Biographical and Prosopographical Analyses of Finnish Academic People 1640–1899 Based on Linked Open Data

Petri Leskinen[1,2], Eero Hyvönen[1,2]

[1]Semantic Computing Research Group (SeCo), Aalto University, Finland
[2]Helsinki Centre for Digital Humanities (HELDIG), University of Helsinki, Finland

Abstract

This paper presents work on prosopographical data analyses using the AcademySampo linked data service and portal. The original primary data, based on ten man-years of digitization work, covers a significant part of the Finnish university history based on the student registries in 1640–1852 and 1853–1899. They contain biographical descriptions of 28 000 students of the University of Helsinki, originally the Royal Academy of Turku. AcademySampo also sheds light to the academic history of Sweden and Baltic countries through their shared history with Finland in the larger Swedish empire. The Finnish student registries have been widely used by genealogists and historians by close reading. The main focus of this article is on the networks connecting the students and on knowledge discovered using this methodology. Networks connecting the students as well as their relatives mentioned in the data can be constructed based on various criteria, e.g., by genealogical relations, or by similarities on career by common vocations and employees. The student records already have a linkage to related Wikidata resources, which have been earlier used for enriching, e.g., the information about the relatives mentioned in the register descriptions. In this paper the biographical data is further extended by using Wikidata and by extracting further information and connections from the textual descriptions in Finnish, Swedish, or English Wikipedia. Although the descriptions in AcademySampo provide detailed data about the academic careers and family relations, the related Wikipedia entries can provide more details about their lifetime events with, e.g., their known locations of work or residence, topics of interest, lifetime events, or acquaintances. Topic of specific interest in this paper include: 1) Inheritance analysis of vocations and social classes in families. This analysis uses correlation matrices based on vocations of the students and their parents. 2) Quantitative analyses and visualizations of the family lines of students, based on automatically created family trees of the students and their parents. Family lines as long as eight generations can be found.

Keywords

Linked Data, Data Analysis, Digital Humanities, Network Analysis, Cultural Heritage

Biographical Data in a Digital World 2022 (BD 2022) Workshop, DH 2022, 25.-29. July, 2022
petri.leskinen@aalto.fi (P. Leskinen)
0000-0003-2327-6942 (P. Leskinen); 0000-0003-1695-5840 (E. Hyvönen)

1. Introduction

AcademySampo[1] [1, 2] consists of two parts: 1) a Linked Open Data (LOD) service[2] published on the Linked Data Finland platform [3] and 1) a semantic portal[3] based on it. The AcademySampo Portal provides intelligent capabilities for searching and browsing with seamlessly integrated data analytical tools and visualizations for biographical and prosopographical [4] research using statistics, networks, timelines, and maps. The open Application Programming Interfaces (API) of the LOD service and its SPARQL endpoint, in turn, provide an easy-to-use opportunity to implement data analyses for DH researchers with some experience in the SPARQL query language[4] and programming. For example, the Yasgui editor [5], Python scripts, Jupyter[5], and Google Colab notebooks[6] can be used.

AcademySampo is part of the Sampo portal series[7] [6] and uses the Linked Open Data Infrastructure for Digital Humanities in Finland (LODI4DH)[8] [7], a part of the Finnish FIN-CLARIAH infrastructure initiative[9]. This paper describes and compares a set of four networks created from the AcademySampo actor data, and introduces various results from analyzing the relationships found in genealogical or academic connections or in lifetime events.

2. Primary Data and Knowledge Graph

AcademySampo's data form an extensive knowledge graph that has been produced algorithmically from the digitized student registers of the Royal Academy of Turku and the University of Helsinki in 1640–1852 and 1853–1899[10] by extracting information from the texts and database structures. The data has been enriched by linking it both internally by artificial intelligence-based reasoning, and externally to other open datasets [8]. The student registers describe all people who have received academic education in Finland in 1640–1899, as there were no other universities in Finland at that time. The descriptions of students tell not only about their studies, but also about their career after studies and relatives, as well as references to the literature. The original register of the Royal Academy of Turku was destroyed in the Great Fire of Turku in 1827, but it was reconstructed in the late 19th century by Vilhelm Lagus. The register was supplemented in the 20th century from various sources, and in the end the information was edited by Yrjö Kotivuori and Veli-Matti Autio in an effort of ca. ten man years.

Since the registers 1640–1852 and 1853–1899 were provided by different authors their tabular CSV data differ to some extent. The source information found in the table of records 1640–1852 includes, in addition to some technical information in the database: 1) the person's registration number, 2) HTML text showing the person's name, places and times of birth and death, parents,

[1] Project homepage: https://seco.cs.aalto.fi/projects/yo-matrikkelit/
[2] The LOD service is available at https://ldf.fi/dataset/yoma
[3] Portal was opened February 2, 2021 at https://akatemiasampo.fi/en/
[4] https://www.w3.org/TR/sparql11-query/
[5] Jupyter Project and Tool: https://jupyter.org
[6] Google Colab: https://colab.research.google.com/notebooks/intro.ipynb#recent=true
[7] See: https://seco.cs.aalto.fi/applications/sampo/
[8] LODI4DH initiative: https://seco.cs.aalto.fi/projects/lodi4dh/
[9] https://seco.cs.aalto.fi/projects/fin-clariah/
[10] Student Registers, University of Helsinki: https://www.helsinki.fi/fi/yliopisto/ylioppilasmatrikkelit-1640-1907

career events, relatives, students, references and 3) the date the record was created. If the person mentioned in the register 1640–1852 is found in either of the registers, a HTML link is manually created connecting this mention to the person's page using the registration number. However, in the register 1853–1899 there are no such links, and the references have been interpreted computationally. In addition, supplementary textual information about a person may be available in other registers. For example, Finnish national poet *Johan Ludvig Runeberg* has further information in the registers of Lagus and Carpelan.

