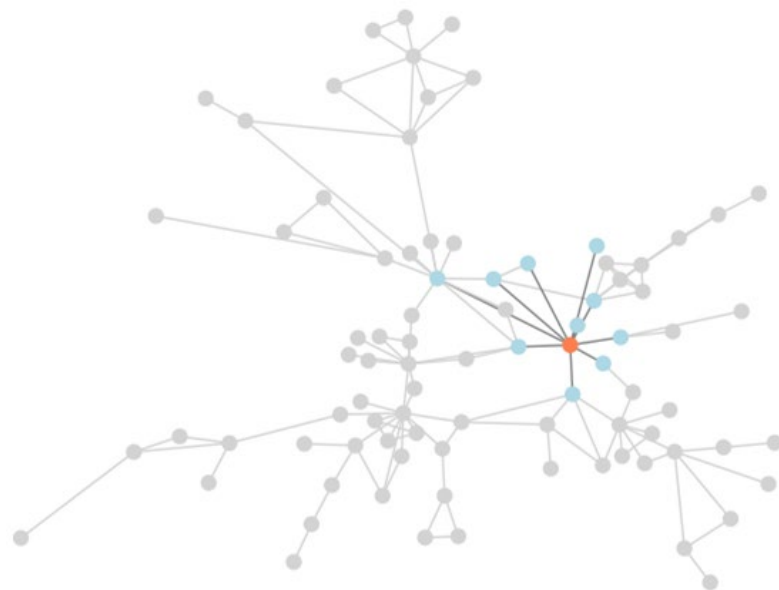


Kirjeenvaihtoverkostojen analysointi Kirjesammon avulla

*Suomalaisen Kirjallisuuden Seura 27.5.2025
Henna Poikkimäki, väitöskirjatutkija, Aalto-yliopisto*

Yleistä kirjeenvaihtoverkostoista

- Ego-verkosto vs. “koko verkosto”
- Analyysin tasoja
 - **Kirjeiden metadata**
 - Kirjeiden sisältö
 - Muu informaatio (biografiat, uutiset,...)
 - Ajallinen ulottuvuus

