

Styles and meanings of sociological knowledge

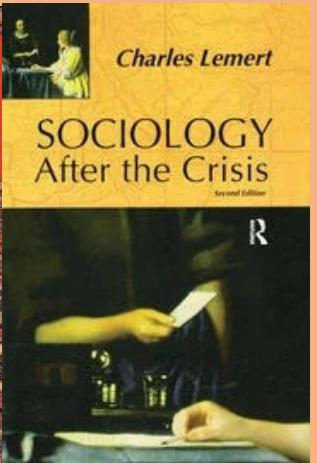
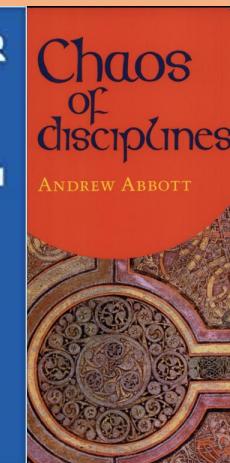
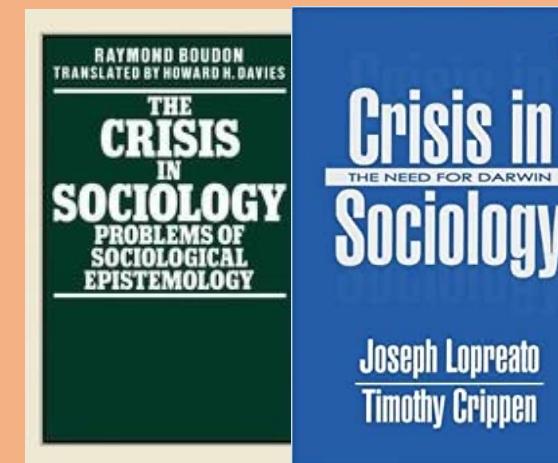
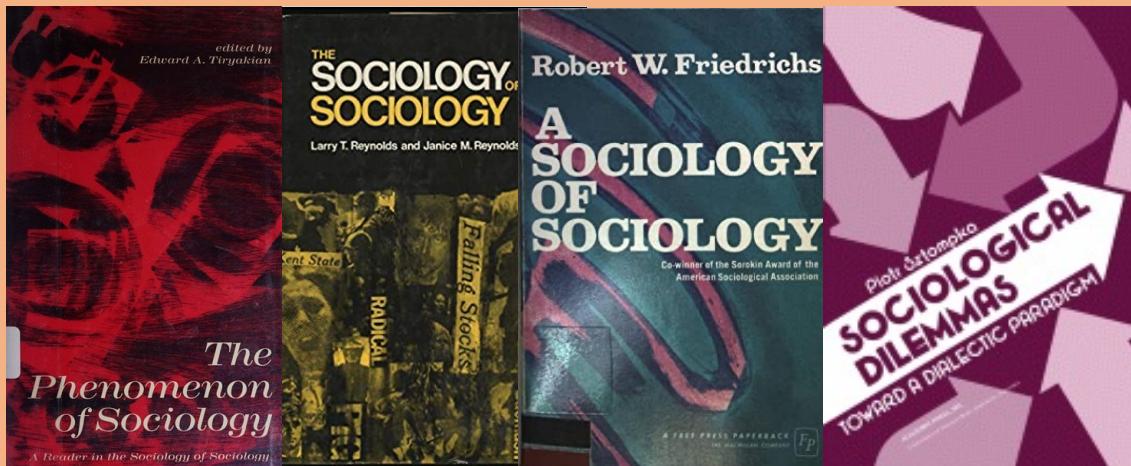
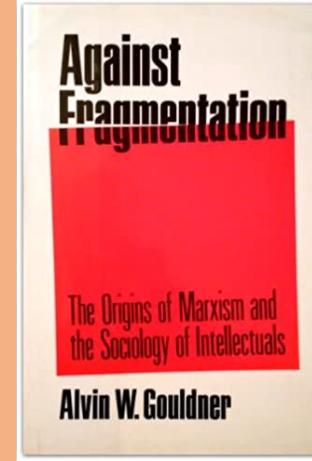
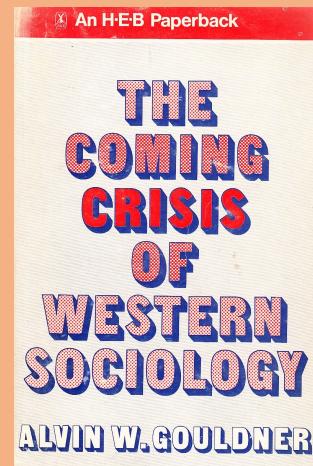
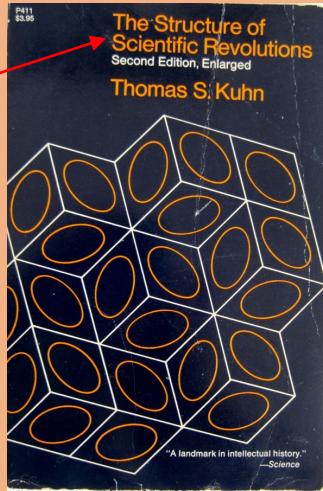
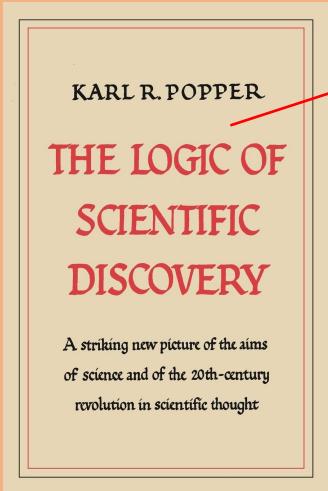
Josef Ginnerskov

PhD candidate in sociology

Department of Sociology

Uppsala University

The Sociology of Sociology

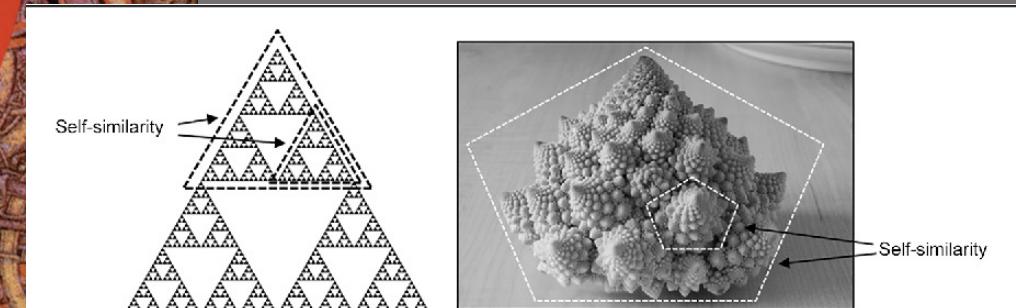
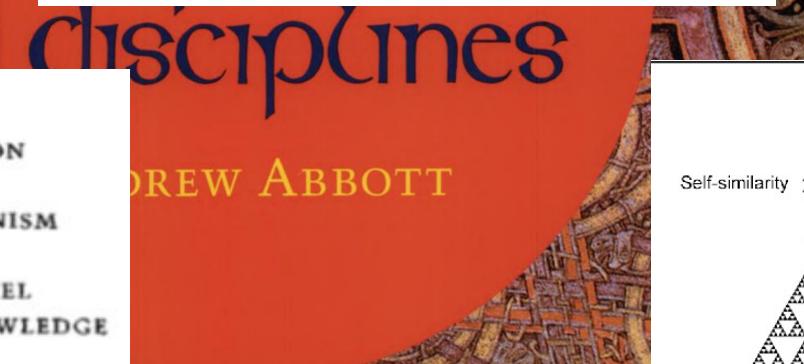
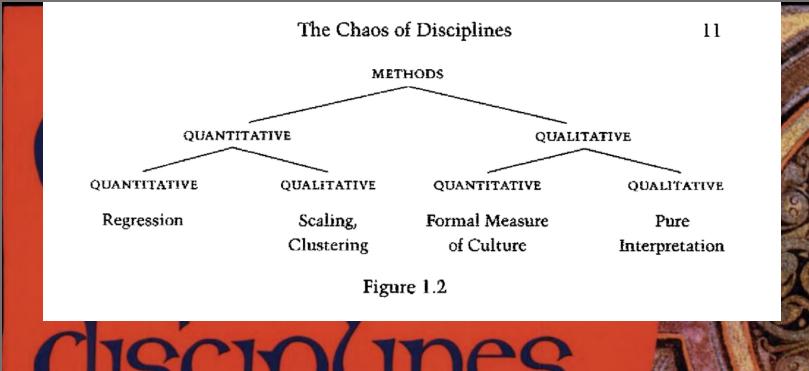
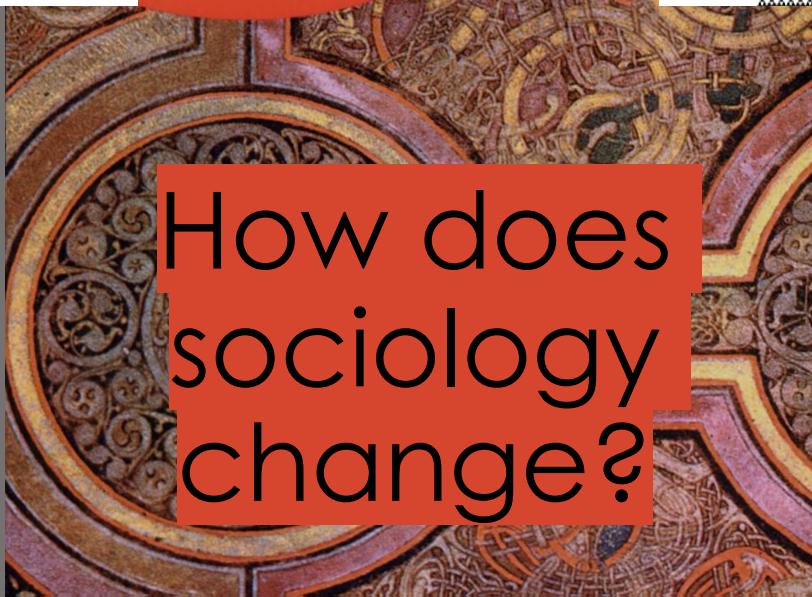


e.g.

versus

Quantitative
POSITIVISM
ANALYSIS
REALISM
SOCIAL STRUCTURE
INDIVIDUAL LEVEL
TRANSCENDENT KNOWLEDGE

Qualitative
INTERPRETATION
NARRATIVE
CONSTRUCTIONISM
CULTURE
EMERGENT LEVEL
SITUATED KNOWLEDGE



The problem with sociology's "evolution"

Can computers shed
new light on sociology
beyond its “lore”?

