

# *Utilizing Annotated Keywords for Prosopographical Data Analysis in Sampo Systems*

*FIN-CLARIAH Roadshow: From Data to Intelligence, 25.3. 2026*

*Petri Leskinen, D.Sc.*

*University of Helsinki, Aalto University, Geneva Graduate Institute*

# Contents

- **Topics**
  - Data and network analysis
  - based on keywords and named entities
- **Case examples**
  - BiographySampo
  - AcademySampo
  - ParliamentSampo





# Biography Sampo

# BiographySampo

1. **Project and Data publication**
2. **Keywords**
3. **Networks**
4. **Similarities**

# BiographySampo

## 1. Project

- a. published in 2018
- b. <http://biografiasampo.fi>

## 2. Data provided by Finnish Literature Society

- a. National biography of Finland
- b. Finnish Clergy 1554–1721
- c. Finnish Clergy 1800–1920
- d. Finnish Generals and Admirals in the Russian armed forces 1809--1917,
- e. Finnish Business Leaders



# BiographySampo

## 1. Keywords

- a. manual
- b. automated
  - i. *based on the free text parts*
  - ii. *analyzed using automatic annotation*
  - iii. *represented using keywords taken from the Finnish General Ontology YSO*

## 2. Networks

- a. manual annotation
- b. automatic annotation
- c. musician example

## 3. Similarities

- a. TF-IDF based recommender system



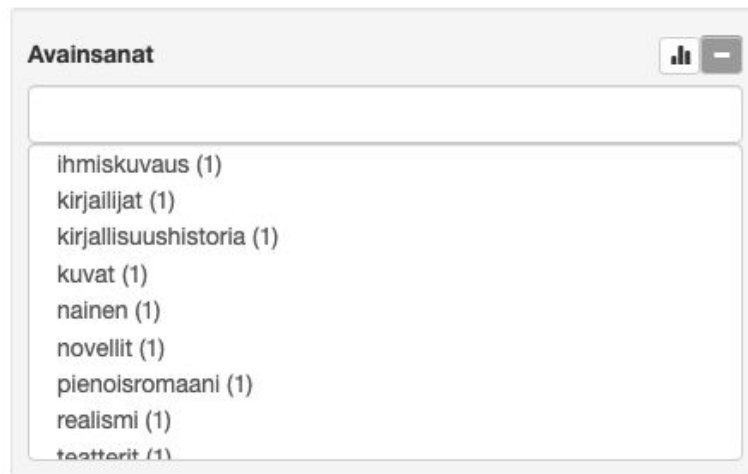
# BiographySampo

TABLE 12: Top ten words used in the biographies of female politicians

	NOUN			ADJ		
	Finnish	English	Count	Finnish	English	Count
1	nainen	woman	557	poliittinen	political	303
2	kuningatar	queen	459	vanha	old	169
3	puolue	political party	456	nuori	young	162
4	kuningas	king	422	seuraava	next	156
5	lapsi	child	378	suomalainen	Finnish	154
6	puoliso	spouse	317	yhteiskunnallinen	societal	122
7	eduskunta	parliament	314	merkittävä	significant	109
8	poika	son	283	sosiaalidemokraattinen	socialdemocratic	100
9	äiti	mother	283	tärkeä	important	97
10	puheenjohtaja	chairperson	278	kansainvälinen	international	94

TABLE 13: Top ten words used in the biographies of male politicians

	NOUN			ADJ		
	Finnish	English	Count	Finnish	English	Count
1	hallitus	government	4066	poliittinen	political	2493
2	puolue	political party	3766	suomalainen	Finnish	1453
3	tehtävä	task	2725	merkittävä	significant	1108
4	puheenjohtaja	chairperson	2649	tärkeä	important	1093
5	jäsen	member	2460	vanha	old	1078
6	kuningas	king	1845	keskeinen	central	995
7	toiminta	action	1840	nuori	young	985
8	eduskunta	parliament	1786	seuraava	next	983
9	sota	war	1742	sanottu	so called or said	693
10	presidentti	president	1718	yhteiskunnallinen	societal	646



# Recommender system

 **Tellervo Koivisto**



+  Punkalaidun 2.1.1929

kansanedustaja, tasavallan presidentin puoliso

Lähteet: Kansallisbiografia, Kansallisbiografian sisällysluettelo

[Eduskunta](#) [Fennica](#) [Geni.com: Kotisivu](#) [Geni.com: Sukupuu](#) [ISNI](#) [Kirjasampo](#) [VIAF](#) [Wikipedia](#) [Wikidata](#)

 **Kansallisbiografia** [Biografiakeskuksen artikkeli](#) [Geni.com, kotisivu](#) [Geni.com, sukupuu](#)

[Kirjasampo](#) [Wikipedia](#)

## KONTEKSTUAALINEN LUKIJA

Alla olevasta tekstistä on tunnistettu  **henkilöt**,  **paikat**, ja  organisaatiot automaattisin menetelmin.