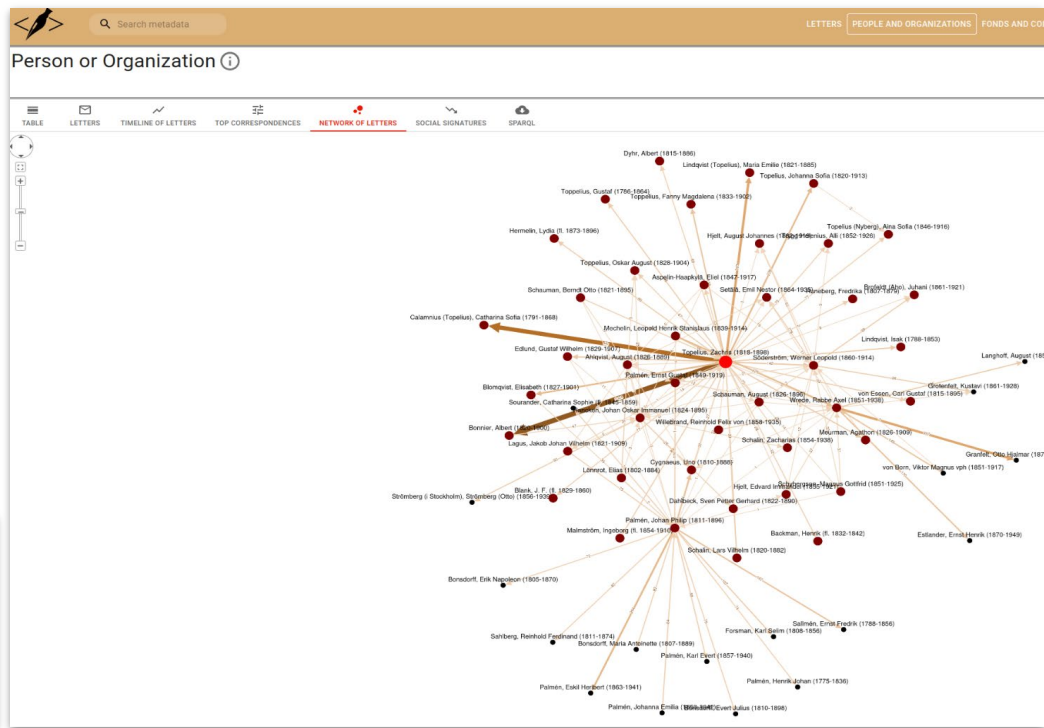


Kirjeenvaihtoverkostot portaalissa

- Interaktiivinen verkostovisualisointi jokaiselle toimijalle
- Yleisimpiä verkostometriikoita ja vastaavat sijoitukset koko verkostossa

Network metrics ⓘ

- Cluster Number: 4.675117e-03 (#1930)
- Core Number: 32 (#1)
- Eigenvector Centrality: 1.752925e-01 (#5)
- Hits_Authority: 2.674315e-06 (#36)
- Hits_Hub: 1.819718e-05 (#68)
- Number of Correspondences: 2302 (#2)
- Pagerank Centrality: 6.40573e-03 (#5)
- Total number of letters: 11887 (#4)
- Weighted In-Degree: 7430 (#6)
- Weighted Out-Degree: 4457 (#3)



SPARQL + Jupyter Notebook

- SPARQL-kysely suoraan koodista
- JSON-muotoinen tulos
muunnetaan sopiviksi datatyypeiksi
ja sopivaan formaattiin

```
import pandas as pd
import rdflib as rdflib
from rdflib.namespace import XSD, RDF, RDFS, Namespace, SKOS, OWL
from SPARQLWrapper import SPARQLWrapper, JSON, POST, N3, REXML
```

```
ENDPOINT = "http://ldf.fi/coco/sparql"
```

```
PREFIXES = ""
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX : <http://ldf.fi/schema/coco/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
""
```

```
sparql = SPARQLWrapper(ENDPOINT)
sparql.setReturnFormat(JSON)
sparql.setMethod(POST)
sparql.setCredentials(user, password)
```

```
source_query = ""
SELECT DISTINCT ?datasource ?label
WHERE {
  [] a :MetadataRecord ;
  dct:source ?datasource .
  ?datasource skos:prefLabel ?label .
  FILTER (lang(?label) = 'en')
}
```

```
sparql.setQuery(PREFIXES + source_query)
results = sparql.query().convert()
```

```
results_as_dict = [dict([(k, v['value']) for k,v in r.items()]) for r in results]
pd.DataFrame(results_as_dict)
```



	datasource	label
0	http://ldf.fi/coco/source/hameenlinna-city-mus...	Hämeenlinna City Museum
1	http://ldf.fi/coco/source/nationalarchive	The National Archives of Finland
2	http://ldf.fi/coco/source/postalmuseum	Postal Museum
3	http://ldf.fi/coco/source/	Serlachius Museums

Yleistä datasta

- Yli 1,2 miljoona kirjettä
- Yli 118000 toimijaa
 - Henkilöt, perheet, organisaatiot
- Kirjeisiin ja toimijoihin liittyvää metadataa sekä linkkejä ulkoisiin lähteisiin

Data Source

	Letters	Actors
Åbo Akademi University Library	366614	27350
The National Archives of Finland	292073	32325
The National Library of Finland	281157	33632
The Society of Swedish Literature in Finland (SLS)	198490	16013
Finnish Literature Society (SKS)	116646	13882
Finnish National Gallery	14402	3528
Elias Lönnrot Letters	6296	1054
J. V. Snellman Letters	1563	4756
Zacharias Topelius Writings	1407	65
Migration Institute of Finland	1342	214
Albert Edelfelt Letters	1310	5802
Theatre Museum	665	171
Hämeenlinna City Museum	438	169
Serlachius Museums	411	136
Aalto University Archives	295	261
Gallen-Kallela Museum	144	3
Postal Museum	81	165
The Archives of President Urho Kekkonen	26	10