The primary data used in creating AcademySampo was therefore mainly text in HTML format without structured metadata, such as places or times of birth, vocation, etc. A major technical challenge in creating the linked data was to unambiguously identify the entities and events mentioned in the text, such as marriages, rewards and promotions, and key concepts, such as vocations. A specific challenge in extracting information was to distinguish between people with the same name, to reason their gender by name, and to infer various relationships, such as little cousin, through other relationships.

The data of the AcademySampo was converted into Linked Data[11] [9] by structuring the text descriptions of the Student register 1640–1852 for about 9500 people and the register 1853–1899 for about 18 450 people. This was done by identifying, through regular expressions, basic biographical information about students, their 47 000 relatives, 120 000 interpersonal relationships, 3000 historical places, 10 000 vocations, and 4000 academic teacher-student relationships. The "semantic glue" of the knowledge graph are the events related to the professional and family life of people identified in the texts, which link the people and organizations involved in different roles with places and times according to the CIDOC CRM[12] ontology and ISO standard. The data has been enriched by linkage to external databases, such as the Finnish National Biography and other biographies of the Finnish Literature Society avaible as LOD in the BiographySampo system [10] and Wikidata[13], and by inferring relationships between people [8]. The public open data service (CC BY 4.0) is available at the Linked Data Finland for accessing and utilizing the data in research and application development, such as the AcademySampo portal.

3. Networks based on different criteria

Social networks can be constructed from a biographical LOD publication with various, different criteria. In this chapter four such networks are introduced and analyzed. The examples are 1) Genealogical family relation network, 2) Teacher-student relation network, 3) A network based on similarity of life events, and 4) a reference network imported from Wikipedia. The analyses presented in this chapter were generated in Google Colab notebooks using the data available at the AcademySampo SPARQL endpoint. The SPARQL results were converted into social networks using the Python module NetworkX[14] [11]. The figures were generated using Python modules Matplotlib[15] and Seaborn[16] or using the network visualization software Gephi [12]

[11] W3C Linked Data: https://www.w3.org/standards/semanticweb/data
[12] CIDOC CRM: http://cidoc-crm.org
[13] Wikidata: https://wikidata.org
[14] https://networkx.org/
[15] https://matplotlib.org/
[16] https://seaborn.pydata.org/

after exporting the network in graphml [13] format.

3.1. Family Relations

The AcademySampo data is rich with detailed family relationships which were manually added by the authors of the 1640–1852 register data. This linkage was later used as training data for linking the relatives in the 1853–1899 dataset [8]. The students are interconnected with 66 types of relations from close relations like parent or child to more distant ones like stepfather-in-law [14]. Approximately 18 900 students have at least one relative among the students. The diagram on the left in Figure 1 depicts the percentage of students who have other relatives among the students during the centuries, and the time series on the right shows the percentage of students whose parent also studied at the University. Generally, during the later half of 19th century the amount of students without an academic family background starts increasing rapidly.

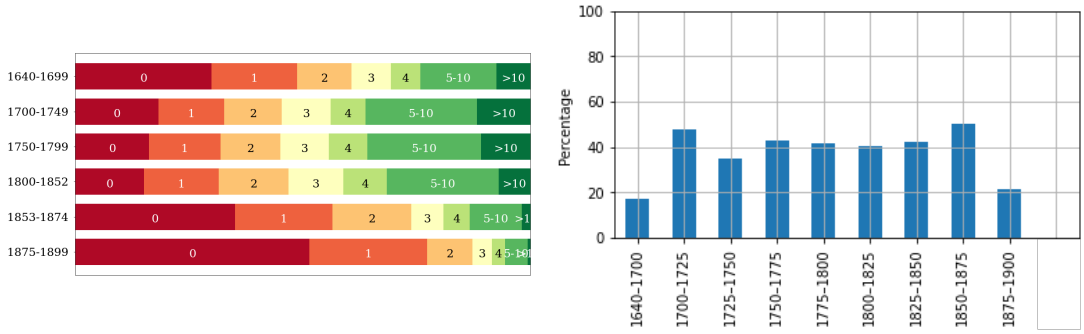

Figure 1: The diagram on the left depicts the number of relatives in the university for each student; the time series on the right depicts the proportion of students whose father also studied at the University.

To further analyze the length of family lines, a genealogical network was created, this time using only the parent-child relations. In this network each kin is represented as a connected component, and the number of generations equals to the length of the longest path in it. In Table 1 the first column is the number of generations in the family line and the second one is the number of kins with that number of generations. For example, there are five kins with a length of eight generations. The largest subgraph has the size 66 nodes. The value 8171 at the bottom row is the number of students without relatives among the students. Notice that a subgraph with a maximum path length 8 contains also subpaths with all shorter lengths (7, 6, 5, ...).

Table 2 lists examples of the student names along two family lines, so that the oldest ancestors are listed first. One could, for example, pay attention to the changes in the spelling of the family names during the centuries. A closer look at the biography of *Nils Abraham Ursin* reveals that in 1845 he was ennobled with the family name *af Ursin*. The history of each family line could be further analyzed by, e.g., looking at the places of birth and death along the family history. Both of the families have their roots in small villages (*Kalvola, Eura, Rantasalmi*) but later they have moved to larger towns (*Turku, Helsinki, Kuopio*) in Finland.

