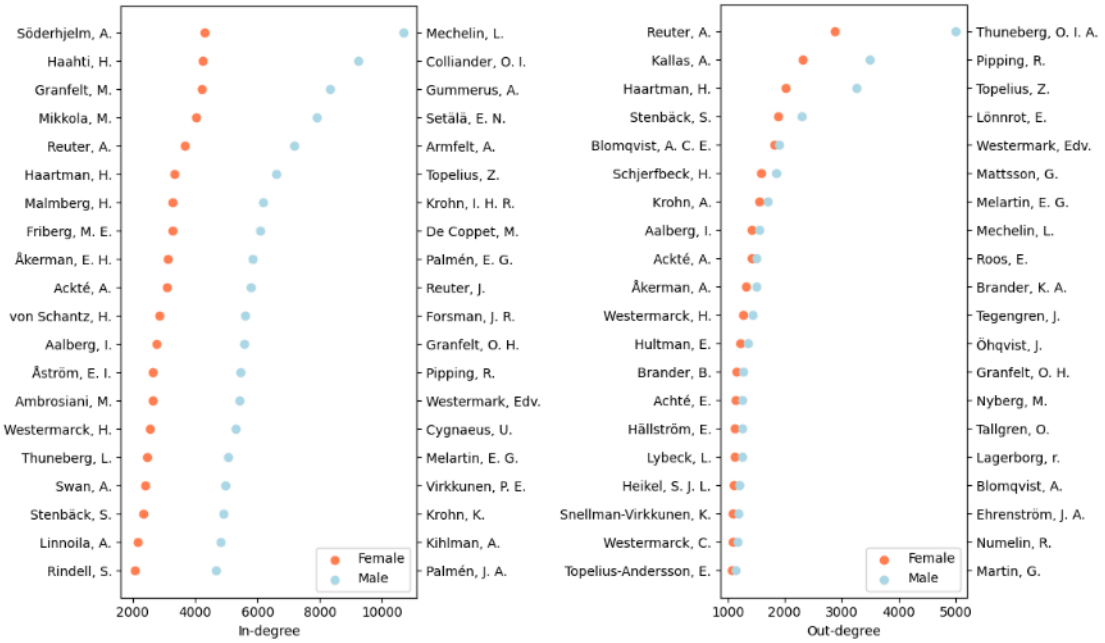
Table 1
Ten pairs of people with most letters sent between them.

Figure 1 shows people with highest weighted in- and out-degrees in the network or people who have sent and received the most letters correspondingly. The gap between genders is greater in high in-degrees, but out-degrees of highest ranking men and women are quite close to each other. In total, women have sent over 30 000 letters from which 49% were sent for other women and 50% for men, and the rest for people whose gender is unknown. On the other hand, men have sent almost 95 000 letters from which 87% are sent for other men and only 11% for women. The gap between in-degrees of men and women might be caused by the more prominent position of the high in-degree men who often have worked as politicians, diplomats, officials or professors among other things, or the letters received by men might be more likely archived and preserved. In general, highest in-degrees are higher than out-degrees, implying together with a large number of people who have not received any letters that the letters sent from a broad set of people are somewhat concentrated towards a smaller number of receivers. Also some people in figure 1 are also in table 1 (e.g. Otto Ivar August Thuneberg and Aline Reuter), indicating that their high in- or our-degree comes mostly from correspondence between family members.

4.2. Ego-network of Elias Lönnrot

As an example of studying an ego-network, we next focus on the ego-network of Elias Lönnrot (1802–1884) extracted from the CoCo KG. Lönnrot was a Finnish polymath who worked in the fields of medicine, literature, and botany among other things. He is best known for compiling the Finnish national epic Kalevala based on poems that he gathered during several expeditions. Lönnrot's letters originate mainly from the Archives of the Finnish Literature Society, but also from some other collections aggregated in the CoCo KG, and constitute one of the best curated parts of the CoCo data collected from different sources. The data contains 3134 letters Lönnrot sent to 498 different actors, 3314 letters he received from 763 different actors, and 94% of the letters have the exact sending date available.

Fig. 2 shows the number of letters Lönnrot sent and received per year during his life. During



(a) Highest in-degrees among males and females. (b) Highest out-degrees among males and females.

