Network regression reveals factors driving the letter correspondence of 16th century reformers

Ramona Roller

Collaborators: Prof. Frank Schweitzer

The European Reformation (1517-1648)

- Transformative movement of society in early modern Europe
  - Division of Catholic Church
  - Major changes in the socio-political system

⚠ Letters were the main means of communication.

⚠ Use them to study the social system in 16th century Europe

Martin Luther's posting of his 95 theses to the church in Wittenberg (1517)

The letter correspondence network of reformers

- **Data**: 20,000 letters, 3,000 people, sending- and (receiving) dates + locations, 1510 - 1575

- **Network**: directed multi-edge network of interactions
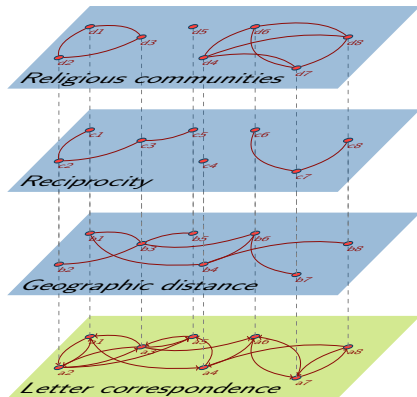 - nodes: reformers
  - edges: letters
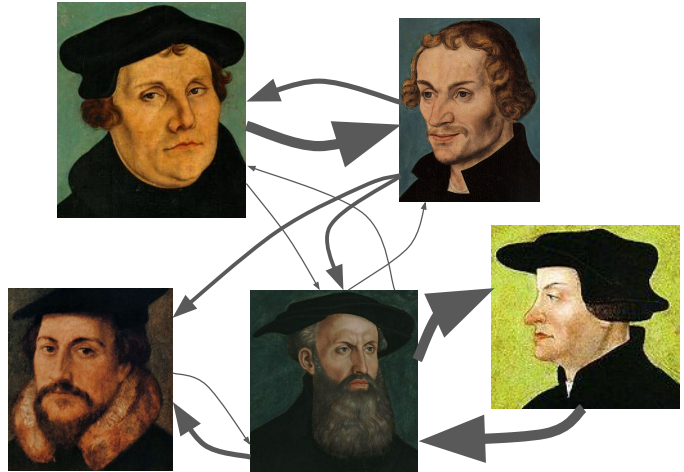
Schematic representation of a sample from the letter correspondence network

The role of geographic distance on letter correspondence

Research question

How do social relations affect the letter correspondence, i.e. the network topology?

Social relations (**R**) between sender and receivers to be tested:

- **Geographic distance** (tested):
  Long distances: letters are convenient but costly;
  Short distances: letters are inconvenient but cheap

- **Reciprocity** (control):
  Social norm of rewarding kind actions

- **Religious communities** (control):
  Support for same/different religious denominations
 E.g. Lutherans, Reformed, Calvinists, Baptists, etc.

Regression approach

Linear regression

- $\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + ... + \beta_p \mathbf{x}_p + \varepsilon$
- E.g. $\mathbf{y}$: number of letters per reformer, $\mathbf{x}_i$: religious denomination, age, etc.
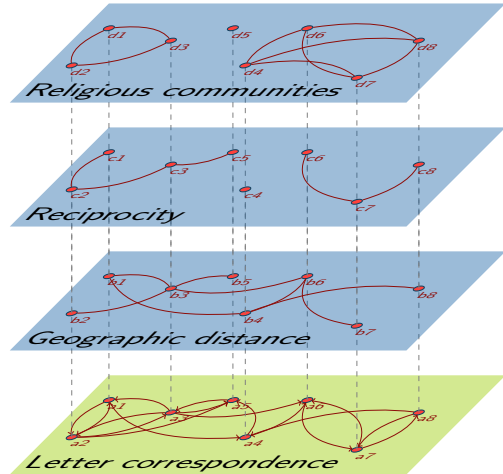- **Problem: Networks do not meet independence assumption**

Network regression (Casiraghi, 2017; Casiraghi et al., 2016)