### Lähisukulaiset

isä  Johannes Einari Kankaanranta  
äiti  Olga Maria Saarinen aviomies  **Mauno Koivisto**  
tytär  Assi Elina Koivisto-Allonen

### Muut sukulaiset

appi  Juho Koivisto

### Samankaltaisia biografioita

 [Sylvi Kekkonen](#)  
[Helvi Boothe](#)  
[Marianne Laxén](#)  
 [Jenni Haukio](#)  
[Risto Ryti](#)  
[Kerppo Kallas Kankaa...](#)  
 [Ellen Svinhufvud](#)  
[Simo Juntunen](#)

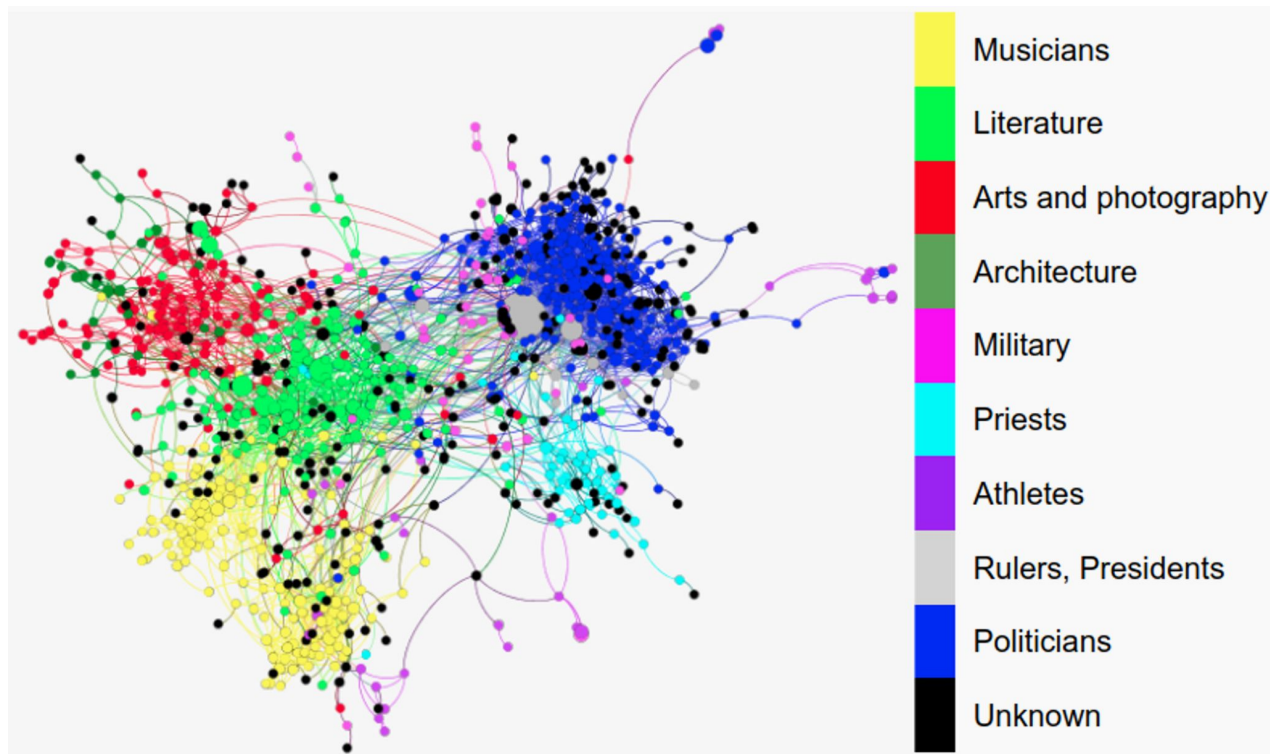
 [Eeva Ahtisaari](#)  
 [Paavo Lipponen](#)  
[Ester Ståhlberg](#)  
[Tarja Halonen](#)  
[Ilmari Kivinen](#)  
 [Alli Paasikivi](#)  
 [Signe Relander](#)