Figure 1: Men and women with highest in-degrees and out-degrees. Names of women are on the left side of the subfigure and names of men on the right side.

his early life and studies until 1832, the number of letters was quite low. After 1832, the number of letters increased in the dataset. This is likely to relate to his graduation in 1832, a piece of information available in, e.g., the interlinked AcademySampo and BiographySampo data. Lönnrot worked first as a doctor in Oulu, and from 1833 until 1854 in the city of Kajaani. From 1853 until his retirement in 1862 Lönnrot worked at the University of Helsinki as he was appointed the Chair of Finnish Literature there. For some reason, the number of letters Lönnrot sent decreased during that time, and during that time Lönnrot received more letters than he sent. A reason for this might be Lönnrot's more central location, or that letters sent by Lönnrot during this time are missing. Lönnrot also did go on various expeditions during 1828–1845, and worked also on other fields, such as journalism.

Figure 3 shows the number of letters between Lönnrot and other letter writers (on the y-axis), as well as the balance between them (on the x-axis). Balance of 0.0 tells that one actor has sent all of the letters and when balance is 1.0 actors have sent equally many letters to each other (Li 2018). In the figure, blue node color tells that Lönnrot has sent all or most of the letters that have been sent between Lönnrot and the other letter writer, and yellow tells that other writer has sent most of the letters. The balance of 1.0 is represented by the green color in the figure. In upper right corner there are actors who sent many letters to Lönnrot and also received many letters from Lönnrot, such as Frans Johan Rabbe, who was also a doctor and worked for "Lääkintöhallitus" (Healt Institute of Finland) during that time, and Carl Gustaf Borg, who worked during same time at the University of Helsinki as Lönnrot. In the upper left side, where the correspondence

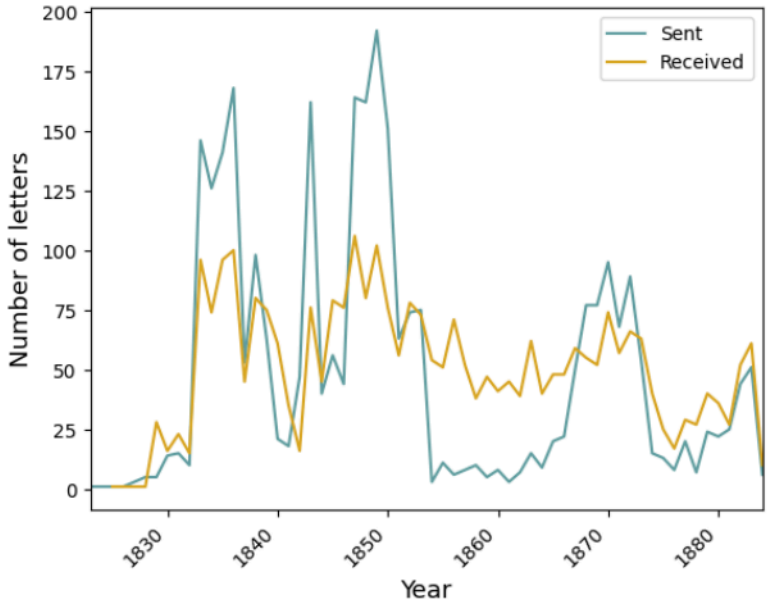


Figure 2: Number of letters sent and received by Elias Lönnrot (1802–1884)

was unbalanced, are groups like the "Suomalaisen Kirjallisuuden Seura" (Finnish Literature Society) for which Lönnrot has sent a lot of letters. Balance of 0.0 can also reveal missing links. For example, according to figure 3 Lönnrot has sent over 25 letters for his supporter Eva Agata Törngren, but how likely it is that Törngren never answered his letters? Deeper analyses are needed with close reading, but the point here is that computationally obtained new analyses – like the ones above – can alert the humanist researcher about potentially interesting historical phenomena for further study.

5. Discussion

This paper discussed benefits and downfalls of applying NA to correspondence networks and provided a few examples of analyses using CoCo KG linked dataset. The paper argued that by aggregating and harmonizing data from distributed collections into a centralized KG, and by transforming it into network structures needed for NA tools, it is possible to analyze correspondence networks both on different levels and scales in new ways. The approach presented can provide humanist researchers novel historical insights into the data. This argument was supported by presenting example analyses of the whole CoCo KG network between over 75 000 people, and of an ego-centric network of a particular prominent person, i.e., Elias Lönnrot.