- Statistical model based on **generalised hypergeometric network ensembles** (gHypE)
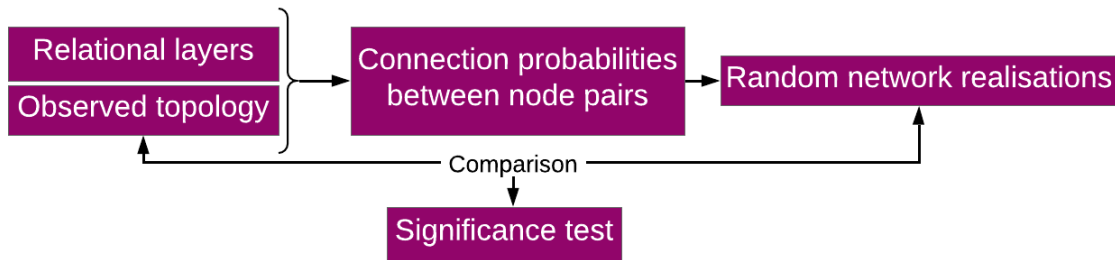
Network regression output

- **Regression coefficients** β_k
  - Quantify importance of relational layers

- **Propensity matrix** $\boldsymbol{\Omega}$
  - **Odds ratio** $\boldsymbol{\Omega}_{ij}/\boldsymbol{\Omega}_{mn}$: How much more likely are nodes i and j to be connected compared to nodes m and n?
  - $\boldsymbol{\Omega} := \prod_{k=1}^{K} \mathbf{R}_k^{\beta_k}$
    where each relational layer corresponds to one $\mathbf{R}_k$

Religious communities

Reciprocity

Geographic distance

Letter correspondence
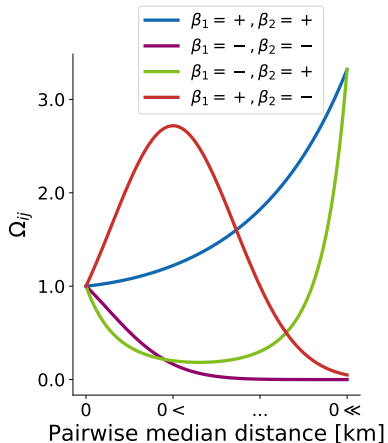
Predictor construction

❶ Geographic distance

- **cost** (distance $\uparrow$, #letters $\downarrow$);
  convenience (distance $\uparrow$, #letters $\uparrow$)
- $\mathbf{R}_{ij}^{(1)} = e^{dist_{ij}}$, $\mathbf{R}_{ij}^{(2)} = e^{dist_{ij}^2}$
- $\mathbf{\Omega} = \mathbf{R}^{(1)\beta_1} * \mathbf{R}^{(2)\beta_2}$: Covers all possible combinations of cost and convenience

❷ Reciprocity

- $\mathbf{R}^{(3)} = \mathbf{A}^T$ (change statistic Snijders, 2006)
- $\mathbf{R}_{ij}^{(3)}$: number of letters i would have to send to j in order to answer each letter of j to i

❸ Religious communities

- Assume homophily
- Same: $\mathbf{R}_{ij}^{(4)} = 10$, different: $\mathbf{R}_{ij}^{(4)} = 1$ Casiraghi, 2017

Legend for plot:
- $\beta_1 = +, \beta_2 = +$
- $\beta_1 = -, \beta_2 = -$
- $\beta_1 = -, \beta_2 = +$
- $\beta_1 = +, \beta_2 = -$

y-axis: Ω_{ij} (3.0, 2.0, 1.0, 0.0)

x-axis: Pairwise median distance [km] (0, 0 <, ..., 0 ≪)

- Only convenience
- Only cost
- Either cost or convenience
- Cost and convenience in balance

Results: reduced model $\Omega_{ij} = \left(e^{dist_{ij}}\right)^{\beta_1} * \left(e^{dist_{ij}^2}\right)^{\beta_2}$

	reduced
Distance	
Linear distance	$7.885 \ (0.159)^{***}$
Quadratic distance	$-17.918 \ (0.405)^{***}$
AIC	43427.830
McFadden $pseudo - R^2$	0.009

$^{***}p < 0.001, \ ^{**}p < 0.01, \ ^{*}p < 0.05$

- **Optimal intermediate distance**: At 168km people are most likely to send letters.