Figure 2 visualizing the changes in the most common places of birth during the years 1650–1900. The places considered are towns and municipalities of Finland and the neighbour countries

Table 1
Lengths and amounts of family lines

generations	#families
8	5
7	21
6	47
5	81
4	205
3	547
2	1816
1	8171

Table 2
Examples of person names in two family lines in format *family name, given names*

Family Line 1	Family Line 2
Homman, Tomas	Ursinus, Jakob
Homeen, Johan	Ursinus, Jakob
Homeen, Johan	Ursinus, Nils
Homén, Johan Jakob	Ursin, Jakob Johan
Homén, Gustaf Vilhelm	af Ursin, Nils Abraham
Homén, Lars	af Ursin, Julius
Homén, Lars Olaf	af Ursin, Nils Robert

Sweden and Russia. Number of student born in other countries was not large enough to have an effect on the results. Turku was the old capital and the largest town in Finland but started to lose its significance during the first half of 19th century when Helsinki first became the new capital and later when the university was moved to Helsinki. The figure also shows that the number of students coming from Sweden was high in the 17th and 18th centuries but decreased significantly in the 19th century when Finland became a part of Russian Empire. Consequently, at that time there is increase in the number of students born in Russia.

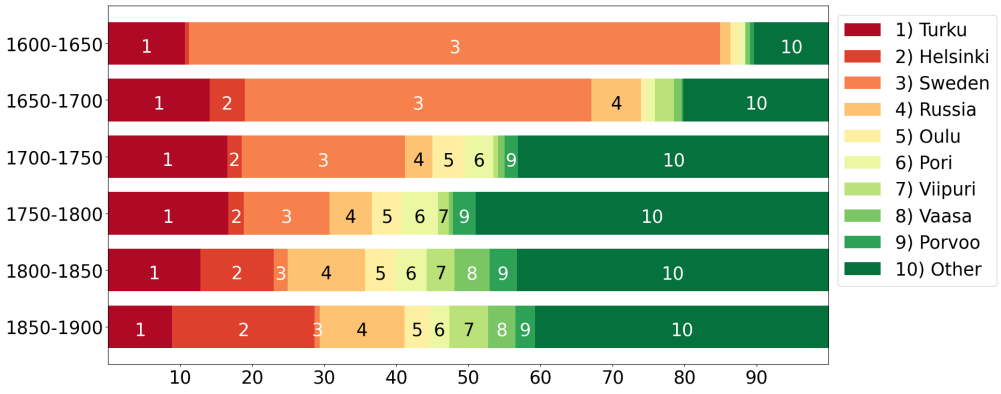

Figure 2: The most common places of birth during the years 1650–1900

Figure 3 depicts the distribution of most common vocational groups during the years 1650–1950. In the early years of the time frame, *Religious work* was the dominant category with a portion of over 50 %. However, that significance decreased to a mere 17 % at the 20th century while vocations related to, e.g., *Public administration* gained more importance. Furthermore, there is a notable growth in the proportion for new occupations in the category *Other*.

Figure 4 shows a correlation matrix between the vocational groups of parent-children pairs. The full labels of the vocational groups are shown on the rows; the columns have the same

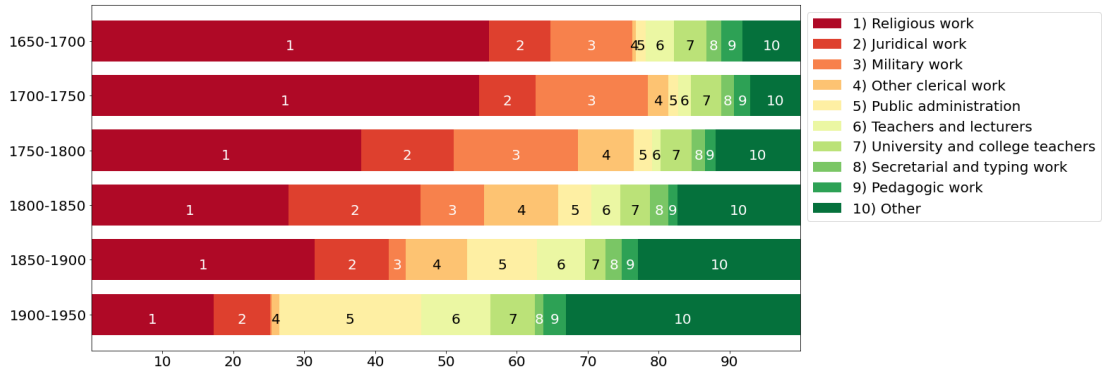

Figure 3: The most common vocational categories during the years 1650–1950

order but only the indices are shown underneath the figures. The values in the cells are the probabilities that a child working in a field represented by the row has had a parent working in the field for the corresponding column. On each row the cell with the largest value has the darkest background color, and all the values on a row sum up 100 %. For example, the uppermost value on the left indicates that 49.1 % of children in the religious work category have had parents in religious work category as well, likewise 53.1 % of the students with parents in the field of agriculture, forestry and fishing (row 6) has chosen a religious work. The values on the matrix diagonal are the probabilities of choosing the same vocational group as one's parent has. However, when looking at these statistics one has to remember the academic context. For instance, on the 6th column all the values are relatively small meaning that independent of the parents' vocation only a few children has chosen a work in agriculture, forestry or fishing, although in 17th–19th centuries Finland was an agricultural country.