Computational text analysis

“The sociological imagination”



“Lost in translation”



<https://i.pinimg.com/236x/2c/52/f9/2c52f94aa34d58fb67747fc2b058efb9--funny-sign-fails-funny-signs.jpg>

“The computational imagination”



<https://cio.ucop.edu/this-computer-scientist-beats-hackers-at-their-own-game/>

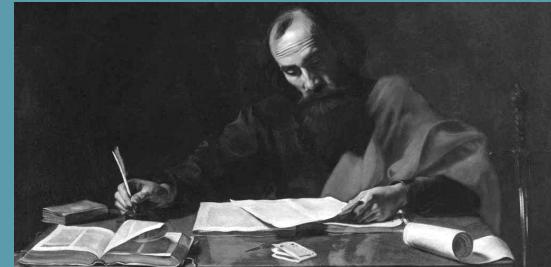
Text mining “might even carry the potential to overcome the qualitative—quantitative divide in the social sciences” (Keuschnigg et al., 2018, p. 8)

Hermeneutists operating machines

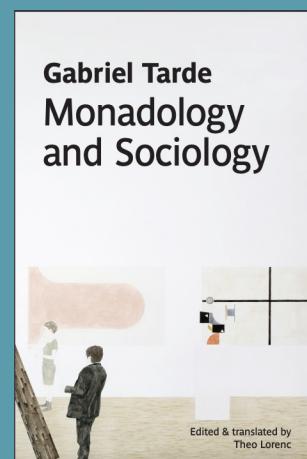
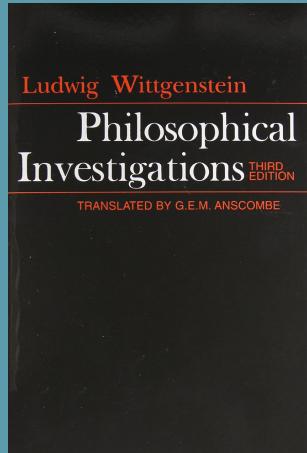
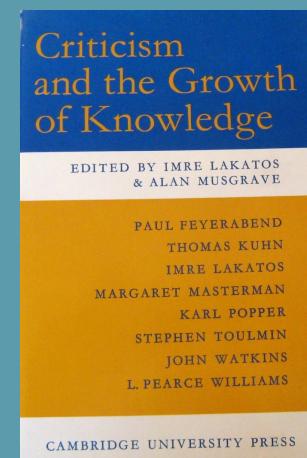
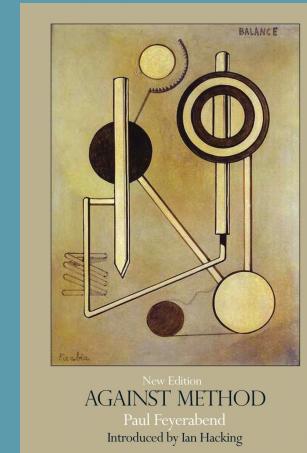


The qualitative interpretation of computational outputs is one of the most underdeveloped parts of Computational Social Science –
David Lazer, 2021-02-04, Institute of Analytical Sociology (Online seminar)

Cultural sociology



VS

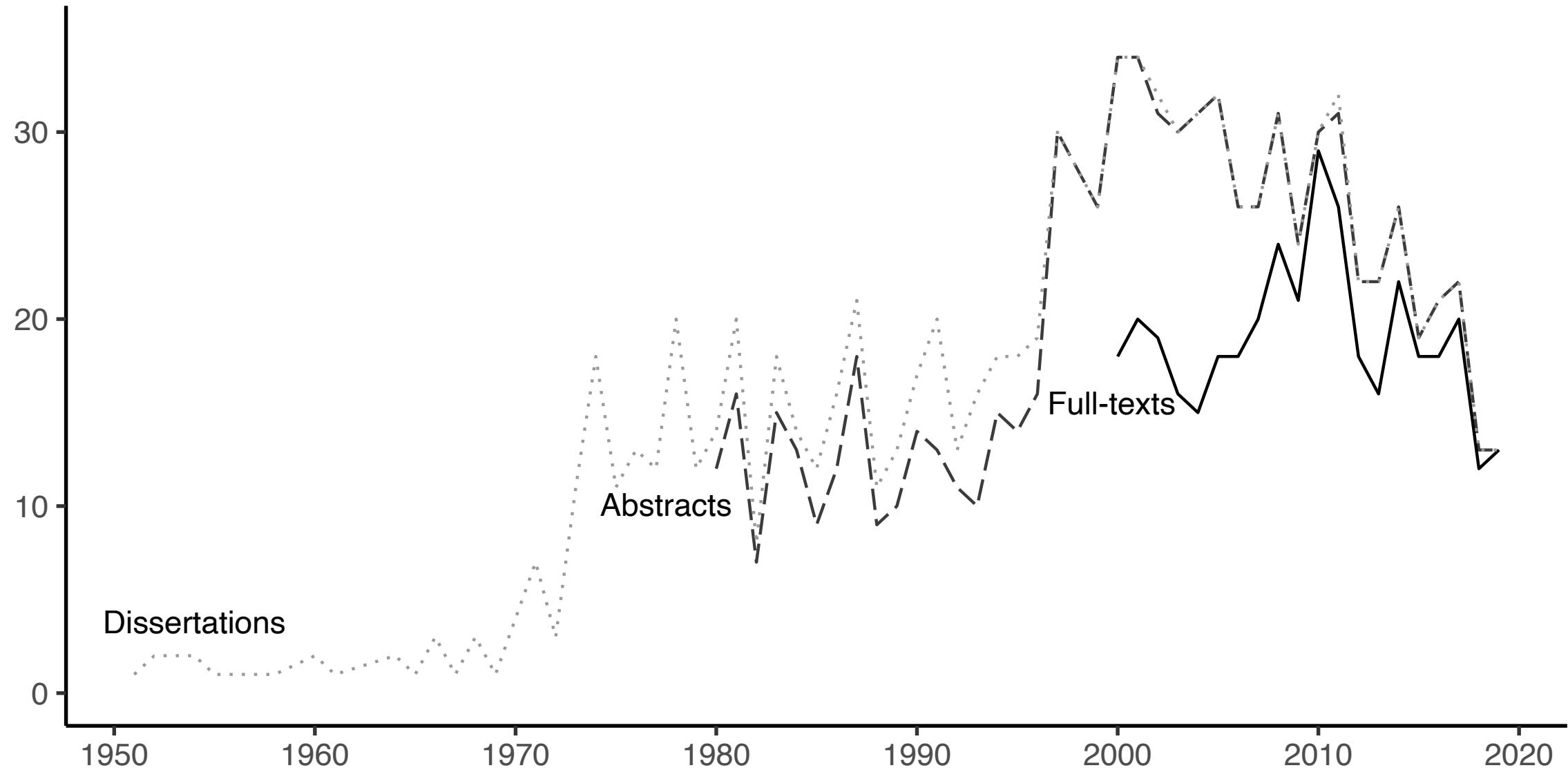


The study
– an empirical revisit to
the sociology of sociology by
mining dissertations
instead of (Anglo-Saxon)
journal article abstracts

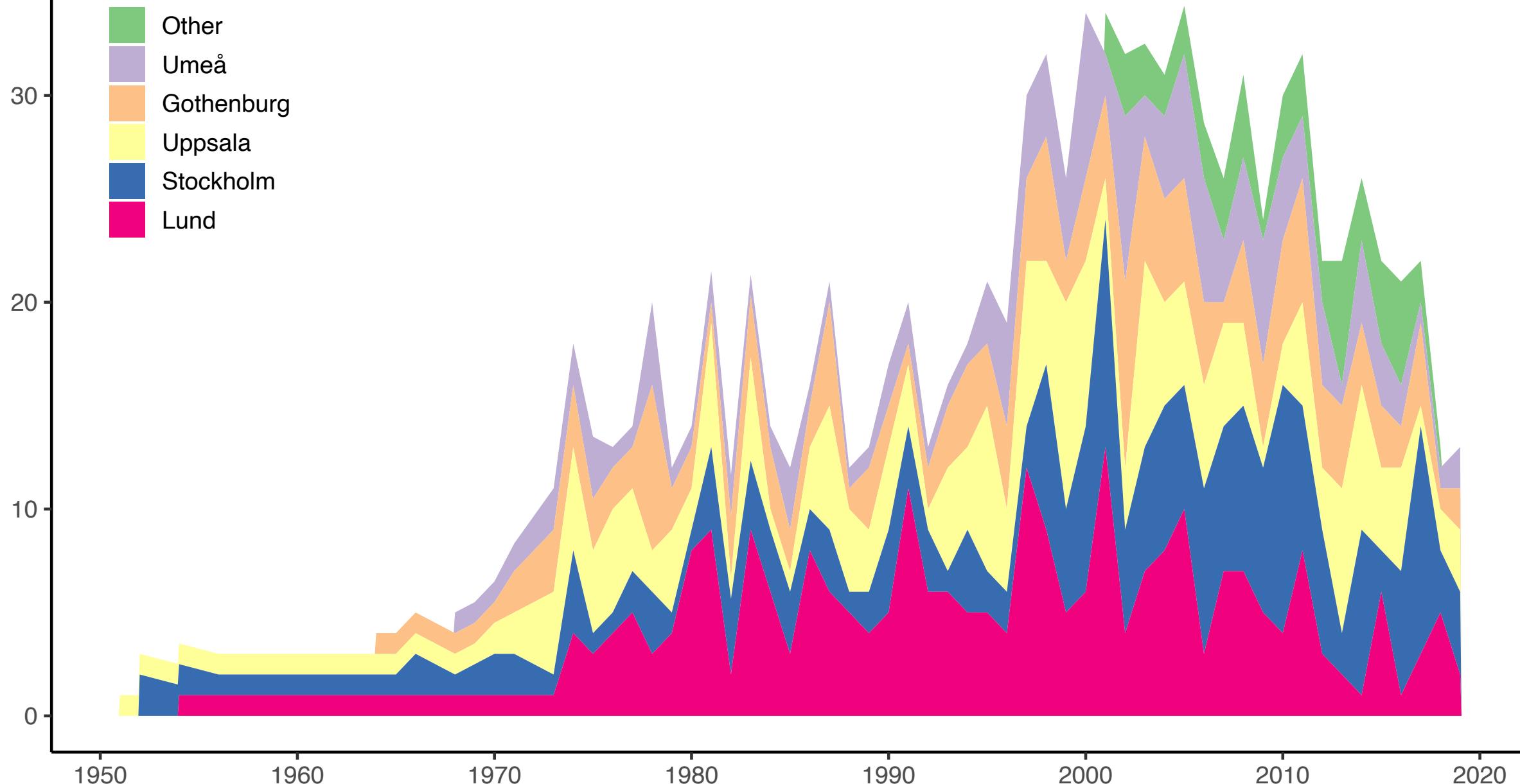
There is no bibliography of
sociology dissertations defended in Sweden...