- **Odds ratio**: $\Omega_{168km}/\Omega_{0km} = 1.29$, $\Omega_{168km}/\Omega_{1000km} = 28809$
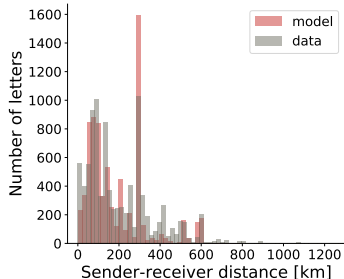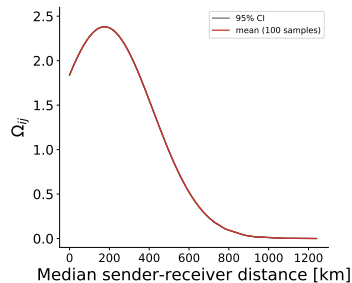
Results: **full model**

	reduced	full
Distance		
Linear distance	$7.885 \, (0.159)^{***}$	$-3.354 \, (0.176)^{***}$
Quadratic distance	$-17.918 \, (0.405)^{***}$	$5.032 \, (0.388)^{***}$
Controls		
Reciprocity		$0.461 \, (0.004)^{***}$
Religious homophily		$0.276 \, (0.016)^{***}$
AIC	43427.307	33989.210
McFadden *pseudo* $- R^2$	0.009	0.224

$^{***}p < 0.001, \, ^{**}p < 0.01, \, ^{*}p < 0.05$

🢒 The **full model is better** than the reduced as the smaller AIC shows.

🢒 The **sign flip** of the distance predictors shows that the controls are essential.

Summary

❶ Insights on the letter correspondence network of reformers

- People are likely to write letters if they ...
  - live close to or far away from each other
  - have high reciprocity
  - support the same religious denomination
- Tested for possible cost-convenience relations

Take home message

Network regression:

Relations explain interactions

❷ Benefits of network regression

- Takes interdependence of samples into account
- Can deal with missing data ($\mathbf{R}_{ij} = 1$ ➜ β has no effect)
- Construction of predictors is not restricted: Use any kind of quantifyable relation, test hypotheses.

❸ Outlook

- Tailor predictor selection towards specific theories of historical research
- Include node attributes as explanatory variables

Network regression

gHypE depends on four $N \times N$ matrices

- **Adjacency matrix A**: given

- **Combinatorial effects matrix Ξ**: covered by configuration model

- **Propensity matrix Ω**: to be computed from predictor matrices **R**'s

$$\boldsymbol{\Omega} := \prod_{k=1}^{K} \mathbf{R}_k^{\beta_k}$$

Religious communities

Reciprocity

Geographic distance

Letter correspondence

- **Odds ratio** Ω_{ij}/Ω_{mn}: How much more likely are nodes i and j to be connected compared to nodes m and n?
- Each **predictor matrix** $\mathbf{R}_k$ encodes one relational network layer
- $\mathbf{R}_{ij}$ can quantify the relation directly or encode some specific assumptions
- The larger $\mathbf{R}_{ij}$ the larger the propensity to be connected of node pair ij
- β_k are the estimated regression coefficients quantifying the importance of one layer
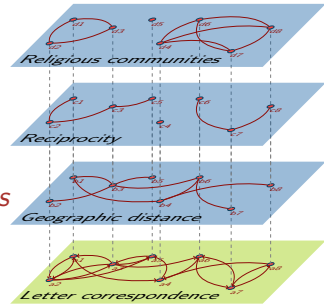
Collinearity causes sign flip

	Reciprocity	Religion
Distance		
Linear distance	$-3.758\ (0.172)^{***}$	$8.283\ (0.164)^{***}$
Quadratic distance	$5.584\ (0.381)^{***}$	$-18.552\ (0.410)^{***}$
Controls		
Reciprocity	$0.457\ (0.004)^{***}$	
Religious homophily		$0.219\ (0.016)^{***}$
AIC	34229.532	43271.460
McFadden $pseudo - R^2$	0.219	0.012

$^{***}p < 0.001,\ ^{**}p < 0.01,\ ^{*}p < 0.05$

- Corr(linear distance, reciprocity) $= 0.265$
- Corr(quadratic distance, reciprocity) $= 0.268$
- Corr(linear distance, religion) $= -0.022$
- Corr(quadratic distance, religion) $= -0.021$
- Corr(reciprocity, religion) $= -0.002$