Figure 5 depicts the correlations between siblings born in the 19th century. Altogether the dataset contains approximately 8800 such sibling pairs. The numbers in the cells are the number of related pairs. By looking at the matrix one can notice that the four most dominant categories, *Public administration*, *Religious work*, *Juridical work*, and *Teachers and lecturers* have high values of correlation. One also has to remember the biases of our data, like for instance the intercorrelation in the field of *Military work* remains low due to the fact that people who chose a military career may not have studied at the university.

3.2. Teacher-Student Relations

The dataset contains a network of teachers and students spanning from 1640's to the year 1853. In a similar way to the family relations, this linkage was manually added by the original dataset authors. There are altogether 4893 links connecting 3159 people. In our work based on the LOD service, network statistics were used to, e.g., locate the most significant individuals. Figure 6 shows how the network spans continuously over the entire time window. On this illustration three most central actors are emphasized, *Henrik Gabriel Porthan*, *Jakob Gadolin*, and *Algot Scarin* who all are famous scholars and professors.

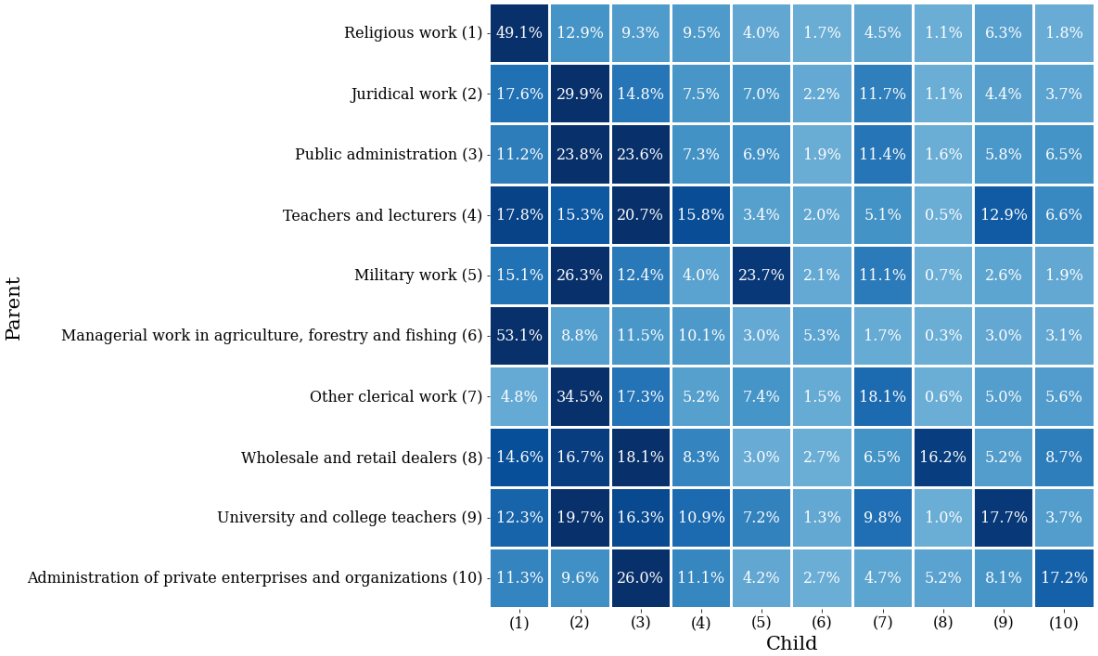

Figure 4: Correlations of the vocational categories between parents and children

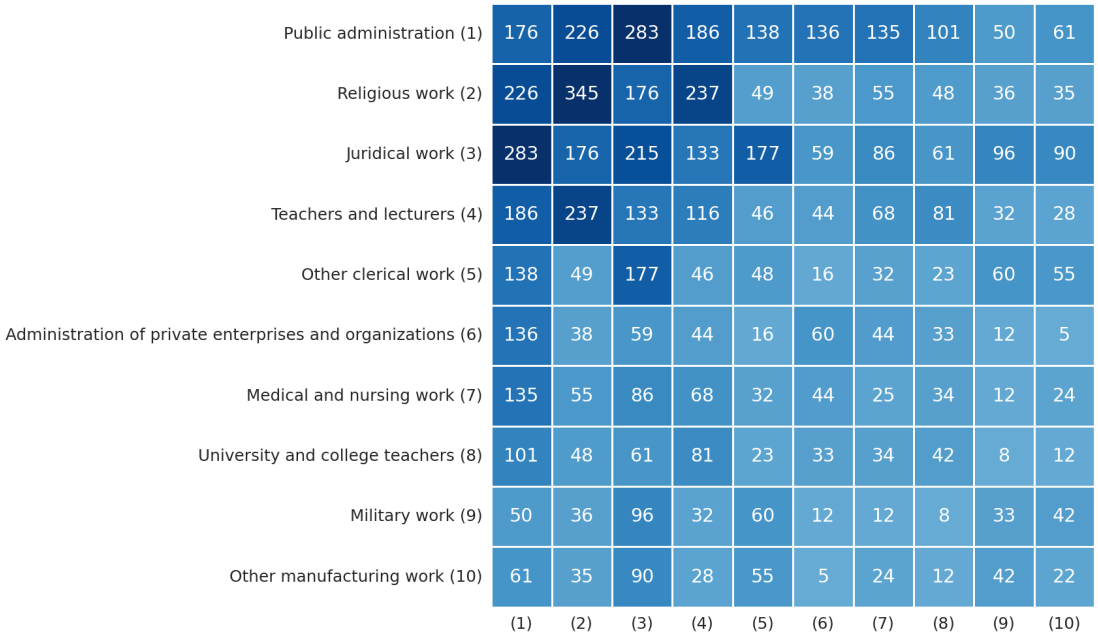

Figure 5: Correlations of the vocational categories between siblings born in the 19th century