Dissertations x data type



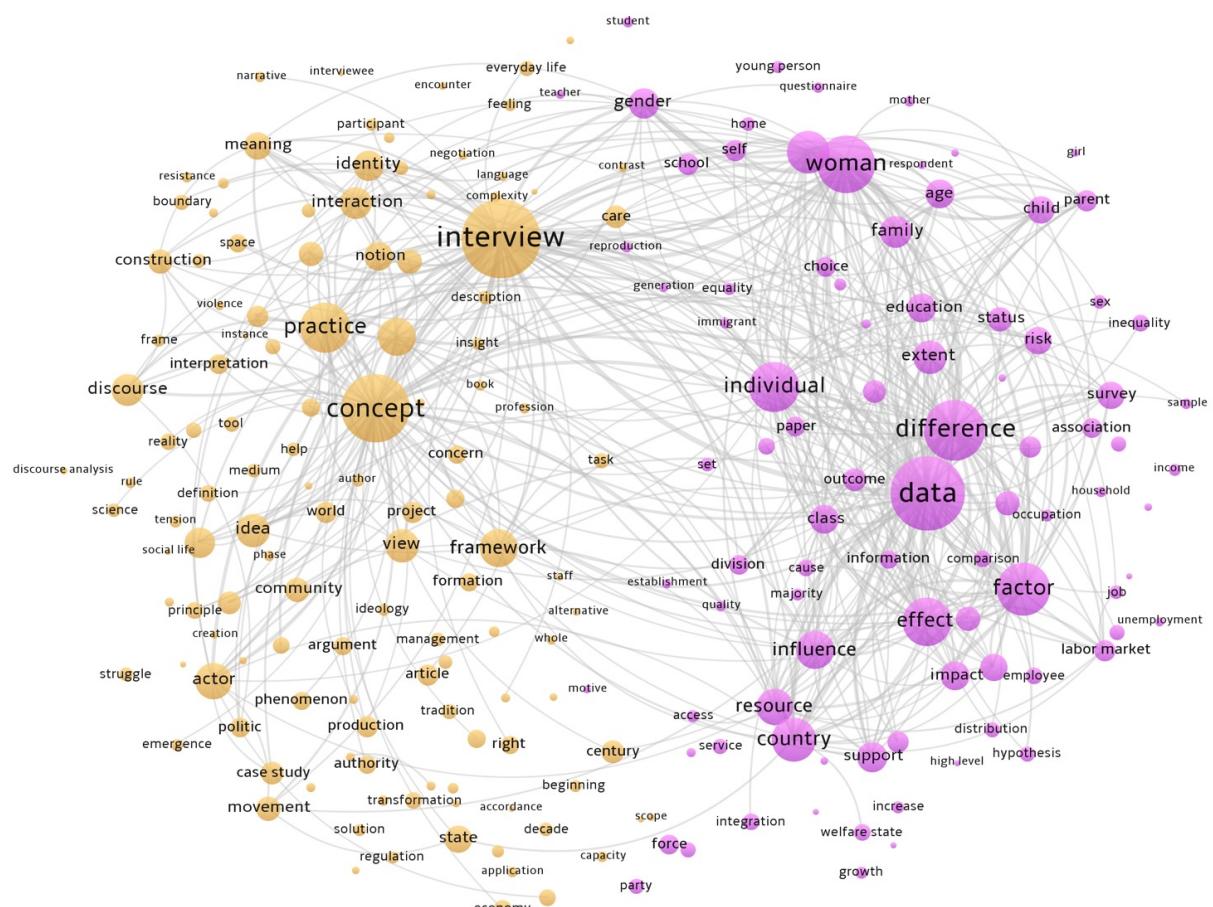
Dissertations x university



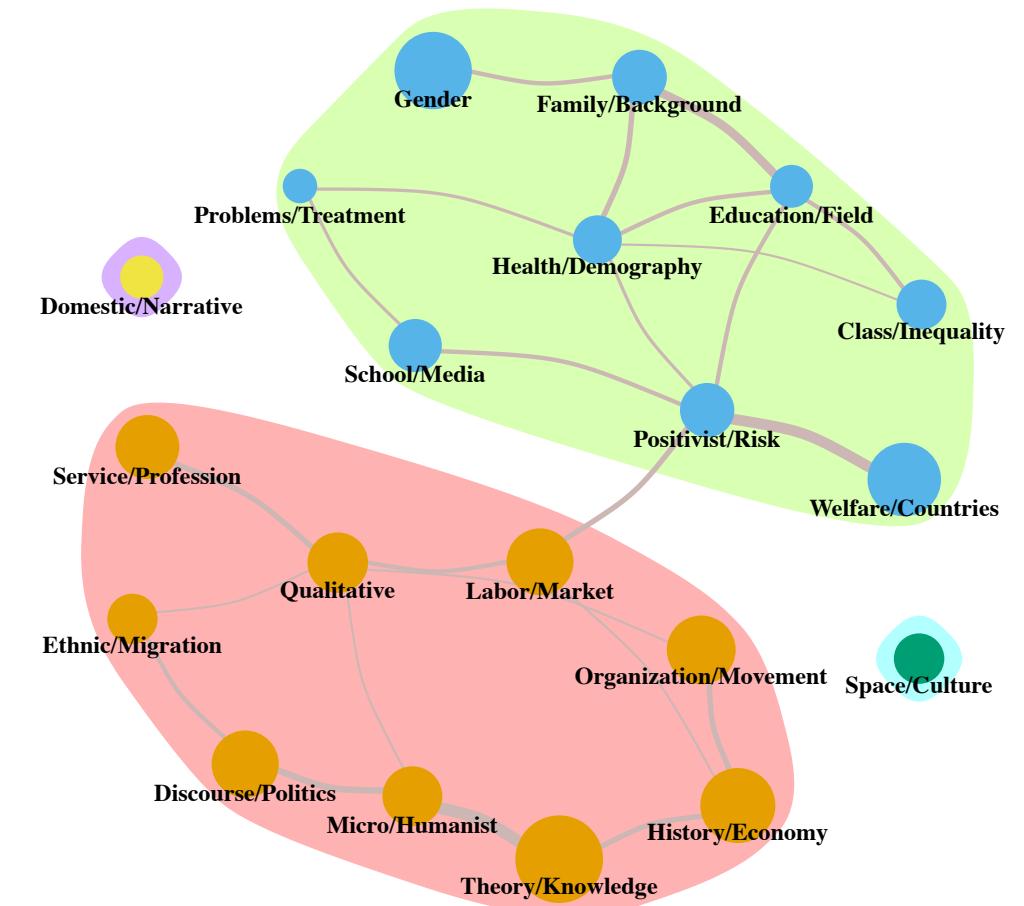
Text analyses performed (1/2)

The abstracts corpus (n=815), 1980-2019

Sub-study 1. Word-correlation networks



Sub-study 2. Structural topic modeling

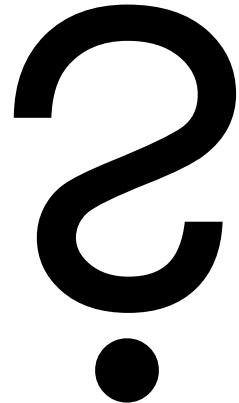
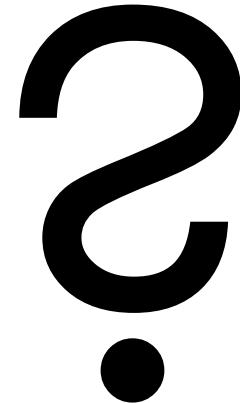


Text analyses performed (2/2)