Aalto University
School of Science



UNIVERSITY OF HELSINKI



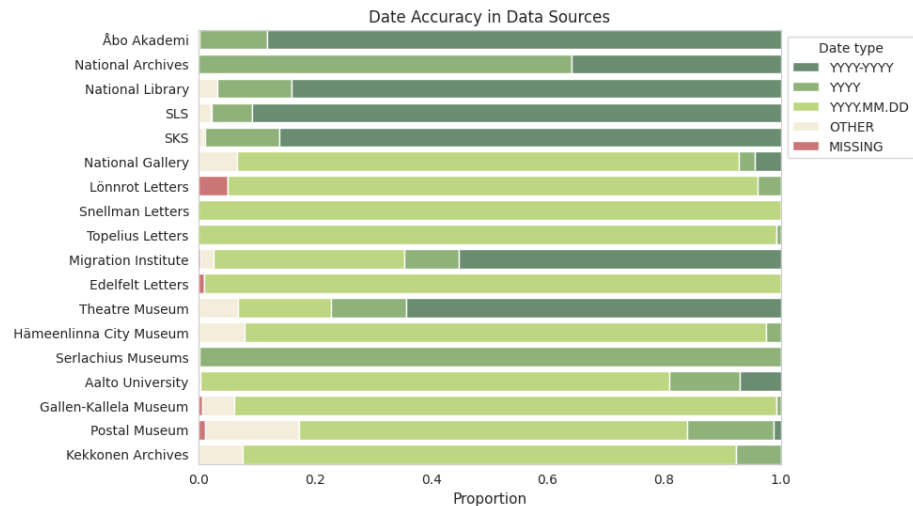
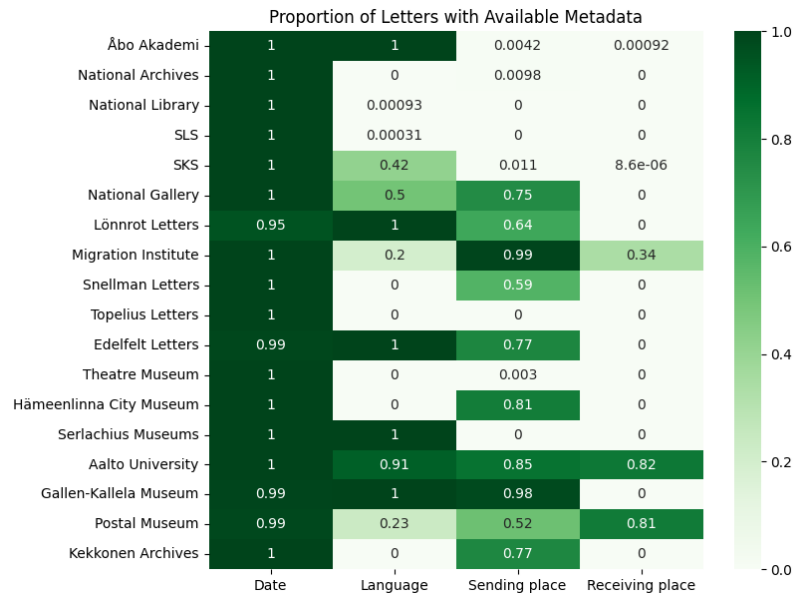
SUOMALAISEN
KIRJALLISUUDEN
SEURA