In analyzing epistolary data, including the CoCo KG, the quality of the metadata determines what kind of networks are possible to construct. For example, the lack of accurate sending dates or even sending years makes studying the historical changes happening during the years in the larger network difficult and unreliable. Studying well-curated subsets of the data, such as temporal networks of individuals is possible with more reliable results. However, in all cases

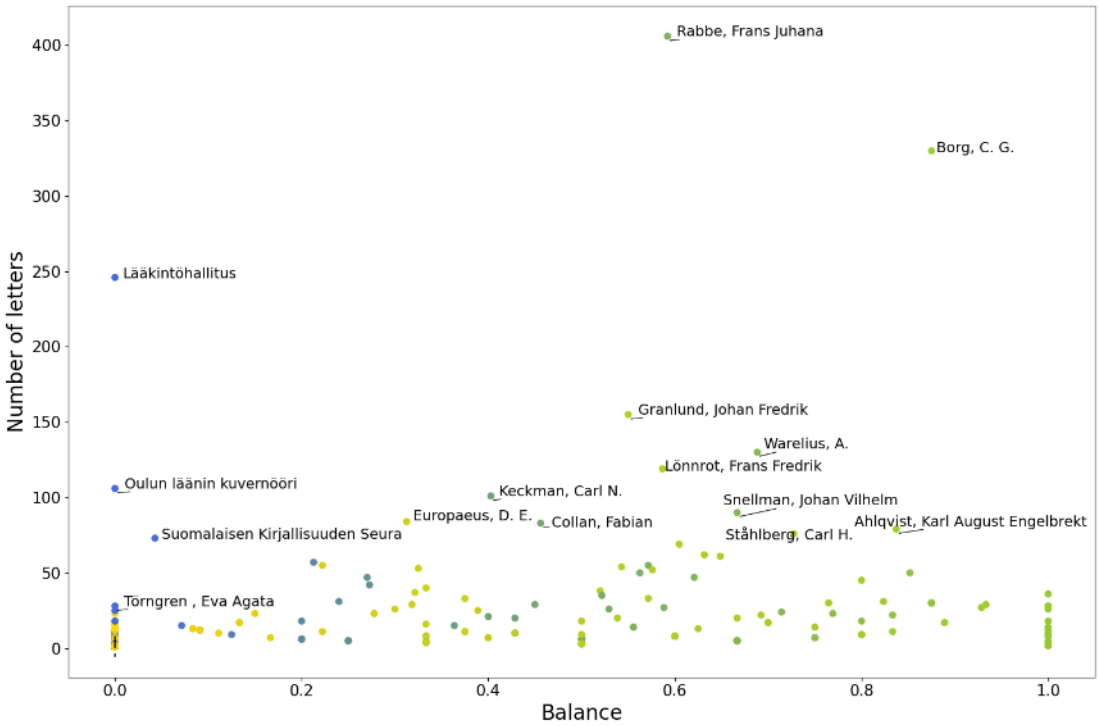


Figure 3: Number of letters and balance between Lönnrot and his correspondents

data literacy (Koltay 2015) is needed for interpreting the results.

Also, benefits of using linked data as the basis for constructing networks was discussed. As KGs are directed graphs by nature, different connections between entities can be queried. Using linked data enriched by other sources is valuable when interpreting the results of NA and in the case of the correspondence networks it can help to fill in the gaps caused by missing letters. Combining letters from multiple data sources like in the CoCo project, allows constructing and studying "whole" networks, instead of only ego-networks, although important nodes and links might still be missing.

Although missing letters and differences in metadata quality limit the possibilities of constructing more comprehensive networks, like networks that take into account sending dates and places of the letters, our example analyses that only scratch the surface of correspondence networks in the Grand Duchy of Finland show that such analyses can still reveal potentially interesting historical phenomena for the humanist researcher for further study. For example, what communities are isolated from the whole network as smaller weakly connected components, how correspondence between family and friends differs from correspondence between other connections, why the number of letters Elias Lönnrot sent decreases when he worked for the University of Helsinki and how did his ego-network evolve over time?

As for further research, combining more prosopographical data into the analyses from the interlinked Sampo portal KGs, such as AcademySampo and BiographySampo, can be studied. Also pointing out potential missing links and bringing in other information based on external

biographical data and NA methods would increase the quality of the networks. We also consider extracting subject matter metadata from the actual letter texts and using letter content as part of the analyses.

Acknowledgements Thanks to Jouni Tuominen, Ilona Pikkanen, and other co-workers in the CoCo project for creating the CoCo KG and fruitful discussions. This work was funded mainly by the CoCo project supported by Research Council of Finland. Partial funding was received from the European Union – NextGenerationEU instrument under grant number P3C3I6 for the national FIN-CLARIAH/DARIAH-FI initiative[28]. Computing resources provided by the CSC – IT Center for Science were used in our work.