The full-text corpus (n=380), 2000-2019

Sub-study 3. Stylometry

Sub-study 2. Word embeddings



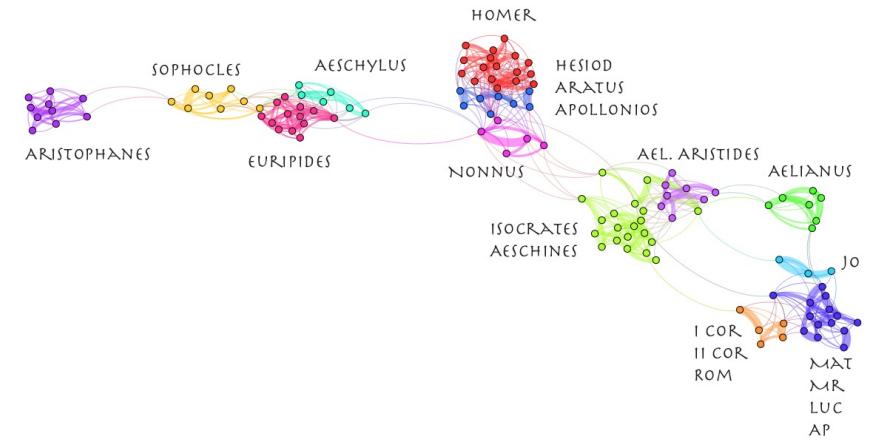
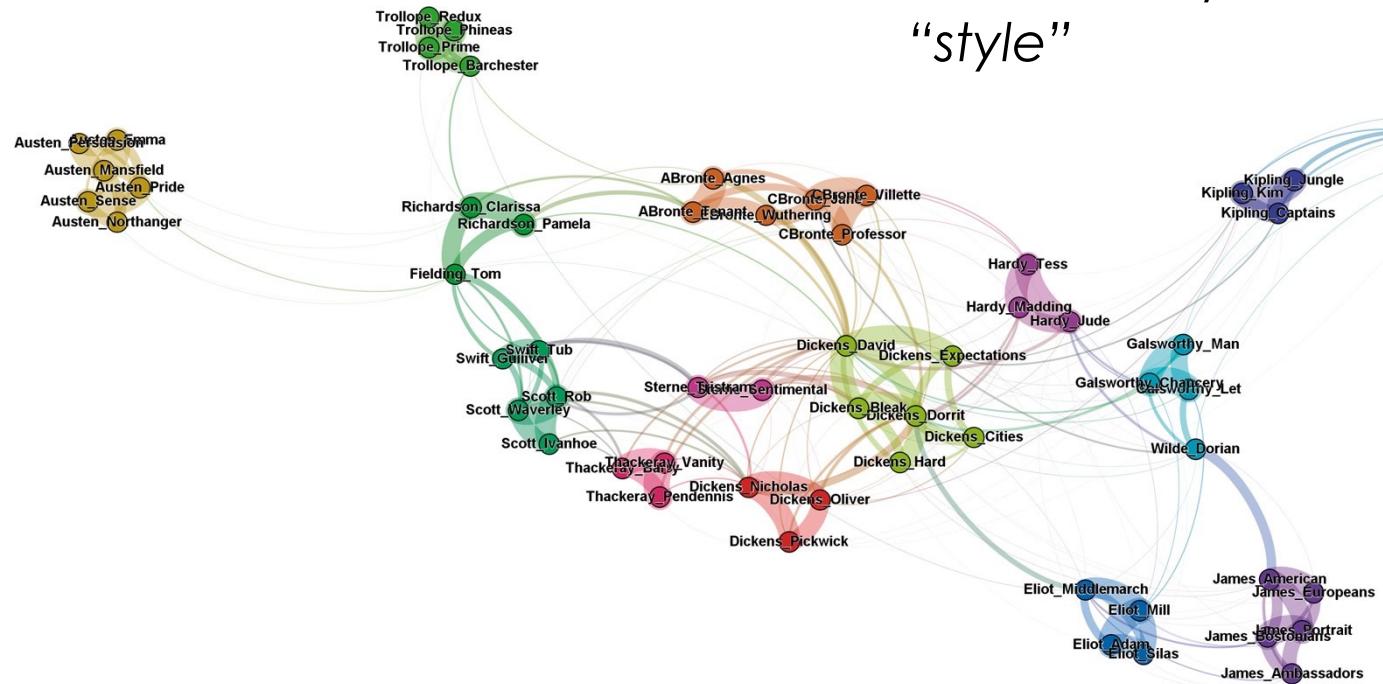
Stylometrics

– are there distinct styles in sociology?

Bootstrap consensus networks

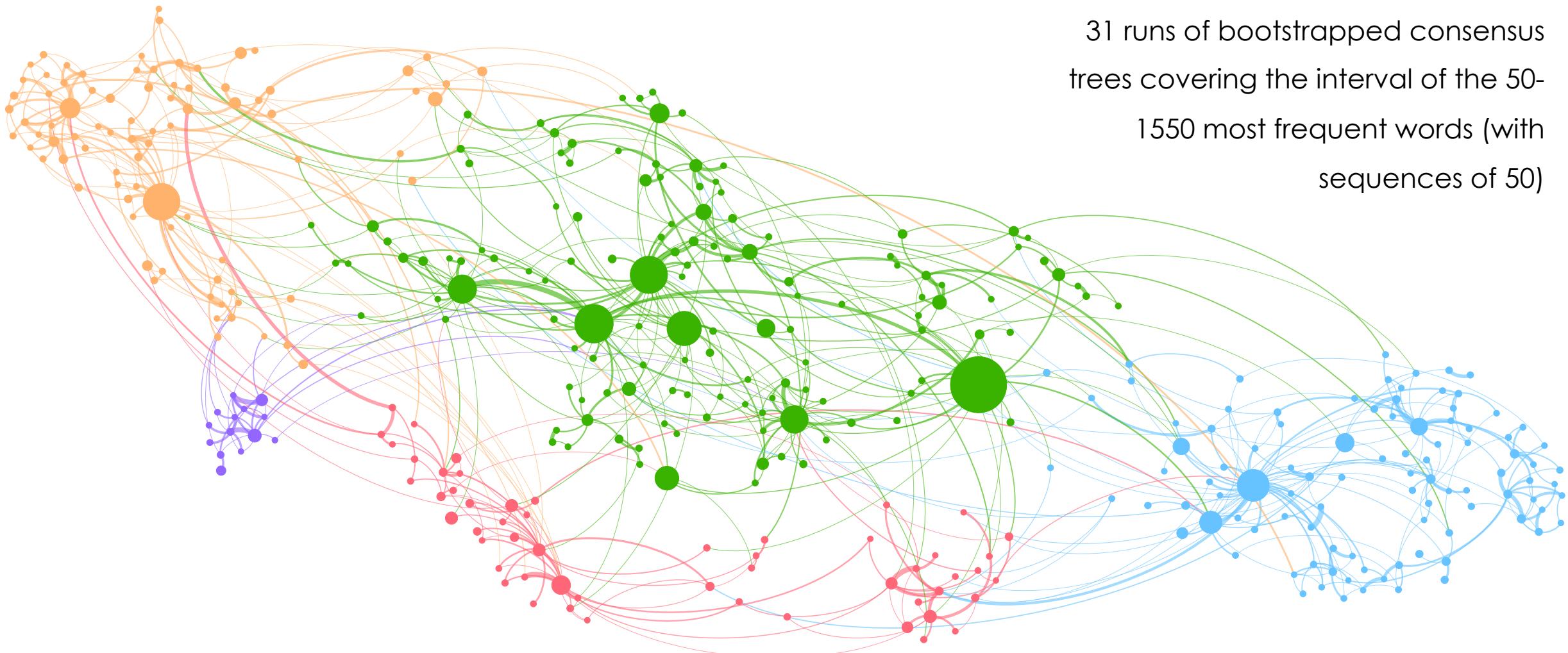
– calculating and visualizing style

The most frequent words used by an author tend to be rather consistent – these constitute his/her “style”



If we go with the hypothesis that authors within a genre (to some extent) have similar styles (compared to authors in other genres) – is sociology constituted by different styles?

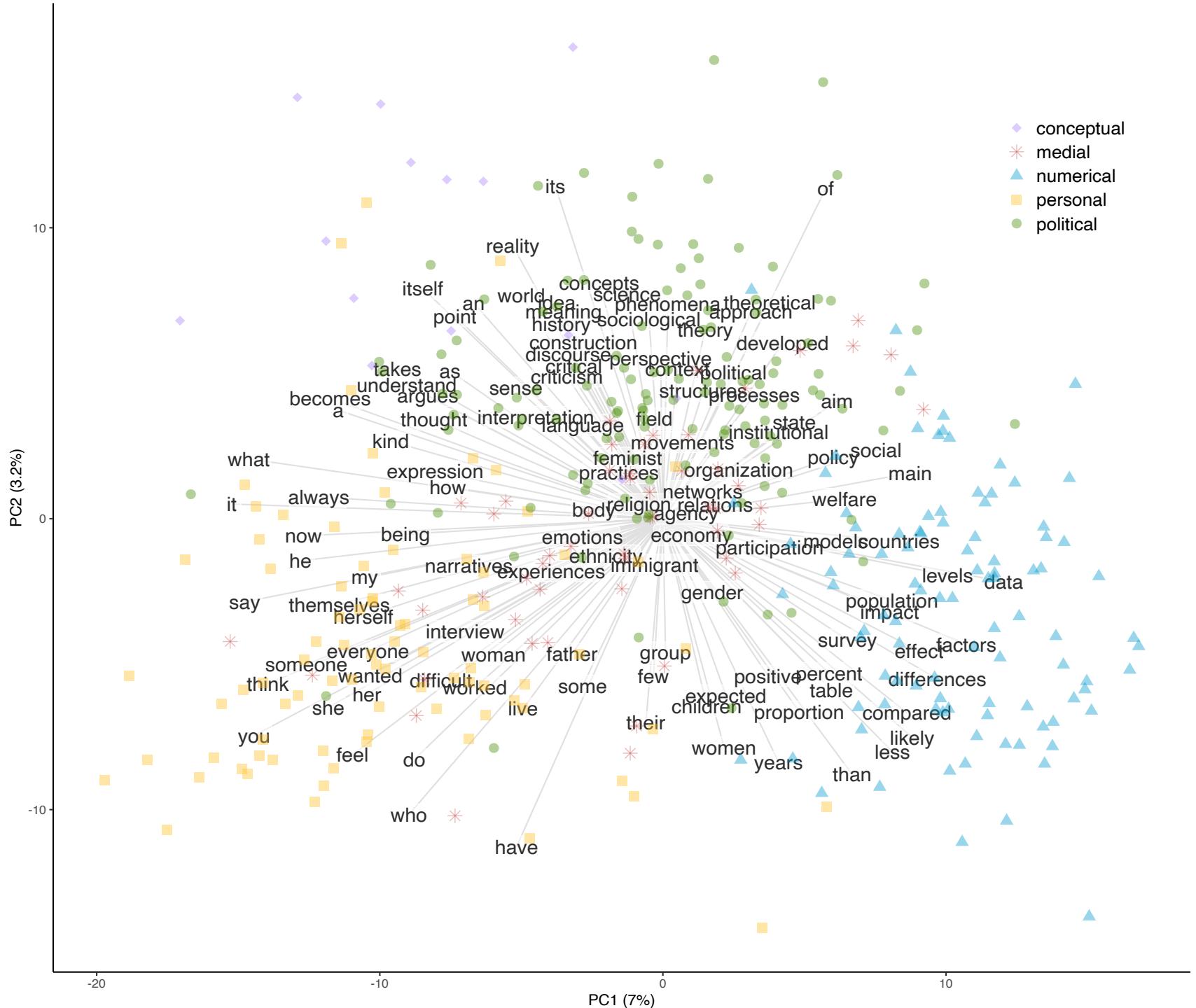
Styles in Swedish 21st century sociology



*The network is constituted of 380 nodes (size=betweenness centrality) and 4292 edges. The average degree of the network (i.e., the mean number of edges per node) is 22.6, with the lowest value being 5 and the highest 126.