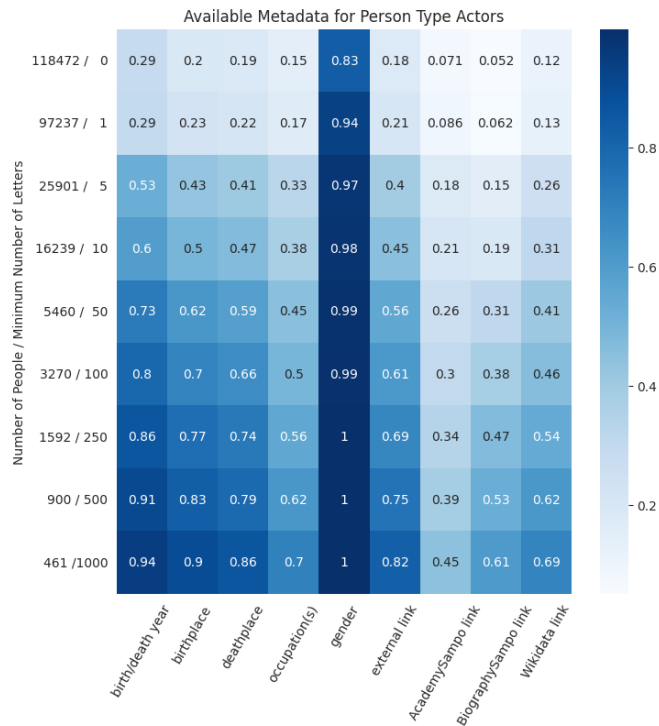
HELDIG
Helsinki Centre for Digital Humanities

HELSINKI INSTITUTE FOR
SOCIAL SCIENCES
AND HUMANITIES

Metadatan kattavuus: Kirjeet



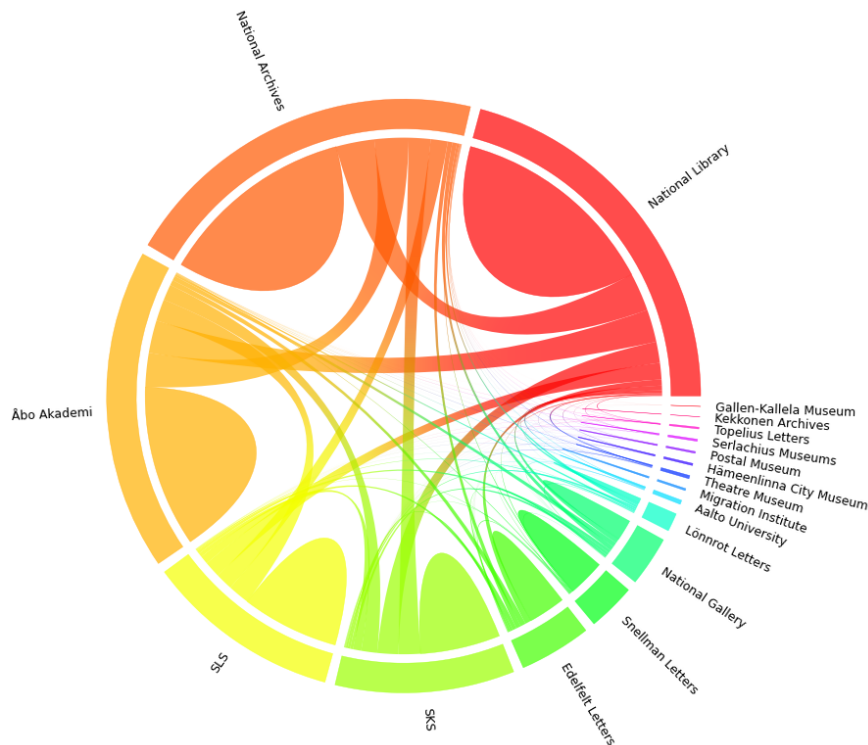
Metadatan kattavuus: Henkilöt



Ihmiset, joiden lähettämiä ja vastaanottamia kirjeitä on paljon löytyvät todennäköisemmin useista eri tietolähteistä ja heistä tiedetään usein myös enemmän.

Henkilöiden linkittäminen lähteiden välillä on haastavaa.