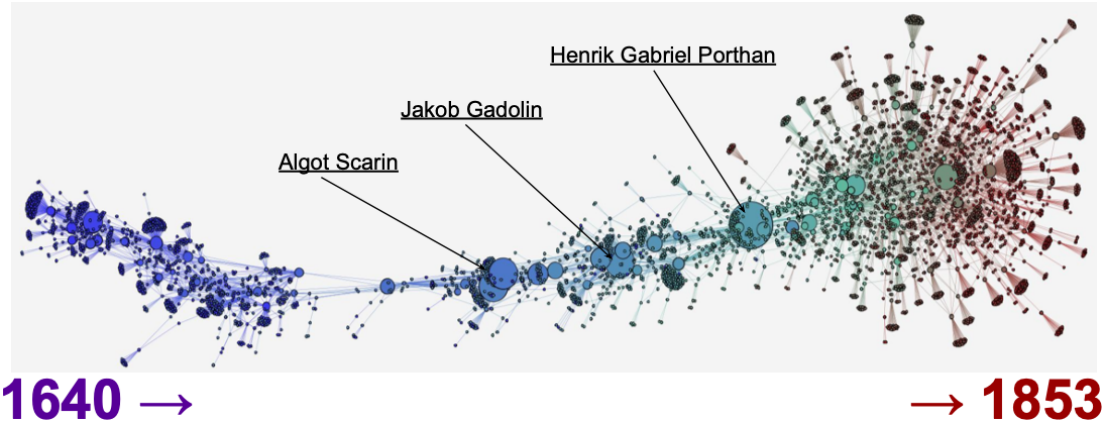

Figure 6: Network based on teacher-student relations

3.3. Similarity of Lifetime Events

Similarity between the students was calculated from the RDF data using the lifetime events. Features like having the same vocation, participating in the same event, being in the same places etc. were considered as links connecting the students. The similarity measure was achieved computationally by 1) a breadth-first search querying all the nodes related to each student including, e.g., the hierarchy of places, time spans, and vocations, 2) by filtering out the nodes that are related to very few or to too many students, 3) by constructing a matrix where each related entity is a feature (column) for a student (row), 4) by applying the TF-IDF measure to reduce the weight of most common terms and to emphasize the rarer ones, and by 5) calculating the similarity using cosine similarity. In addition to this method, also *RDF2VEC embeddings* [15] were tested on the data. However, the similarities and recommendations achieved with embeddings did not seem feasible. Furthermore, the approach above allowed to adjust weights by the class of a feature, e.g., to enhance the importance of, e.g., related organizations or inversely reduce the importance of a common time span.

In the data publication the recommended similarities are modeled as RDF resources connecting the two students, indicating the similarity value, and containing links to the database entries having the highest effect on the found similarity. For example, the people similar to chemist and physicist Johan Gadolin are found based on terms like *mineralogist*, *Uppsala University*, and *Royal Swedish Academy of Sciences*. Another example of similarities is a cluster of engineers[17] who worked in Baku, Azerbaijan, for the oil company Branobel[18] that was run by the Nobel brothers. In the AcademySampo Portal these connections are shown as a network visualization[19].

[17] https://akatemiasampo.fi/en/people/page/p17642/table
[18] https://en.wikipedia.org/wiki/Branobel
[19] https://akatemiasampo.fi/en/people/page/p17642/connections

Table 3
Typical classes of links found in Wikipedia pages

Class of Wikipedia Entity	Count
human	1443
municipality of Finland	236
former municipality of Finland	174
city	102
newspaper	97
town	70
academic discipline	65
position	64
profession	63
organization	59

3.4. Enriching Data from an External Databases

Register descriptions of people are often short, and an external database can provide more detailed information about their lifetime. The AcademySampo Data Service contains also a linkage to external data publications, such as the Finnish BiographySampo[20] [10], Members of Finnish Parliament, Ministers of Finland, as well as the international Wikidata. Using the linkage to Wikidata allows to also access the related Wikipedia pages written in various languages. Out of the total of 28 000 students approximately 2700 have an entry in the Finnish Wikipedia. The description text from the Finnish Wikipedia page was queried for each person. The graph is constructed based on the links on the pages so that two entities having the same link get interconnected. The Python module MediaWiki[21] was used for scraping the pages.

In this graph also the properties of the links can be analyzed, e.g., what are the connections based on related people, places, vocations, or organizations. The most frequent classes of the links are shown in Table 3 indicating that in most cases two people are connected by a mutual reference to a third person. Many of the referenced people are Finnish contemporaries but this information can also reveal clusters of students who became authors and were influenced, e.g., by *Goethe*, *Ovid* or *Aesop*. Besides references to people, in other cases the links are generated by references to places or more rarely by an organization, academic discipline or degree, ideology, historical event, a work of art, or a style of art or literature. Among the organizations is the Finnish Medical Society *Duodecim*[22] which was founded in 1881 by 12 Finnish physicians who are all in the AcademySampo database.

In addition to the links in Wikipedia, the information about the categories can be utilized in analyses. Figure 7 depicts a graph where the weight of an edge connecting two categories equals the number of people belonging to both. Observing the graph shows that the most common categories are *Professors*, *Nobility*, and *Member of (Finnish) Parliament*.