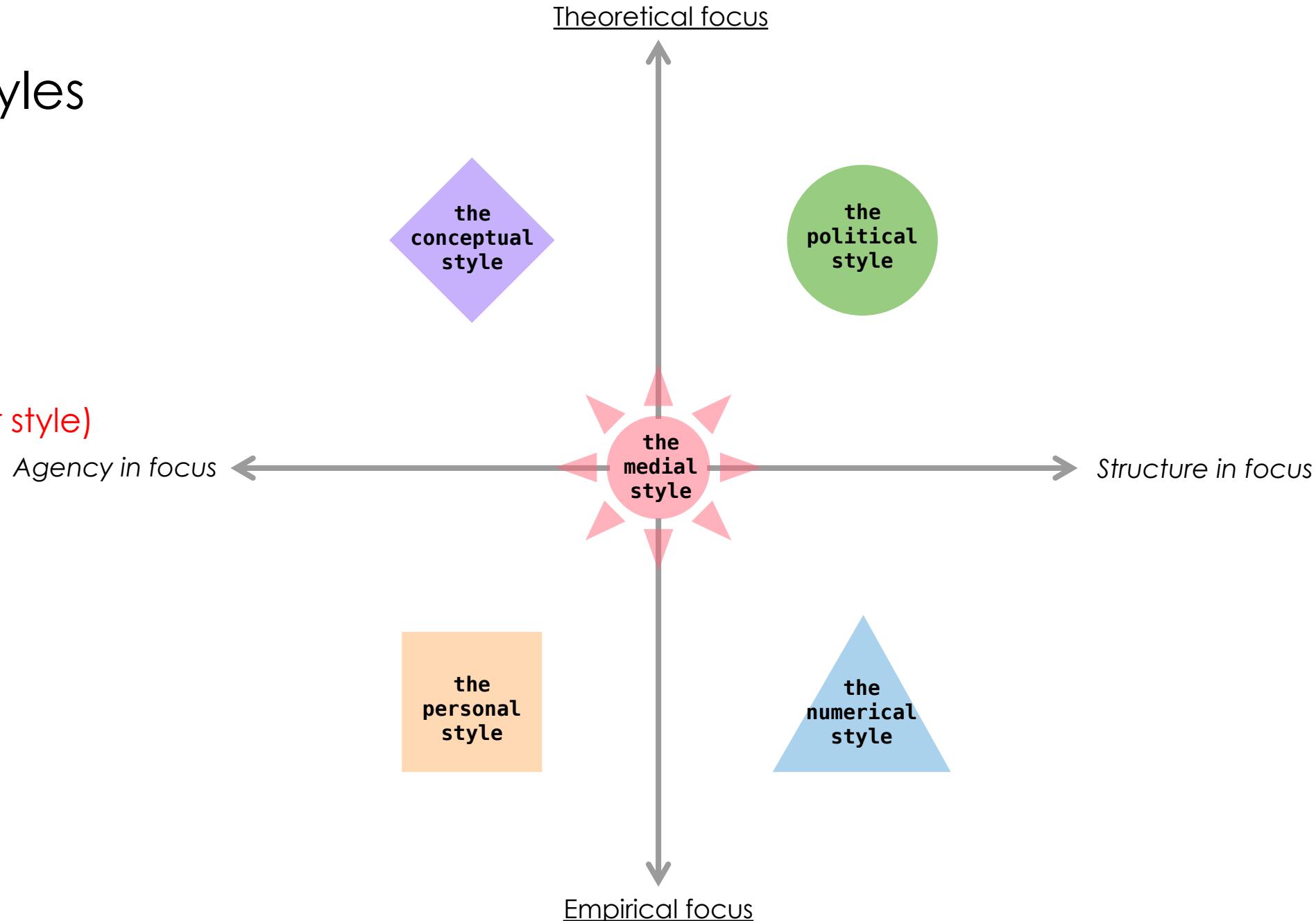
Labelling the styles

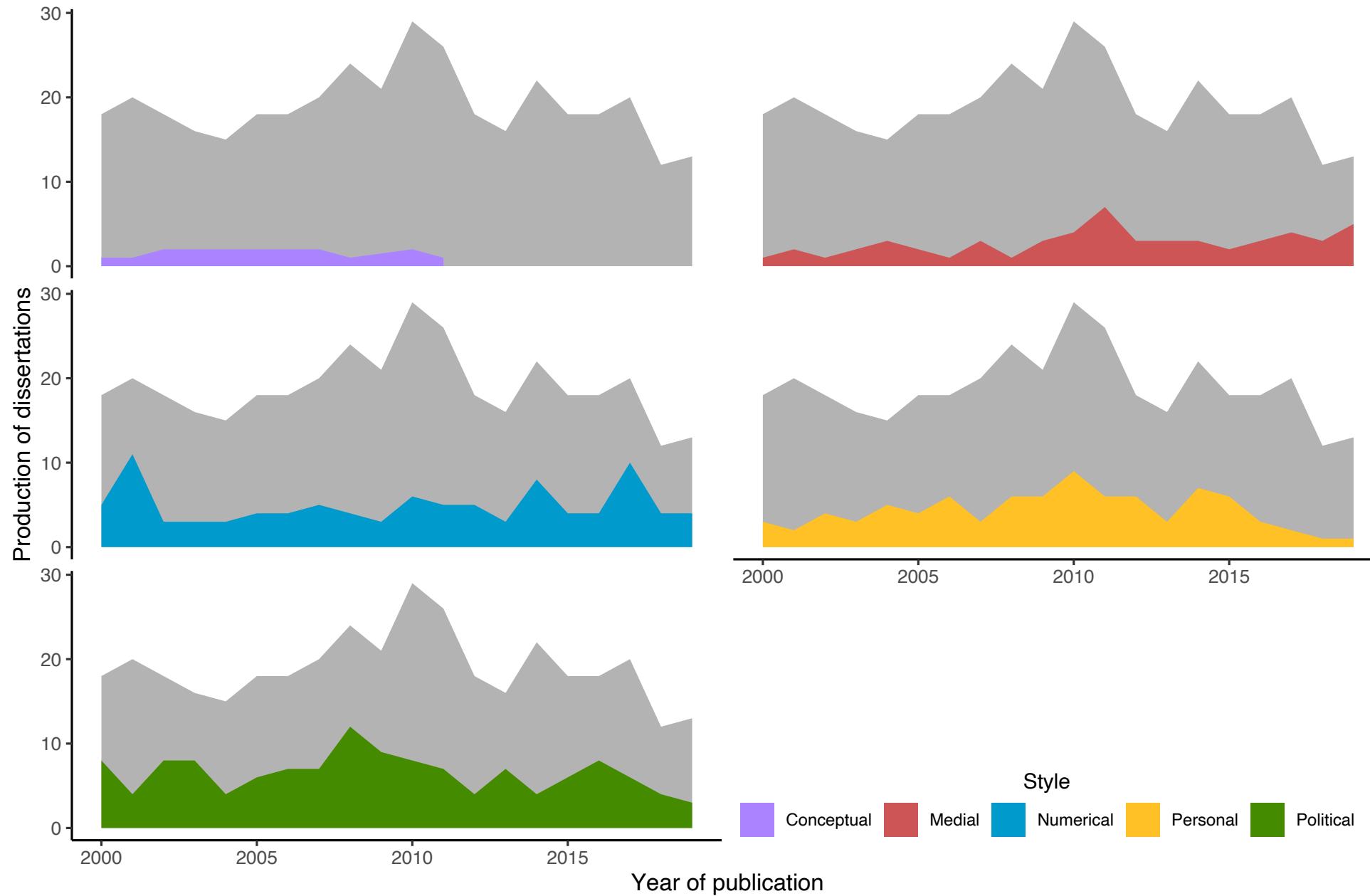
- Burrows's Zeta (MFW)
- PCA (MFW)
- Tf-idf (bigrams)
- Close readings (5 per style)