Henkilöiden jakautuminen tietolähteisiin

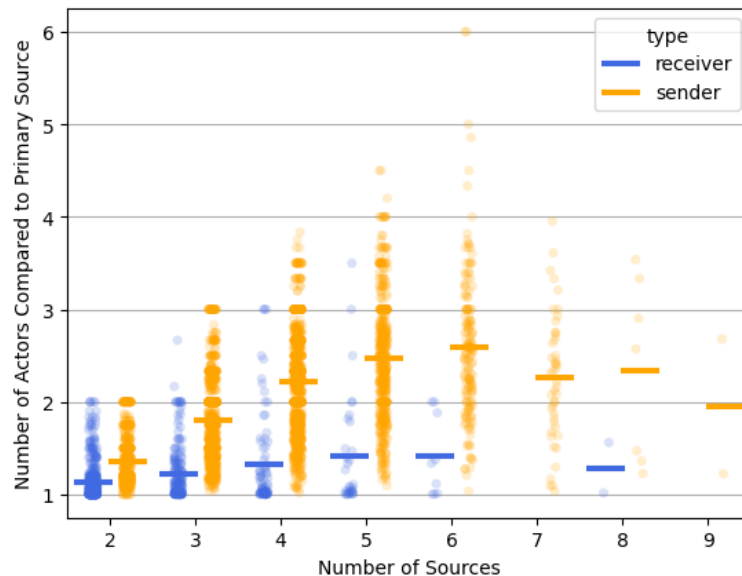
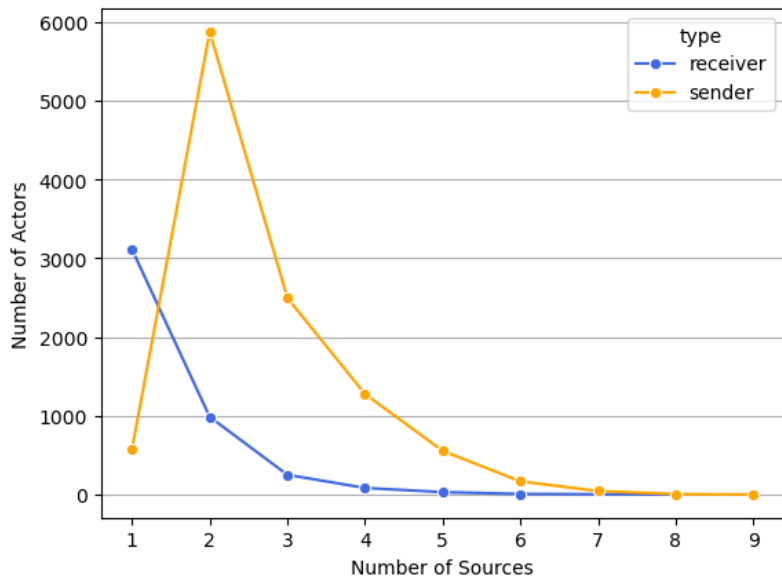


Jotkin toimivat esiintyvät useissa lähteissä

- Noin 10% kaikista toimijoista
- Heistä 61% vaihtanut kirjeitä alle kuuden toimijan kanssa
- Muutoksia ego-verkostossa, kun useampi lähde huomioidaan