### Tänne viittaavat sivut (2)

[Mauno Koivisto](#) [Kati Peltola](#)

# BiographySampo

## Network Analysis



Aalto University  
School of Science



UNIVERSITY OF HELSINKI



# BiographySampo

## Articles

– Minna Tamper, Petri Leskinen, Eero Hyvönen, Risto Valjus and Kirsi Keravuori: **Analyzing Biography Collection Historiographically as Linked Data: Case National Biography of Finland**. *Semantic Web – Interoperability, Usability, Applicability*, vol. 14, no. 2, pp. 385-419, IOS Press, 2023. [bib](#) [pdf](#) [link](#)

– Minna Tamper, Petri Leskinen, Kasper Apajalahti and Eero Hyvönen: **Using Biographical Texts as Linked Data for Prosopographical Research and Applications**. *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection. 7th International Conference, EuroMed 2018, Nicosia, Cyprus* (Marinos Ioannides, Eleanor Fink, Raffaella Brumana, Petros Patias, Anastasios Doulamis, João Martins and Manolis Wallace (eds.)), pp. 125-137, Springer-Verlag, November, 2018. [bib](#) [pdf](#) [link](#)

... and more on the [project pages](#)



# AcademySampo

# AcademySampo

1. **Project details**
2. **Keywords**
3. **Networks**
4. **Similarities**

# AcademySampo

## 1. Project

- a. published in 2021
- b. <https://akatemiasampo.fi/en/>

## 2. Source Data

- a. Registers of students of Royal Academy of Turku and University of Helsinki
- b. 28000 students during the years 1640–1899
  - i. *Ylioppilasmatrikkeli 1640–1852: containing approx. 18000 student records collected by Yrjö Kotivuori.*
  - ii. *Ylioppilasmatrikkeli 1853–1899: containing approx. 9500 student records collected by Veli-Matti Autio.*



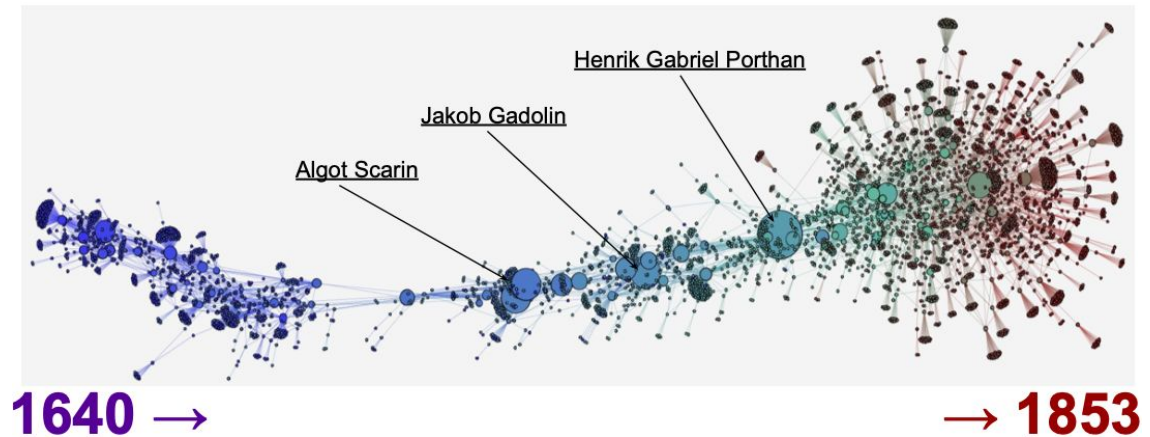
# AcademySampo

1. **Free texts and Datasheet**
2. **Annotations**
  - a. manual
    - i. *Places of birth*
    - ii. *Elementary Schools*
    - iii. *Student Nations*
3. **Networks ja similarities**
  - a. manual annotation
  - b. automatic extraction of named entities
  - c. Wikipedia links
  - d. TF-IDF recommender system based on RDF data