[20] https://biografiasampo.fi
[21] https://pymediawiki.readthedocs.io/en/latest/code.html#api
[22] https://fi.wikipedia.org/wiki/Suomalainen_L%C3%A4%C3%A4k%C3%A4riseura_Duodecim

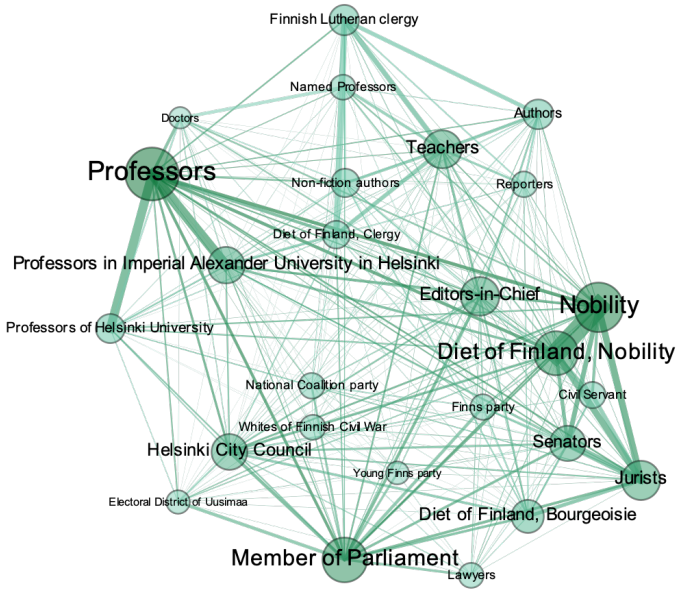

Figure 7: Categories based on the Wikipedia linkage

Table 4
Top actors in example networks by pagerank centrality

	Student–teacher	Wikipedia
1	Porthan, Henrik G.	Mannerheim, Carl G. E.
2	Gyldenstolpe, Mikael	Paasikivi, Juho K.
3	Scarin, Algot	Mechelin, Leopold H. S.
4	Hassel, Henrik	Leino, Eino
5	Gadolin, Jakob	Sibelius, Jean

3.5. Analyzed Networks

Table 4 depicts the top five actors by their Pagerank centrality in the teacher-student and Wikipedia networks analyzed in the previous section. The central actors in the teacher-student network are the same as in Figure 6 while the central actors in the Wikipedia network are well-known Finnish people of politics and culture. The networks constructed by genealogical relations and similarity values are different in their nature, so applying a social network statistic to them would not reveal useful results.

Table 5 contains general metrics of the four networks, (1) The teacher-student relations, (2) the genealogical network, (3) the network based on actor similarities, and (4) network

Table 5
Comparison between the four networks (Similarity, teacher-student, Families, and Wikipedia) in the AcademySampo, *BiographySampo*, and *Email* datasets using network measures

Measure	Similarity	Student–teacher	Families	Wikipedia	*Biographies*	*Email*
edges	20000	4893	9380	24988	2741	2396
nodes	20418	3159	12183	2628	-	-
density	0.98	1.55	0.77	9.50	5.48	4.79
average degree	1.01	3.36	1.54	1.52	-	-
HD	17.62	231	10	35.73	323	499
max clique	4	4	2	17	-	-
diameter	103	15	12	8	5	7
GCC	0.04	0.05	0	0.34	0.35	0.54
components	2269	6	2806	3	-	-
giant component	9338	3146	66	2624	-	-
APL	31.70	6.24	6.08	3.09	2.76	1.98
alpha (α)	1.63	1.38	2.01	1.33	1.43	1.87

constructed based on Wikipedia pages. For comparison also the available measures from the BiographySampo reference network (*Biographies*) described in [16] as well as the EU Email Communication Network (*Email*) analyzed by Hashmi et al. [17] are included in the table. This table contains first the numbers of nodes and edges in the network. The Average degree indicates the average amount of links for a single node and highest degree (HD) is the highest node degree in the network. Max clique size is the largest size of a clique. For example, the value 17 indicates that there exists a subgroup of 8 people who all are linked to one another. The table shows the number of separated components in the network, and the size of the largest connected component. The genealogical network is scattered into numerous separated components, while the three reference networks are all more connected having giant components connecting most of the data points. The Diameter is the number of edges along the longest path between any two nodes in the network. The Alpha (α) is the constant obtained when a power-law distribution is fitted on the degree distribution of the network [18]. The Global Clustering Coefficient (GCC) is the measure of connected triples; the Average Path Length (APL) is the average number of edges traversed along the shortest paths for all possible pairs of nodes.

The measures provided here are the same as introduced in Hashmi et al., and which were later analyzed in the context of BiographySampo data. The four networks are different in their nature: in the network of similarities each node is forced to find a similar pair; the genealogical network is a directed acyclic graph consisting of relatively small connected components; also in the network built by teacher-student relationships the triadic closure (GCC) remains low. However, the measures of the Wikipedia network, specially density, GCC, or diameter, show a small-world behavior of a social network, as measured for *Biographies* and *Email*.

4. Discussion

Work on AcademySampo is continuation to our earlier biographical LOD systems on Norssit Alumni register [19], the U.S. Congress Prosopographer [20], and BiographySampo [10]. Our earlier articles provide examples of analyses for BiographySampo data [16] and for the Members of the Finnish Parliament in the ParliamentSampo system [21, 22]. Extracting Linked Data from texts has been studied in several works [23]. In [24] language technology was used for extracting entities from biographies and in [25] from news. With epistolary data the social networks can be constructing from the letter exchange information [26, 27, 28, 29].