References

Ahnert, Ruth, and Sebastian E Ahnert. 2019. "Metadata, Surveillance and the Tudor State." *History Workshop Journal* 87 (January): 27–51. ISSN: 1363-3554. https://doi.org/10.1093/hwj/dby033. eprint: https://academic.oup.com/hwj/article-pdf/doi/10.1093/hwj/dby033/28296164/dby033.pdf.

Bellingeri, M., D. Bevacqua, F. Sartori, M. Turchetto, F. Scotognella, R. Alfieri, N. K. K. Nguyen, T. T. Le, Q. Nguyen, and D. Cassi. 2023. "Considering weights in real social networks: A review." *Frontiers in Physics* 11. ISSN: 2296-424X. https://doi.org/10.3389/fphy.2023.1152243.

Brookes, Stuart, and Hoai Nguyen Huynh. 2018. "Transport networks and towns in Roman and early medieval England: An application of PageRank to archaeological questions." *Journal of Archaeological Science: Reports* 17:477–490. ISSN: 2352-409X. https://doi.org/https://doi.org/10.1016/j.jasrep.2017.11.033.

Bruneau, Olivier, Nicolas Lasolle, Jean Lieber, Emmanuel Nauer, Siyana Pavlova, and Laurent Rollet. 2021. "Applying and Developing Semantic Web Technologies for Exploiting a Corpus in History of Science: the Case Study of the Henri Poincaré Correspondence." *Semantic Web* 12 (2): 359–378. https://doi.org/10.3233/SW-200400.

De Weerdt, Hilde. 2007. "Mapping Communication from Mingzhou: Networks of Correspondence." In *Prosopography of Middle Period China: Using the Chinese Biographical Database" Workshop, Warwick.*

Decker, Stefan, Prasenjit Mitra, and Sergey Melnik. 2000. "Framework for the semantic Web: an RDF tutorial." *IEEE Internet Computing* 4 (6): 68–73.

Doerr, Martin, Stefan Gradmann, Steffen Hennicke, Antoine Isaac, Carlo Meghini, and H. Sompel. 2010. "The Europeana Data Model (EDM)." *World Library and Information Congress: 76th IFLA General Conference and Assembly* (January): 10–15. https://cdn.ifla.org/past-wlic/2010/149-doerr-en.pdf.

Dumont, Stefan. 2016. "correspSearch -– Connecting Scholarly Editions of Letters." *Journal of the Text Encoding Initiative* 10. https://doi.org/10.4000/jtei.1742.

[28] FIN-CLARIAH/DARIAH-FI, Linked Open Data part: https://seco.cs.aalto.fi/projects/fin-clariah/

Düring, Marten. 2016. "How Reliable are Centrality Measures for Data Collected from Fragmentary and Heterogeneous Historical Sources? A Case Study." In *The Connected Past: Challenges to Network Studies in Archaeology and History.* Oxford University Press, March. ISBN: 9780198748519. https://doi.org/10.1093/9780198748519.003.0011. eprint: https://academic.oup.com/book/0/chapter/346732411/chapter-pdf/43164862/isbn-9780198748519-book-part-11.pdf.

Edwards, Gemma, and Nick Crossley. 2009. "Measures and Meanings: Exploring the Ego-Net of Helen Kirkpatrick Watts, Militant Suffragette." *Methodological Innovations Online* 4 (1): 37–61. https://doi.org/10.1177/205979910900400104. eprint: https://doi.org/10.1177/205979910900400104.

Erickson, Bonnie H. 1997. "Social Networks and History: A Review Essay." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 30 (3): 149–157. https://doi.org/10.1080/01615449709601182.

Eskelinen, H., T. Helén, J. Kajander, M. Kinnunen, S. Linnavalli, V. Oittinen, E. Ojanen, et al. 2001. "J. V. Snellman: Kootut teokset 1–24." Http://snellman.kootutteokset.fi/, Accessed 22 September 2022, *Helsinki: Edita.*

Fernandez Riva, Gustavo. 2019. "Network Analysis of Medieval Manuscript Transmission: Some Basic Principles and Methods." *Journal of Historical Network Research* 3, no. 1 (November): 30–49. https://doi.org/10.25517/jhnr.v3i1.61.

Fiscarelli, Antonio Maria. 2022. "Social network analysis for digital humanities." In *Digital History and Hermeneutics. Between Theory and Practice,* edited by Andreas Fickers and Juliane Tatarinov, 23–42. Berlin, Boston: De Gruyter Oldenbourg. ISBN: 9783110723991. https://doi.org/doi:10.1515/9783110723991-002.