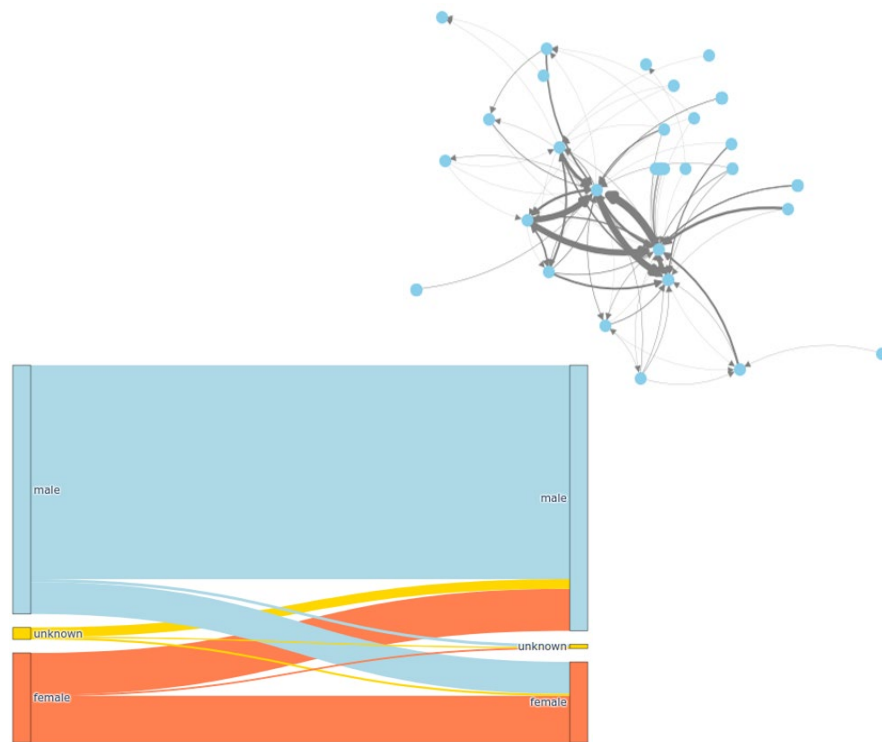
Vaikka ego-verkostoon ei tulisi muutoksia, henkilön asema laajemman verkoston kontekstissa voi selkiytyä

Useaan tietolähteeseen kuuluvat toimijat



Esimerkki: henkilöiden kirjeenvaihtoverkosto

- Suunnattu painotettu verkosto
- Lähes 89000 henkilöä, joista
 - 66% miehiä, 27% naisia
 - 92% on vastaanottanut enintään yhden kirjeen
 - 50% on lähettänyt enintään yhden kirjeen

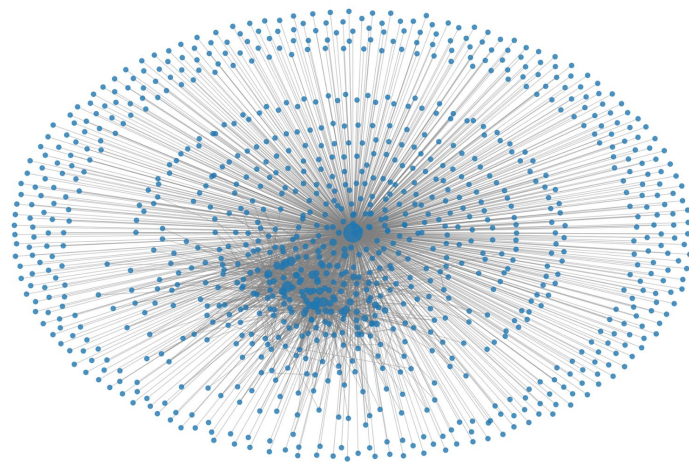
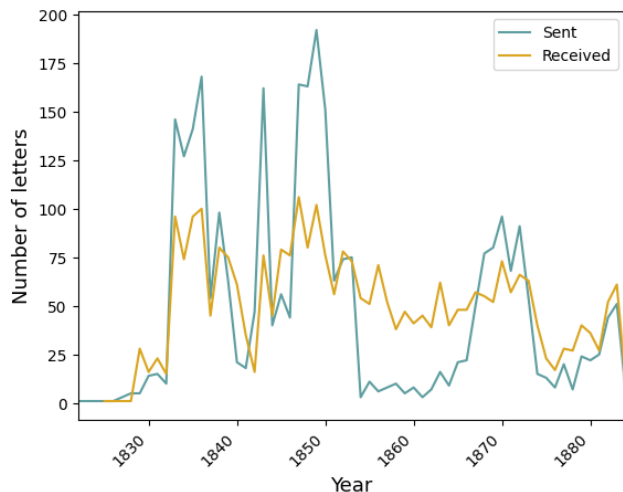


