Identity & Representation: Patterns in Historical Letter Metadata

Research Questions
1) Epistolary cultures.
2) Individuals in correspondence structures.
3) Data profiles of collections.

Methodologies
1) Python and SPARQL queries.
2) Gephi visualisation & network measurements.
4) Qualitative & historical research.

The Shaped of the Archive

Out of 35,652 people, only six appear in all four datasets contributing to CoCo. Among them was Johan Snellman, one of the most qualitatively important actors in CoCo. His ego archive (approx. 6,300 letters and 1,200 correspondents) contains his ego archive (approx. 6,300 letters and 1,200 correspondents) and his ego archive (approx. 6,300 letters and 1,200 correspondents). Although he was generally regarded as a reserved and humble person, his correspondence reveals a nature of "information gatherer" within the network. This being reflected in the genesis of his work's main subject, the national epic "Kalevala", for which he gathered Finnish folk poetry and literature passed down orally during several expeditions in Finland, Russian Karelia, the Kola Peninsula and the Baltic countries.

Egos in the Network

Using Gephi with NetworkX in Python to visualise the overlap between qualitatively important archives and personal collections in CoCo and CKCC.

Network of correspondences in the CoCo dataset with nodes coloured by the gender (red for female; blue for male; and black for unknown).

Genders in the Network

Machine learning algorithms used to determine the likely gender for the known first name of the actor across datasets.

Epistolary Fingerprints

Challenges & Future Research
- Disambiguation (CoCo) - the amount of actors will become more accurate/the unknowns.
- Using complete ego networks or archival collections as a 'prototype' to model 'complete' networks and estimate missing data and correspondence connections.
- Building more network profiles (like 'Gatherer' vs 'Redundant') to consider similar roles in network structures, both in these datasets and beyond.
- Thinking critically about the role and development of letter metadata in cultural heritage study and preservation.

Identity & Representation: Patterns in Historical Letter Metadata

Research Questions
1) Epistolary cultures.
2) Individuals in correspondence structures.
3) Data profiles of collections.

Methodologies
1) Python and SPARQL queries.
2) Gephi visualisation & network measurements.
4) Qualitative & historical research.

The Shaped of the Archive

Out of 35,652 people, only six appear in all four datasets contributing to CoCo. Among them was Johan Snellman, one of the most qualitatively important actors in CoCo. His ego archive (approx. 6,300 letters and 1,200 correspondents) contains his ego archive (approx. 6,300 letters and 1,200 correspondents) and his ego archive (approx. 6,300 letters and 1,200 correspondents). Although he was generally regarded as a reserved and humble person, his correspondence reveals a nature of "information gatherer" within the network. This being reflected in the genesis of his work's main subject, the national epic "Kalevala", for which he gathered Finnish folk poetry and literature passed down orally during several expeditions in Finland, Russian Karelia, the Kola Peninsula and the Baltic countries.

Egos in the Network

Using Gephi with NetworkX in Python to visualise the overlap between qualitatively important archives and personal collections in CoCo and CKCC.

Network of correspondences in the CoCo dataset with nodes coloured by the gender (red for female; blue for male; and black for unknown).

Genders in the Network

Machine learning algorithms used to determine the likely gender for the known first name of the actor across datasets.

Epistolary Fingerprints

Challenges & Future Research
- Disambiguation (CoCo) - the amount of actors will become more accurate/the unknowns.
- Using complete ego networks or archival collections as a 'prototype' to model 'complete' networks and estimate missing data and correspondence connections.
- Building more network profiles (like 'Gatherer' vs 'Redundant') to consider similar roles in network structures, both in these datasets and beyond.
- Thinking critically about the role and development of letter metadata in cultural heritage study and preservation.