Freire, Nuno, René Voorburg, Roland Cornelissen, Sjors De Valk, Enno Meijers, and Antoine Isaac. 2019. "Aggregation of Linked Data in the Cultural Heritage Domain: A Case Study in the Europeana Network." *OpenReview.net,* https://doi.org/10.3390/info10080252.

Gerlach, Lisa, and David B. Blumenthal. 2023. "On the role of network topology in German-Jewish recommendation letter networks in the early twentieth century." *Applied Network Science* 8 (24). https://doi.org/https://doi.org/10.1007/s41109-023-00550-x.

Hagberg, A.A., D.A. Schult, and P.J. Swart. 2008. "Exploring network structure, dynamics, and function using NetworkX." In *Proceedings of the 7th Python in Science Conference,* edited by G. Varoquaux, T. Vaught, and J. Millman, 11–15. Pasadena, CA USA.

Heuvel, Charles van den. 2015. "Mapping Knowledge Exchange in Early Modern Europe: Intellectual and Technological Geographies and Network Representations." *researchgate.net* 9, no. 1 (March): 95–114. https://doi.org/10.3366/ijhac.2015.0140.

Hyvönen, Eero. 2023. "Digital Humanities on the Semantic Web: Sampo Model and Portal Series." *Semantic Web* 14 (4): 729–744. https://doi.org/10.3233/SW-223034.

Hyvönen, Eero. 2024. "How to Create a National Cross-domain Ontology and Linked Data Infrastructure and Use It on the Semantic Web." Accepted, *Semantic Web,* https://doi.org/10.3233/SW-243468.

Hyvönen, Eero, Petri Leskinen, Minna Tamper, Heikki Rantala, Esko Ikkala, Jouni Tuominen, and Kirsi Keravuori. 2019. "BiographySampo – Publishing and Enriching Biographies on the Semantic Web for Digital Humanities Research." In *The Semantic Web. ESWC 2019,* 574–589. Springer–Verlag. https://doi.org/10.1007/978-3-030-21348-0_37.

Hyvönen, Eero, Petri Leskinen, and Jouni Tuominen. 2023. "LetterSampo – Historical Letters on the Semantic Web: A Framework and Its Application to Publishing and Using Epistolary Data." *Journal on Computing and Cultural Heritage* 16 (1). https://doi.org/10.1145/3569372.

Ji, Shaoxiong, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S. Yu. 2022. "A Survey on Knowledge Graphs: Representation, Acquisition, and Applications." *IEEE Transactions on Neural Networks and Learning Systems* 33 (2): 494–514. https://doi.org/10.1109/TNNLS.2021.3070843.

Koltay, T. 2015. "Data literacy for researchers and data librarians." *Journal of Librarianship and Information Science* 49 (1): 3–14. https://doi.org/10.1177/0961000615616450. https://doi.org/10.1177/0961000615616450.

Leskinen, Petri, and Eero Hyvönen. 2021. "Using the AcademySampo Portal and Data Service for Biographical and Prosopographical Research in Digital Humanities." In *ISWC-Posters-Demos-Industry 2021 International Semantic Web Conference (ISWC) 2021: Posters, Demos, and Industry Tracks,* vol. 2980. CEUR Workshop Proceedings, October. http://ceur-ws.org/Vol-2980/paper330.pdf.

Leskinen, Petri, Eero Hyvönen, and Jouni Tuominen. 2021. "Sparql2GraphServer: a Server-side Tool for Extracting Networks from Linked Data for Data Analysis." In *ISWC-Posters-Demos-Industry 2021 International Semantic Web Conference (ISWC) 2021: Posters, Demos, and Industry Tracks,* vol. 2980. CEUR Workshop Proceedings, October. http://ceur-ws.org/Vol-2980/paper343.pdf.

Leskinen, Petri, Heikki Rantala, and Eero Hyvönen. 2022. "Analyzing the Lives of Finnish Academic People 1640–1899 in Nordic and Baltic Countries: AcademySampo Data Service and Portal." In *DHNB 2022 The 6th Digital Humanities in Nordic and Baltic Countries Conference,* vol. 3232. CEUR Workshop Proceedings, March. http://ceur-ws.org/Vol-3232/paper07.pdf.