Suurimmat kirjeiden lähettäjät ja vastaanottajat

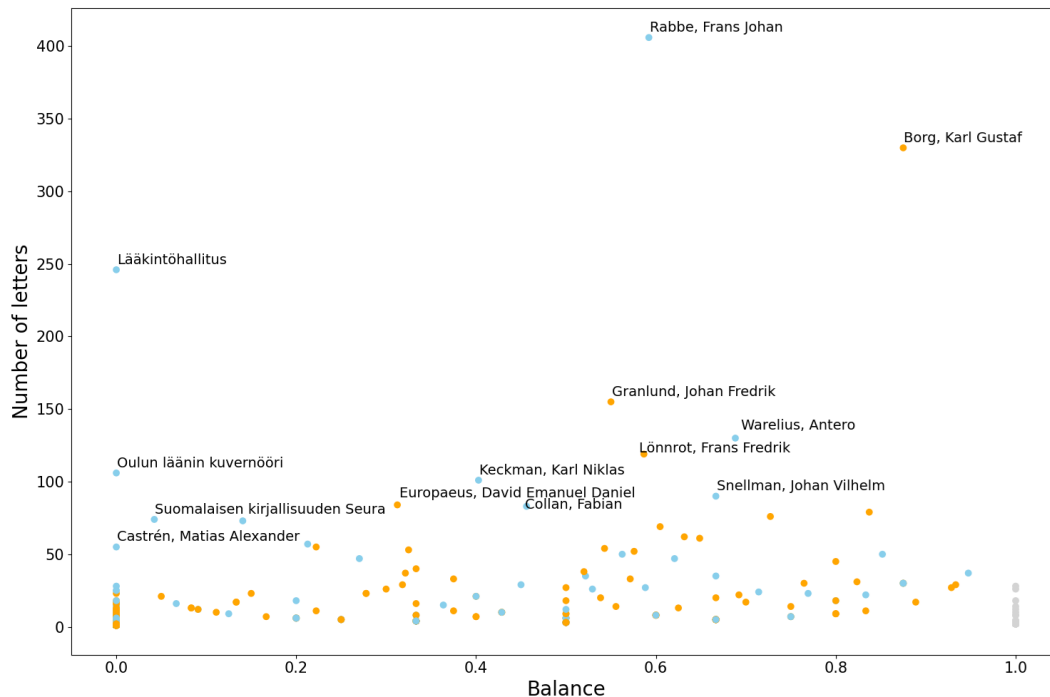


Esimerkki: Elias Lönnrotin ego-verkosto

- 3211 lähetettyä kirjettä 499 toimijalle
- 3312 vastaanotettua kirjettä 764 toimijalta
- Tarkat lähetyspäivät yleensä tunnettu



Esimerkki: Elias Lönnrotin ego-verkosto



“Tasapaino”

- 1 = Lönnrot ja toinen osapuoli ovat lähettäneet yhtä paljon kirjeitä toisilleen
- 0 = vain Lönnrot (sininen) tai toinen osapuoli (oranssi) on lähettänyt kirjeitä

Alhainen tasapaino voi kertoa mahdollisesti puuttuvista kirjeistä.

Yhteenveto

- **Kirjesammon kaltaiset systeemit**
 - Helpottaa laajempien, metadata tason verkostojen analysointia
 - Rikastaa mahdollisesti ego-verkostoja
- **Metadataan perustuvien verkostojen analyysi**
 - Metadata saatavuus ja tarkkuus vaikuttaa analyysiin
 - Potentiaalisesti kiinnostavien ilmiöiden tunnistaminen datasta
 - Voi auttaa valitsemaan sopivan joukon kirjeitä lähilukua varten



Aalto University
School of Science



UNIVERSITY OF HELSINKI



SUOMALAISEN
KIRJALLISUUDEN
SEURA



HELDIG
Helsinki Centre for Digital Humanities

HELSINKI INSTITUTE FOR
SOCIAL SCIENCES
AND HUMANITIES



Kiitos!