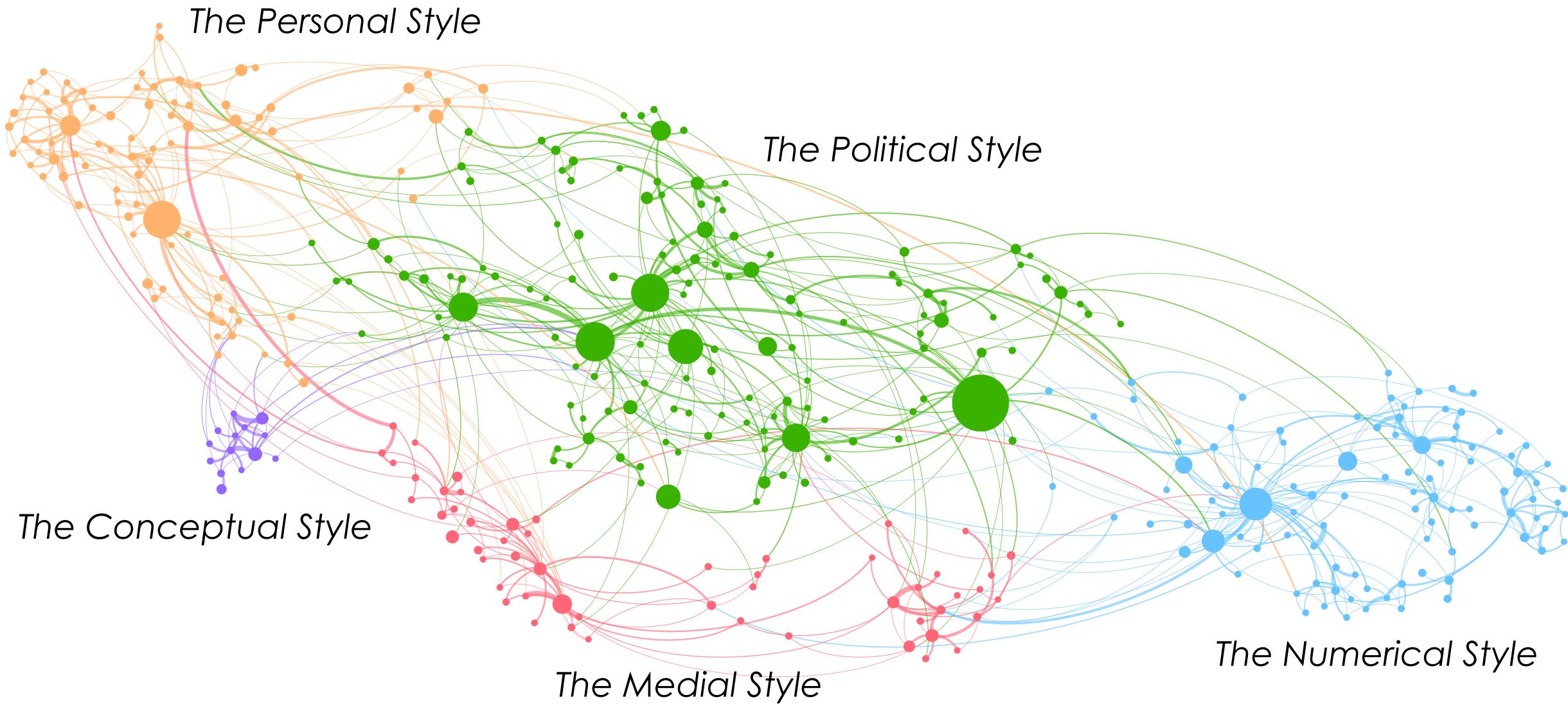
Labelling the styles

- Burrows's Zeta (MFW)
- PCA (MFW)
- Tf-idf (bigrams)
- Close readings (5 per style)





Styles in Swedish 21st century sociology



Word embeddings

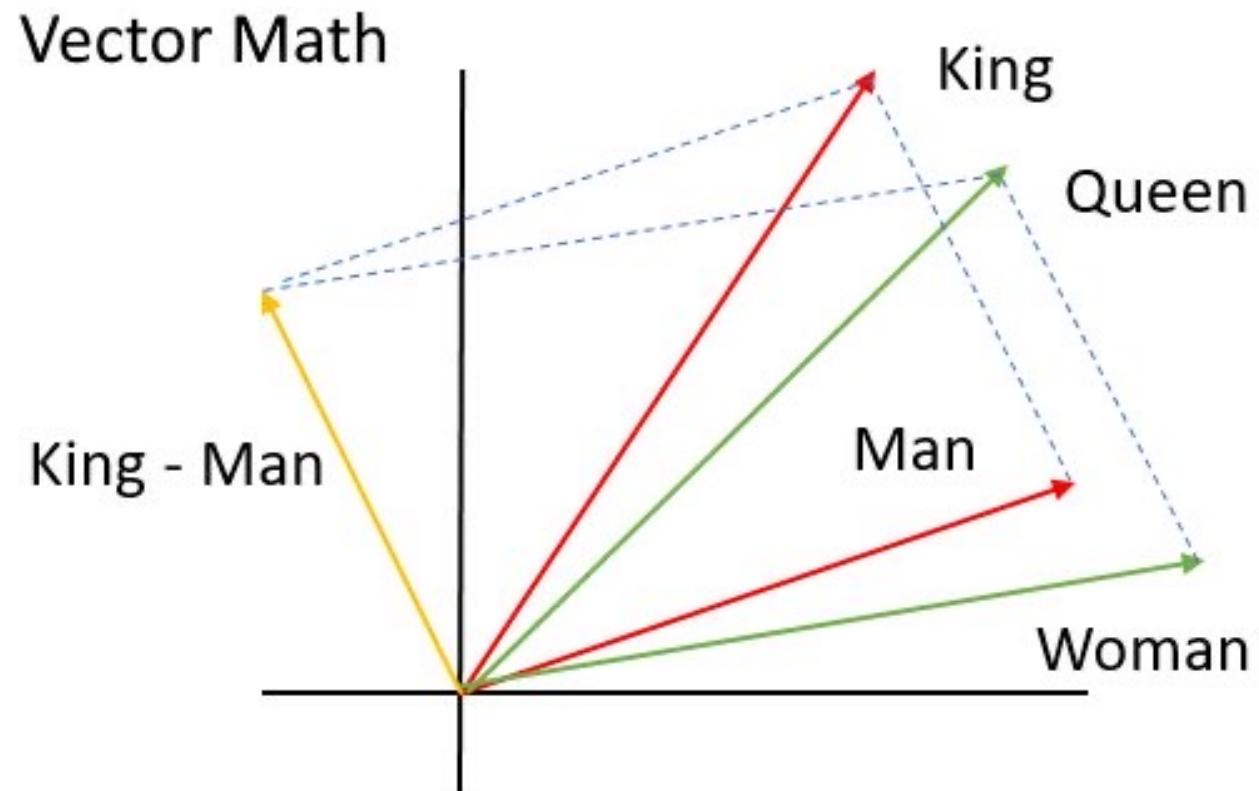
- what is the meaning of sociology?

Word2vec

– representing words with vectors

*“You shall know a word by
the company it keeps!”
(Firth, J. R. 1957:11)*

Word embeddings are a type of word representation in the form vector array, that allows words with similar meaning to have a similar representation.