Leskinen, Petri, Javier Ureña-Carrion, Jouni Tuominen, Mikko Kivelä, and Eero Hyvönen. 2023. "Knowledge Graphs and Data Services for Studying Historical Epistolary Data in Network Science on the Semantic Web." Under evaluation, *Semantic Web,* https://www.semantic-web-journal.net/content/knowledge-graphs-and-data-services-studying-historical-epistolary-data-network-science-1.

Li, Hui. 2018. "Social Network Extraction and Exploration of Historic Correspondences." PhD diss., Heidelberg University.

Menczer, Filippo, Santo Fortunato, and Clayton Davis. 2020. *A First Course in Network Science.* Cambridge University Press. ISBN: 9781108471138. https://doi.org/10.1017/9781108653947.

Miert, D van. 2008. "What was the Republic of Letters?" *Historisch Tijdschrift,* https://ugp.rug.nl/groniek/article/view/27601.

Miert, Dirk van. 2016. "What was the Republic of Letters? A brief introduction to a long history (1417–2008)." *Groniek* 204/205 (November): 269–287.

Moravec, Michelle. 2017. "Network Analysis and Feminist Artists." *Artl@s Bulletin* 6 (3). ISSN: 0191-6599.

Morrissey, Robert Michael. 2015. "Archives of Connection." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 48 (2): 67–79. https://doi.org/10.1080/01615440.2014.962208.

Onnela, J-P, Jari Saramäki, Jorkki Hyvönen, György Szabó, David Lazer, Kimmo Kaski, János Kertész, and A-L Barabási. 2007. "Structure and tie strengths in mobile communication networks." *Proceedings of the national academy of sciences* 104 (18): 7332–7336.

Otte, Evelien, and Ronald Rousseau. 2002. "Social network analysis: a powerful strategy, also for the information sciences." *Journal of Information Science* 28 (6): 441–453. https://doi.org/10.1177/016555150202800601.

Perry, Brea L., Bernice A. Pescosolido, and Stephen P. Borgatti. 2018. *Egocentric Network Analysis: Foundations, Methods, and Models.* Structural Analysis in the Social Sciences. Cambridge University Press.

Petz, Cindarella, Raji Ghawi, and Jürgen Pfeffer. 2022. "Tracking the Evolution of Communities in a Social Network of Intellectual Influences." *Journal of Historical Network Research* 7, no. 1 (November): 114–154. https://doi.org/10.25517/jhnr.v7i1.146. http://jhnr.uni.lu/index.php/jhnr/article/view/146.

Ravenek, W, C van den Heuvel, and G Gerritsen. 2017. "The ePistolarium: Origins and Techniques." *JSTOR,* https://www.jstor.org/stable/j.ctv3t5qjk.33.

Riehle, Alexander, and Johannes Preiser-Kapeller. 2020. *Letters and Network Analysis,* 7:431–465. Leiden; Boston : BRILL, ISBN: 90-04-41369-3.

Roller, Ramona. 2022. "Social Differentiation during the European Reformation. Statistical analyses of letter correspondence networks" [in en]. Doctoral Thesis, ETH Zurich. https://doi.org/10.3929/ethz-b-000587458.

Tamper, Minna, Petri Leskinen, Eero Hyvönen, Risto Valjus, and Kirsi Keravuori. 2023. "Analyzing Biography Collection Historiographically as Linked Data: Case National Biography of Finland." *Semantic Web* 14 (2): 385–419. https://doi.org/10.3233/SW-222887.

Tuominen, Jouni, Mikko Koho, Ilona Pikkanen, Senka Drobac, Johanna Enqvist, Eero Hyvönen, Matti La Mela, Petri Leskinen, Hanna-Leena Paloposki, and Heikki Rantala. 2022. "Constellations of Correspondence: a Linked Data Service and Portal for Studying Large and Small Networks of Epistolary Exchange in the Grand Duchy of Finland." In *6th Digital Humanities in Nordic and Baltic Countries Conference,* vol. 3232. CEUR Workshop Proceedings. https://ceur-ws.org/Vol-3232/paper41.pdf.

Ureña-Carrion, Javier, Petri Leskinen, Jouni Tuominen, Charles van den Heuvel, Eero Hyvönen, and Mikko Kivelä. 2022. "Communication Now and Then: Analyzing the Republic of Letters as a Communication Network." *Applied Network Science* 7 (May). https://doi.org/10.1007/s41109-022-00463-1.

Wetherell, Charles. 1998. "Historical Social Network Analysis." *International Review of Social History* 43:125–144. ISSN: 00208590, 1469512X, accessed August 19, 2024. http://www.jstor.org/stable/26405516.