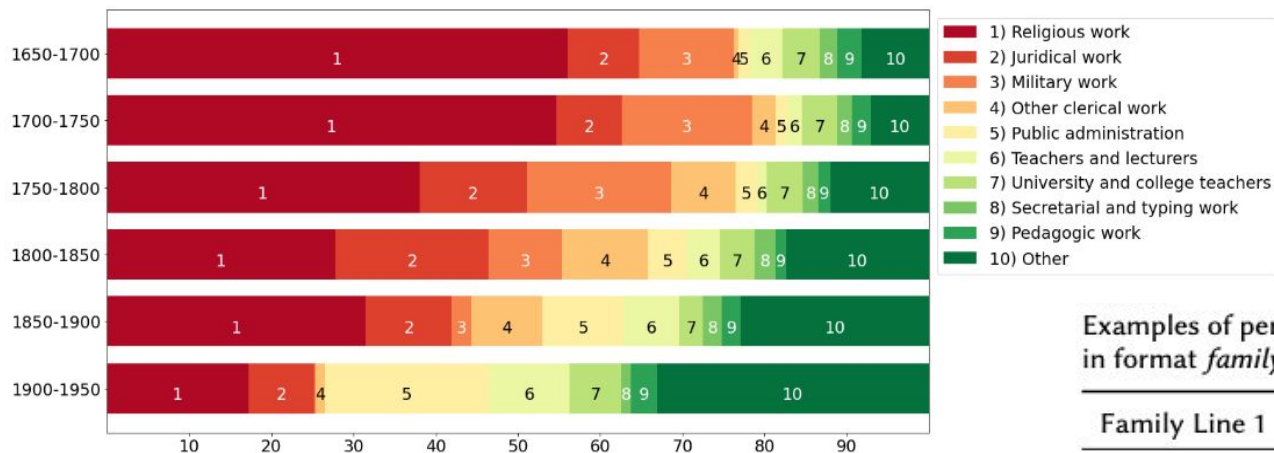


# Network Analysis

- 1) Genealogical family relation network  
long family lines of 7–8 generations
- 2) Teacher-student relation network
- 3) A network based on similarity of life events  
e.g. working for Nobel Company
- 4) Reference network extracted from Wikipedia  
e.g. Duodecim



# Data Analysis



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

# ParliamentSampo

The image depicts the interior of the Finnish Parliament building, known as the Eduskunta. The architecture is semi-circular, featuring a central wooden podium with a speaker's chair. The seating is arranged in a semi-circle, with a balcony level above. The walls are adorned with statues and decorative elements. The text 'ParliamentSampo' is overlaid in white, centered in the image.

# ParliamentSampo

1. **Project details**
2. **Keywords**
  - a. manual
  - b. automatic
3. **Networks**



Aalto University  
School of Science



UNIVERSITY OF HELSINKI



HELDIG  
Centre for Digital Humanities

# Project *ParliamentSampo*

**Parlamenttisampo.fi** published in 2023

**Data publication consists of two interlinked knowledge graphs**

- 1) **Speeches** over 1.000.000  
plenary sessions of the Parliament of Finland
- 2) **Actors**  
~2.900 Members of Parliament of Finland + related groups and organizations

... more on the [project page](#)

# Speech content enrichment 2015–

- **Named entities**
  - Identifying people and places
- **Word lemmatization and part of speech**
  - “Eduskunnan” → “eduskunta”, “eduskunta”/NOUN
  - TurkuNLP Neural parser pipeline
- **Keywords**
  - Annif (a tool developed by the National Library)
    - Keywords from YSO – General Finnish Ontology
  - TF-IDF
- **Library classification (EKS) categories**
  - The Library of Parliament subject headings
  - Unsupervised classifier



Uusimmat puheenvuorot ⓘ	
Tulokset: 135305 puheenvuoroa	
<b>Rajaa:</b>	
Sisältö ⓘ	▼
Puheenvuoron tyyppi ⓘ	▼
Puhuja ⓘ	▼
Puhujan puolue ⓘ	▼
Puhujan sukupuoli ⓘ	▼
Puheenvuoron kieli ⓘ	▼
Päivämäärä ⓘ	▼
Valtiopäivät ⓘ	▼
Vaalikausi ⓘ	▼
Käsiteltävä asiakirja ⓘ	▼
Asiasana (autom. tunnistettu) ⓘ	▼
Mainittu henkilö (autom. tunnistettu) ⓘ	▼
Mainittu paikka (autom. tunnistettu) ⓘ	▼
Asiakohdan kirjastoluokitus (autom. tunnistettu) ⓘ	▼

# Speech content enrichment: YSO and EKS keywords

Asiasana (autom. tunnistettu) ⓘ



Hae...

- poliitikot [55001]
- ministerit [50365]
- kunnat [45334]
- lainsäädäntö [41220]
- kokemukset [37690]
- historia [36811]
- kritiikki [36605]
- henkilöhistoria [36352]
- vaikutukset [35473]
- uudistukset [32780]

Asiakohdan kirjastoluokitus (autom. tunnistettu)



Hae...