Representing and analyzing biographical data is a new research and application field. In 2015, the first Biographical Data in Digital World workshop BD2015 was held presenting several works on studying and analyzing biographies as data [30], and the proceedings of BD2017 contain more similar works [31]. In [32], analytic visualizations were created based on U.S. Legislator registry data. The idea of biographical network analysis is related to the Six Degrees of Francis Bacon system[23] [33, 34] that utilizes data of the Oxford Dictionary of National Biography. In earlier research, sociocentric and egocentric networks connecting the actors could be constructed from texts based on, e.g., mentioned names, hypertext links, genealogical relations, or similarities in characteristics, such as lifetime events [2, 35].

This paper presented the idea of creating various networks of academic students based on linkage in Linked Data. It was also shown that networks based on different approaches can reveal different phenomena from the data.

Acknowledgements Yrjö Kotivuori and Veli-Matti Autio authored the original data publications used in our work. Our work is related to the EU project InTaVia: In/Tangible European Heritage[24]. CSC – IT Center for Science has provided computational resources for the work.

References

[1] P. Leskinen, E. Hyvönen, Linked open data service about historical Finnish academic people in 1640–1899, in: DHN 2020 Digital Humanities in the Nordic Countries. Proceedings of the Digital Humanities in the Nordic Countries 5th Conference, CEUR Workshop Proceedings, Vol. 2612, 2020, pp. 284–292. URL: http://ceur-ws.org/Vol-2612/short14.pdf.

[2] P. Leskinen, H. Rantala, E. Hyvönen, 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 , CEUR Workshop Proceedings, long papers, Vol. 3232, 2022. URL: http://ceur-ws.org/Vol-3232/paper07.pdf.

[3] E. Hyvönen, J. Tuominen, M. Alonen, E. Mäkelä, Linked Data Finland: A 7-star Model and Platform for Publishing and Re-using Linked Datasets, in: ESWC 2014: The Semantic Web: ESWC 2014 Satellite Events, Springer–Verlag, 2014, pp. 226–230. doi:10.1007/978-3-319-11955-7_24.

[23] http://www.sixdegreesoffrancisbacon.com
[24] https://intavia.eu/

[4] K. Verboven, M. Carlier, J. Dumolyn, A short manual to the art of prosopography, in: Prosopography approaches and applications. A handbook, Unit for Prosopographical Research (Linacre College), 2007, pp. 35–70. doi:1854/8212.

[5] L. Rietveld, R. Hoekstra, The YASGUI family of SPARQL clients, Semantic Web – Interoperability, Usability, Applicability 8 (2017) 373–383. doi:10.3233/SW-150197.

[6] E. Hyvönen, Digital humanities on the semantic web: Sampo model and portal series, Semantic Web – Interoperability, Usability, Applicability 14 (2023) 729–744. doi:10.3233/SW-223034.

[7] E. Hyvönen, Linked open data infrastructure for digital humanities in Finland, in: DHN 2020 Digital Humanities in the Nordic Countries. Proceedings of the Digital Humanities in the Nordic Countries 5th Conference, CEUR Workshop Proceedings, vol. 2612, 2020, pp. 254–259. URL: http://ceur-ws.org/Vol-2612/short10.pdf.

[8] P. Leskinen, E. Hyvönen, Reconciling and using historical person registers as linked open data in the AcademySampo knowledge graph, in: The Semantic Web – ISWC 2021, Springer–Verlag, 2021, pp. 714–730. doi:10.1007/978-3-030-88361-4_42.

[9] T. Heath, C. Bizer, Linked Data: Evolving the Web into a Global Data Space (1st edition), Synthesis Lectures on the Semantic Web: Theory and Technology, Morgan & Claypool, 2011. URL: http://linkeddatabook.com/editions/1.0/.

[10] E. Hyvönen, P. Leskinen, M. Tamper, H. Rantala, E. Ikkala, J. Tuominen, K. Keravuori, BiographySampo – Publishing and Enriching Biographies on the Semantic Web for Digital Humanities Research, in: The Semantic Web. ESWC 2019, Springer–Verlag, 2019, pp. 574–589. doi:10.1007/978-3-030-21348-0_37.

[11] A. A. Hagberg, D. A. Schult, P. J. Swart, Exploring Network Structure, Dynamics, and Function using NetworkX, in: G. Varoquaux, T. Vaught, J. Millman (Eds.), Proceedings of the 7th Python in Science Conference, Pasadena, CA USA, 2008, pp. 11 – 15.

[12] M. Bastian, S. Heymann, M. Jacomy, Gephi: An Open Source Software for Exploring and Manipulating Networks, in: Third international AAAI conference on weblogs and social media, 2009. URL: https://www.academia.edu/download/3244556/gephi-bastian-feb09.pdf.

[13] U. Brandes, M. Eiglsperger, I. Herman, M. Himsolt, M. S. Marshall, GraphML progress report structural layer proposal, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 2265 LNCS (2002) 501–512. doi:10.1007/3-540-45848-4_59.

[14] P. Leskinen, E. Hyvönen, Extracting Genealogical Networks of Linked Data from Biographical Texts, in: The Semantic Web: ESWC 2019 Satellite Events, Springer, 2019, pp. 121–125. doi:10.1007/978-3-030-32327-1_24.

[15] P. Ristoski, J. Rosati, T. Di Noia, R. De Leone, H. Paulheim, RDF2Vec: RDF graph embeddings and their applications, Semantic Web 10 (2019) 721–752.

[16] M. Tamper, P. Leskinen, E. Hyvönen, R. Valjus, K. Keravuori, Analyzing biography collection historiographically as linked data: Case National Biography of Finland, Semantic Web – Interoperability, Usability, Applicability 14 (2023) 385–419. doi:10.3233/SW-222887.