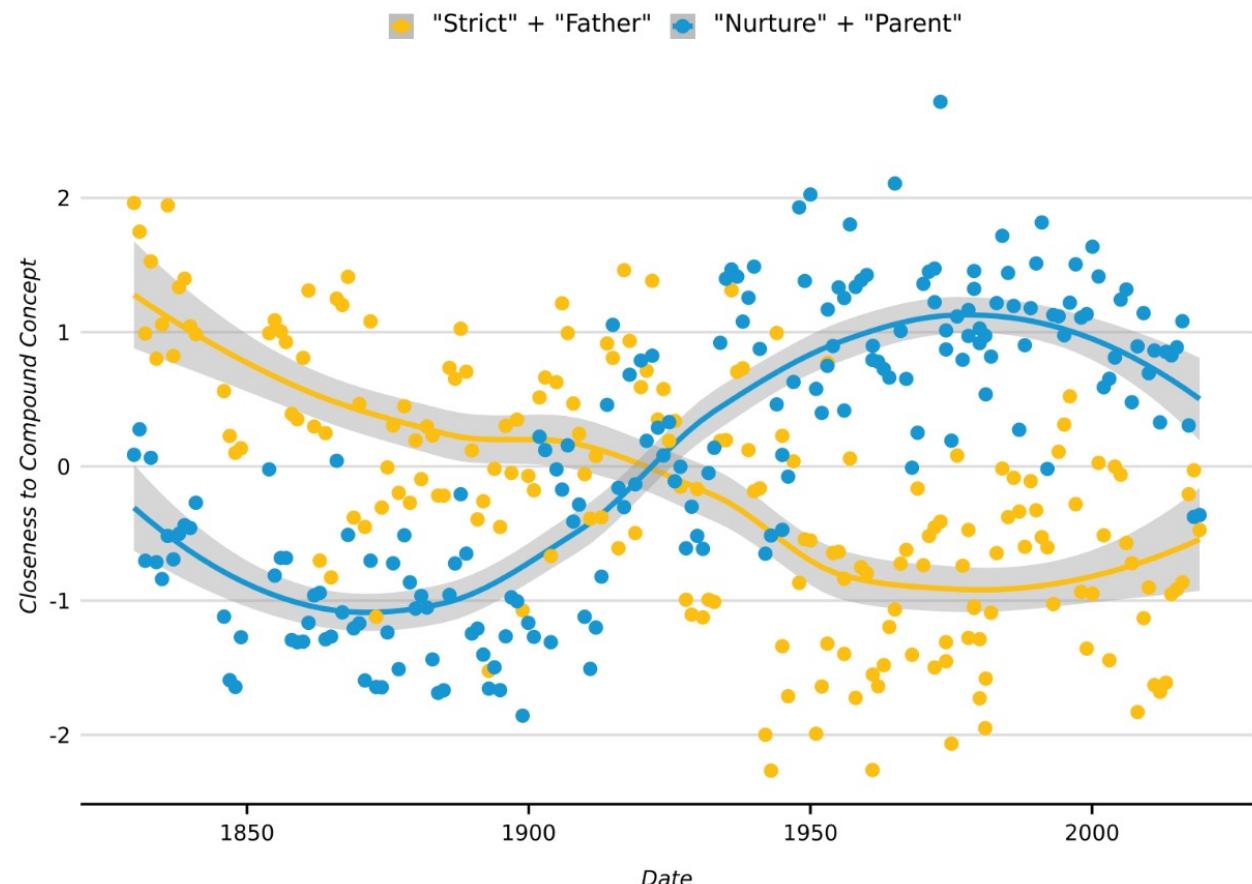


CMD

– Concept Mover's Distance

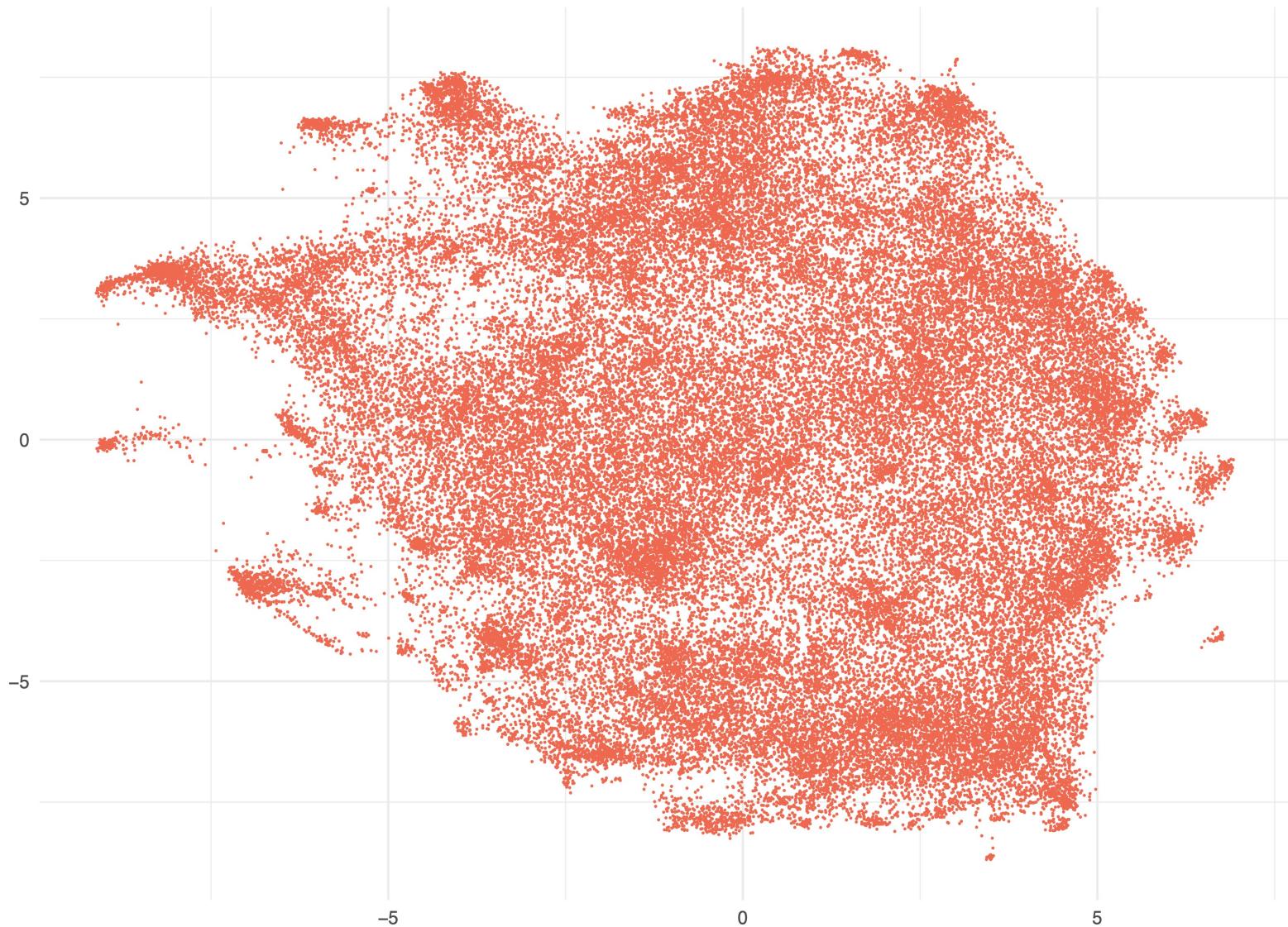
Measuring a text's engagement with a focal concept

"the minimum cost that a document's embedded words need to travel to arrive at the position of all the words in an ideal 'pseudo document' consisting of only words denoting a specified concept" (Stoltz and Taylor 2019, p. 294).



The sociology dissertation vector space

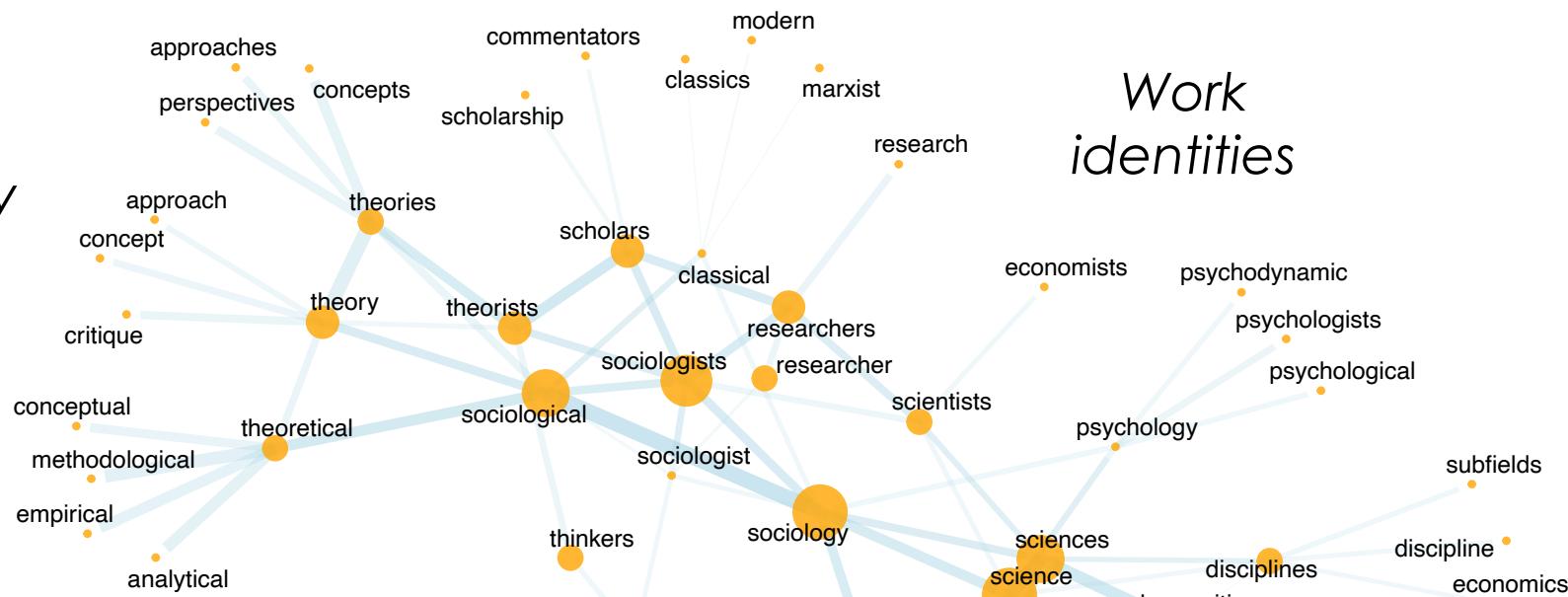
22b total words
174k unique words
66k words occurring 3 times
300 dimensions
20b elements
word-window of ± 8



The meaning of sociology?

5 closest neighbors of “sociology”,
their 5 closest neighbors,
and their 5 closest neighbors