- ▼  Elinkeinoelämä [29853]
  - Kauppa, teollisuus, kilpailu, kuluttajansuoja [266]
  - Pankki, valuutta [3812]
- ▶  Hallinto [55759]
- ▶  Hallinto-oikeus [44189]
- ▶  Historia [1083]
- ▶  Kansainvälinen oikeus [14064]
- ▶  Kansainvälinen politiikka [4844]
- ▶  Kansainvälinen yksityisoikeus [191]
- ▶  Kansantalous, julkinen talous [37411]

# YSO Keywords by parties

Speeches 2015–

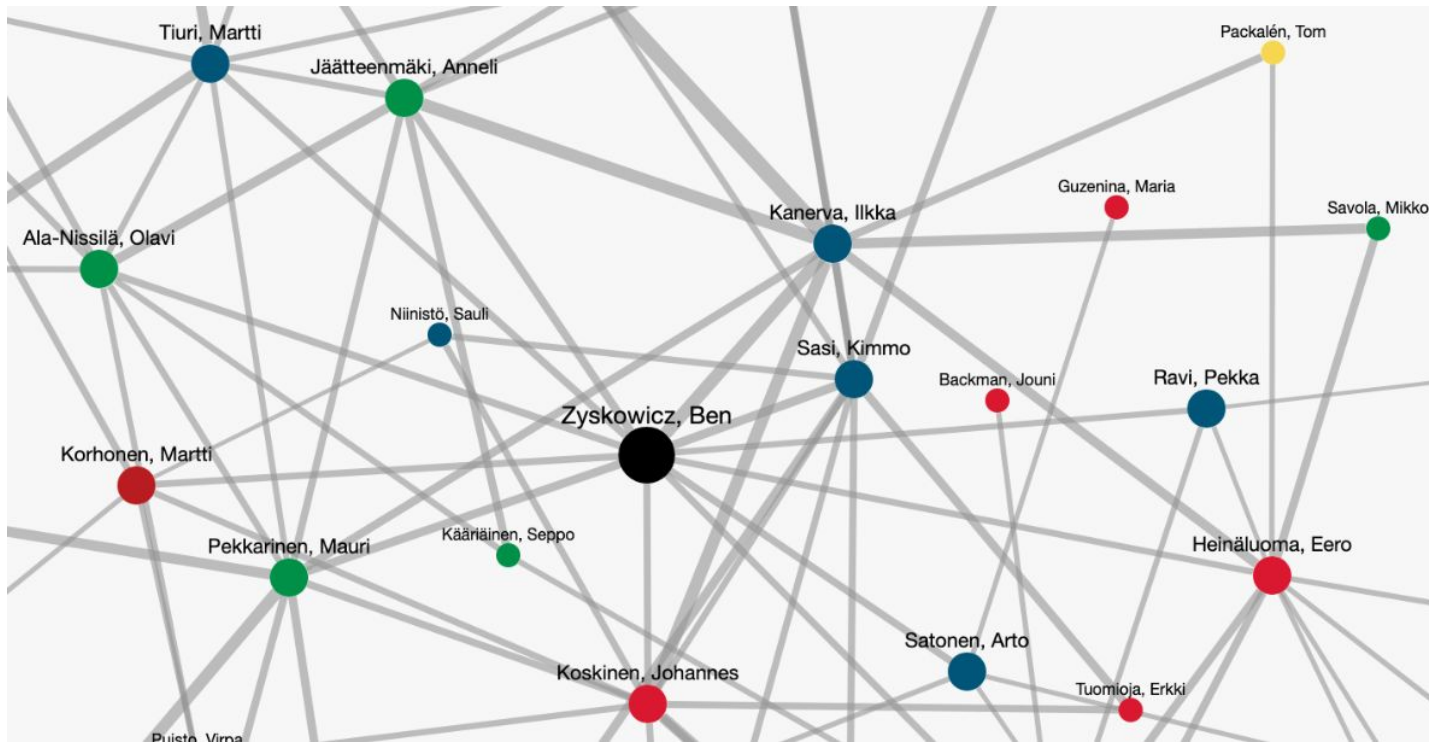
	Social Democratic Party of Finland	Centre Party	Finns Party	Green League	Christian Democrats
1	people with memory disorders	wood as building material	peat resources	loss of nature	freedom of religion
2	intermediate labour market	private roads	expulsion (expatriation)	carbon neutrality	abortion
3	media culture	broadband networks	illegal immigration	natural diversity	evangelical Lutheran churches
4	job security	small towns	aliens policy	nature reserves	euthanasia
5	memory disorders	thinning	migrants	recreation areas	pregnancy
6	orphans	bioeconomy	peat production	air protection	christianity
7	bargaining system	wood (natural materials)	- synon visual communication	protected areas	church (phenomena)
8	form (structural)	county land use planning	immigration	biodiversity	marriage
9	trades union congresses	methane	representation (mutual activity)	living environment	maternity
10	employers' organisations	privately owned forests	fuel tax	recreational use	religions (belief systems)
11	protection against arbitrary dismissal	forest ownership	Aliens Act	natural resources	paternity
12	non-socialist parties	forest owners	fuel peat	environmental awareness	problems with alcohol
13	wage and salary earners	forest resources	party subsidies	environmental policy	self-determination
14	incomes policy	wood-processing industry	support of kinship care	climate	medicine (science)
15	labour market agreements	biogas	thanks	diversity	healthcare personnel