[17] A. Hashmi, F. Zaidi, A. Sallaberry, T. Mehmood, Are all social networks structurally similar?, in: Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on, IEEE, 2012, pp. 310–314. doi:10.1109/asonam.2012.59.

[18] A. Clauset, C. R. Shalizi, M. E. Newman, Power-Law Distributions in Empirical Data, http://dx.doi.org/10.1137/070710111 51 (2009) 661–703. URL: https://epubs.siam.org/doi/abs/10.1137/070710111. doi:10.1137/070710111.

[19] E. Hyvönen, P. Leskinen, E. Heino, J. Tuominen, L. Sirola, Reassembling and Enriching the Life Stories in Printed Biographical Registers: Norssi High School Alumni on the Semantic Web, in: Language, Technology and Knowledge, Springer–Verlag, 2017, pp. 113–119. doi:10.1007/978-3-319-59888-8_9.

[20] G. Miyakita, P. Leskinen, E. Hyvönen, Using Linked Data for Prosopographical Research of Historical Persons: Case U.S. Congress Legislators, in: 7th International Conference, EuroMed 2018, Proc., Part II, Springer-Verlag, 2018, pp. 150–162. doi:10.1007/978-3-030-01765-1_18.

[21] P. Leskinen, E. Hyvönen, J. Tuominen, Members of Parliament in Finland Knowledge Graph and Its Linked Open Data Service, in: Further with Knowledge Graphs. Proceedings of the 17th International Conference on Semantic Systems, 6-9 September 2021, Amsterdam, The Netherlands, IOS Press, 2021, pp. 255–269. URL: https://ebooks.iospress.nl/volumearticle/57420. doi:10.3233/SSW210049.

[22] H. Poikkimäki, P. Leskinen, M. Tamper, E. Hyvönen, Analyses of Networks of Politicians Based on Linked Data: Case ParliamentSampo – Parliament of Finland on the Semantic Web, in: Semantic Web and Ontology Design for Cultural Heritage (SWODCH 2022), Turin, Italy, Proceedings, CEUR WS Proceedings, 2022. URL: https://seco.cs.aalto.fi/publications/2022/poikkimaki-et-al-2022.pdf, accepted.

[23] J. L. Martinez-Rodriguez, A. Hogan, I. Lopez-Arevalo, Information Extraction Meets the Semantic Web: A Survey, Semantic Web – Interoperability, Usability, Applicability 11 (2020) 255–335.

[24] A. Fokkens, S. ter Braake, N. Ockeloen, P. Vossen, S. Legêne, G. Schreiber, V. de Boer, BiographyNet: Extracting Relations Between People and Events, in: Europa baut auf Biographien, New Academic Press, Wien, 2017, pp. 193–224.

[25] M. Rospocher, M. van Erp, P. Vossen, A. Fokkens, I. Aldabe, G. Rigau, A. Soroa, T. Ploeger, T. Bogaard, Building event-centric knowledge graphs from news, Web Semantics: Science, Services and Agents on the World Wide Web 37 (2016) 132–151.

[26] W. Ravenek, C. v. d. Heuvel, G. Gerritsen, The ePistolarium: Origins and Techniques, JSTOR (2017). URL: https://www.jstor.org/stable/j.ctv3t5qjk.33.

[27] A. Rockenberger, E. Nessheim Wiger, M. Refslund Witting, H. Bøe, E. Irene Thor, O. Wolden, M. Paasche, O. Søndenå, NorKorr – Norwegian Correspondences and Linked Open Data, in: Digital Humanities in the Nordic Countries 2019, 2019. URL: https://munin.uit.no/handle/10037/15862, poster paper.

[28] S. Dumont, correspSearch -– Connecting Scholarly Editions of Letters, Journal of the Text Encoding Initiative (2016). URL: https://doi.org/10.4000/jtei.1742. doi:10.4000/jtei.1742.

[29] E. Hyvönen, P. Leskinen, J. Tuominen, LetterSampo – Historical Letters on the Semantic Web: A Framework and Its Application to Publishing and Using Epistolary Data, Journal on Computing and Cultural Heritage 14 (2023) 1–24. doi:10.1145/3569372.

[30] S. ter Braake, R. S. Anstke Fokkens, T. Declerck, E. Wandl-Vogt (Eds.), BD2015, Biographical Data in a Digital World 2015, CEUR Workshop Proceedings, Vol-1399, 2015. URL: http:

//ceur-ws.org/Vol-1399/.
[31] A. Fokkens, S. ter Braake, R. Sluijter, P. Arthur, E. Wandl-Vogt (Eds.), BD2017 Biographical Data in a Digital World 2015, CEUR Workshop Proceedings, Vol-2119, 2017. URL: http://ceur-ws.org/Vol-2119/.
[32] R. Larson, Bringing Lives to Light: Biography in Context, 2010. Final Project Report, University of Berkeley, http://metadata.berkeley.edu/Biography_Final_Report.pdf.
[33] C. Warren, D. Shore, J. Otis, L. Wang, M. Finegold, C. Shalizi, Six degrees of Francis Bacon: A statistical method for reconstructing large historical social networks, Digital Humanities Quarterly 10 (2016).
[34] A. Langmead, J. Otis, C. Warren, S. Weingart, L. Zilinski, Towards Interoperable Network Ontologies for the Digital Humanities, Int. J. of Humanities and Arts Computing 10 (2016) 22–35.
[35] D. K. Elson, K. McKeown, N. J. Dames, Extracting Social Networks from Literary Fiction, aclweb.org (2010). URL: https://www.aclweb.org/anthology/P10-1015.pdf.