# MP data

- **Information of approx. 2900 current or past members**
- **Biographical data**
  - Family and given names, birth, death, profession, place of living, ...
- **Career and membership events**
  - Education, Career, Electoral Districts, Parliamentary Groups, Parliamentary Bodies and Committees, Positions of Trust, Governmental Memberships, ...
- **Further information**
  - Member's publications, publications about the member, private interests...
- **... extracts for other sources**
  - from databases of Finnish Government and Wikidata



# MP network



Aalto University  
School of Science

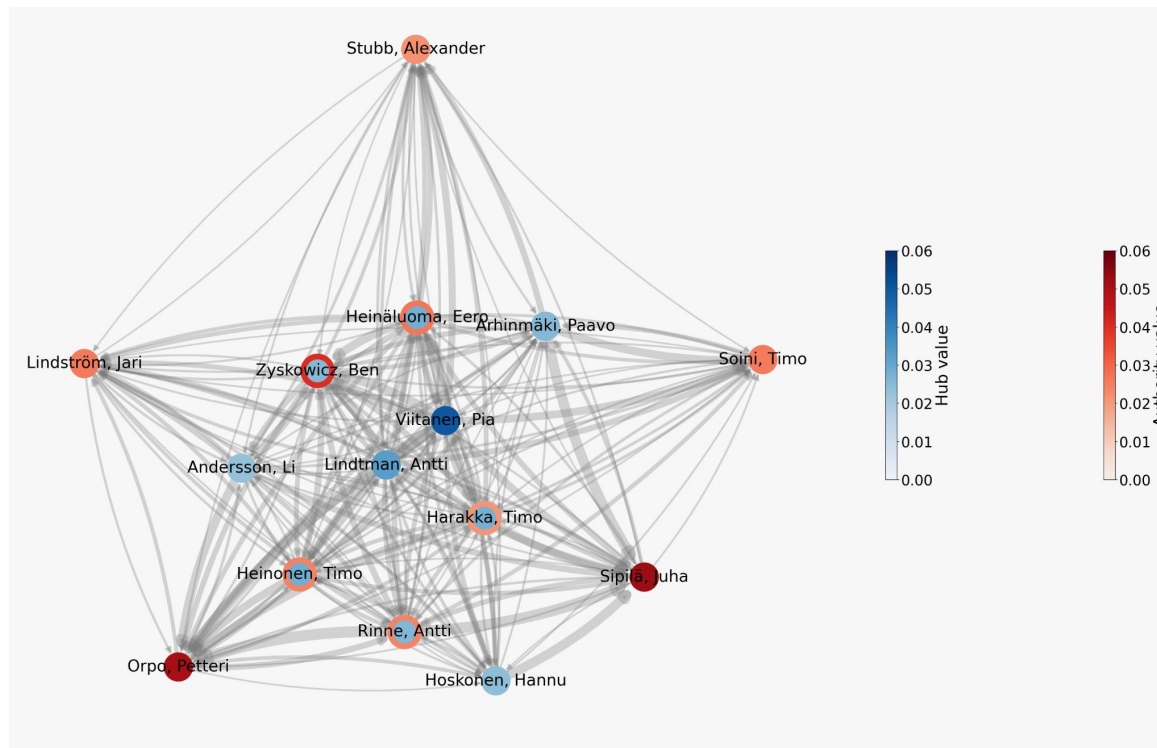


UNIVERSITY OF HELSINKI



HELDIG  
Centre for Digital Humanities

# Reference network



Thank you