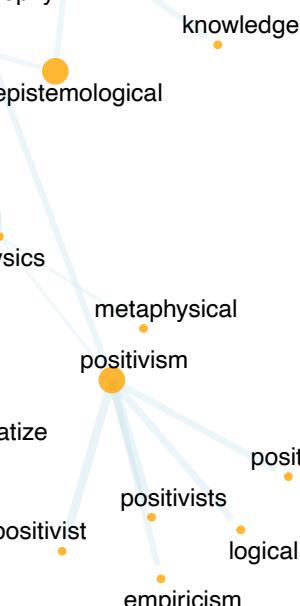
Theory

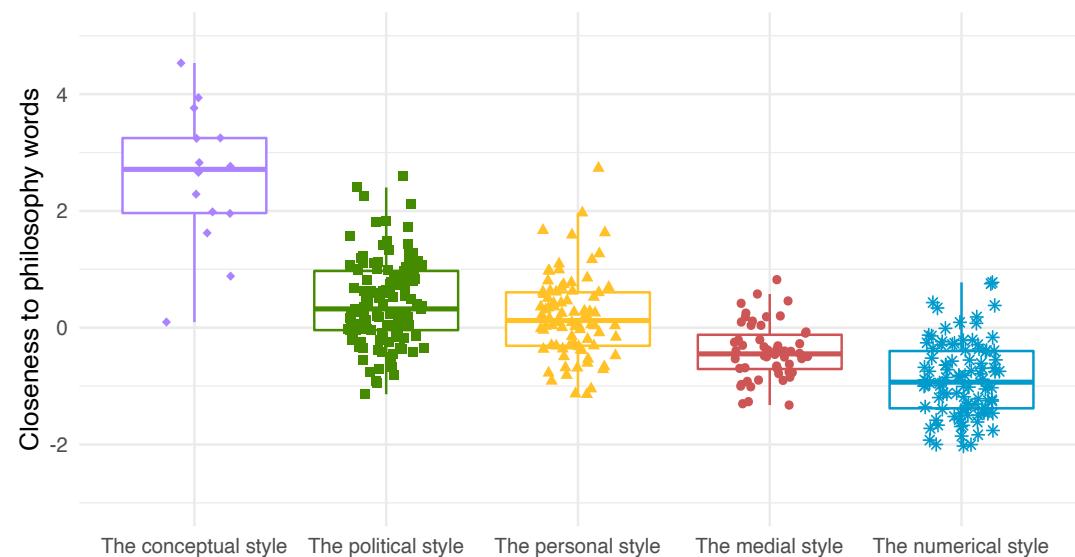
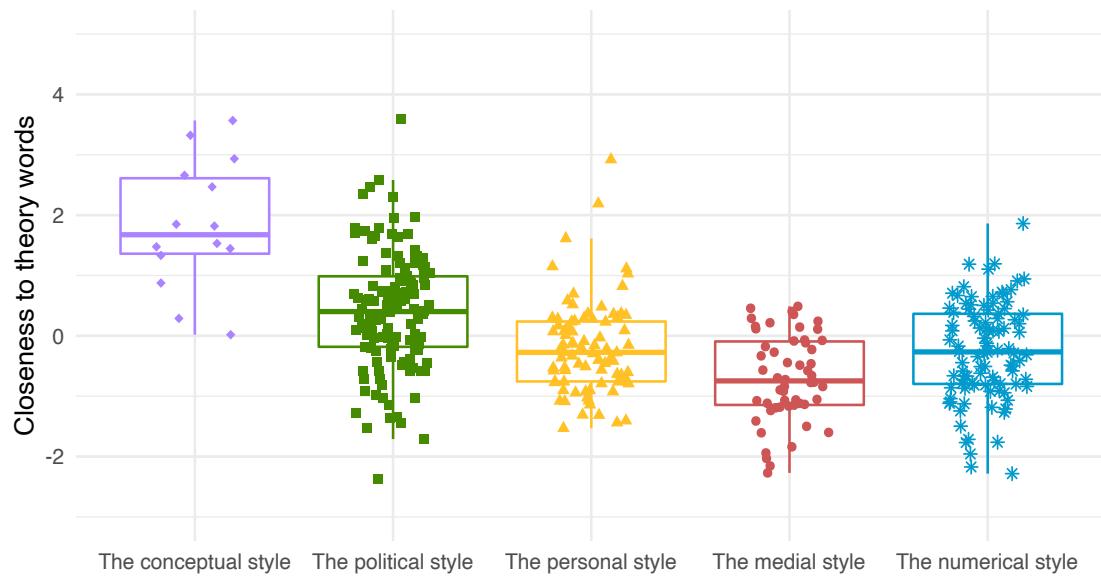
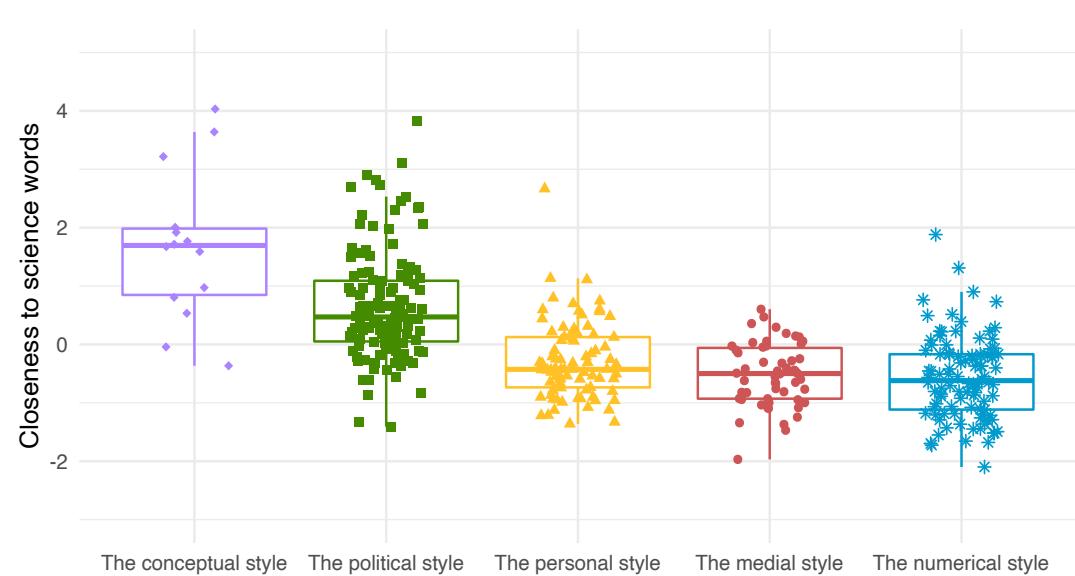
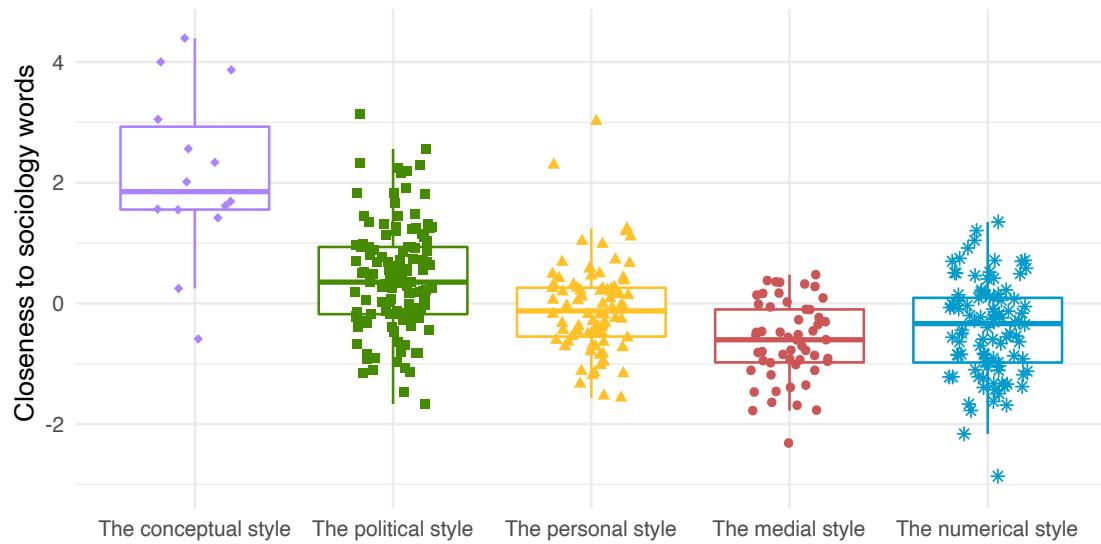


Philosophy

Work
identities

Science



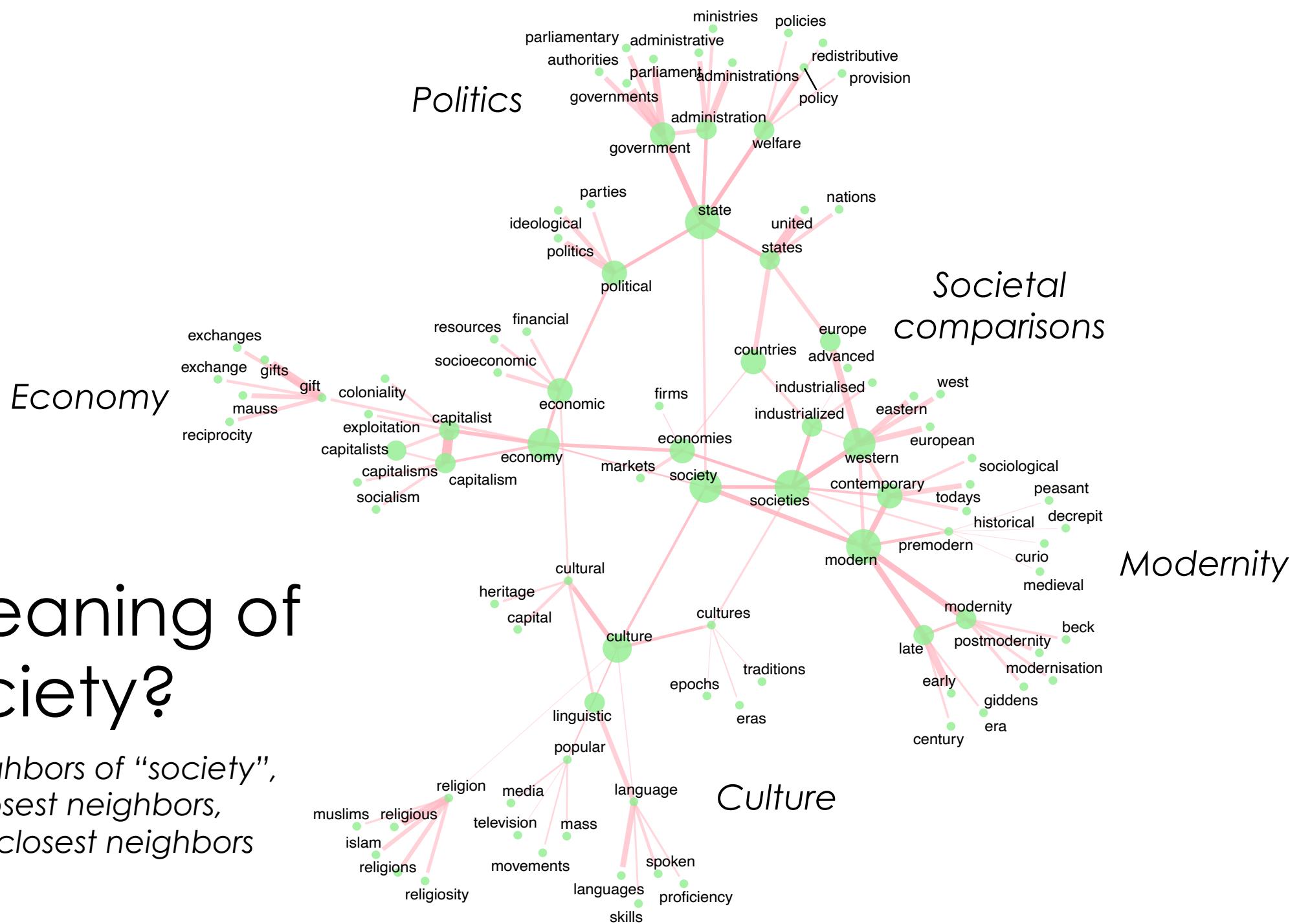


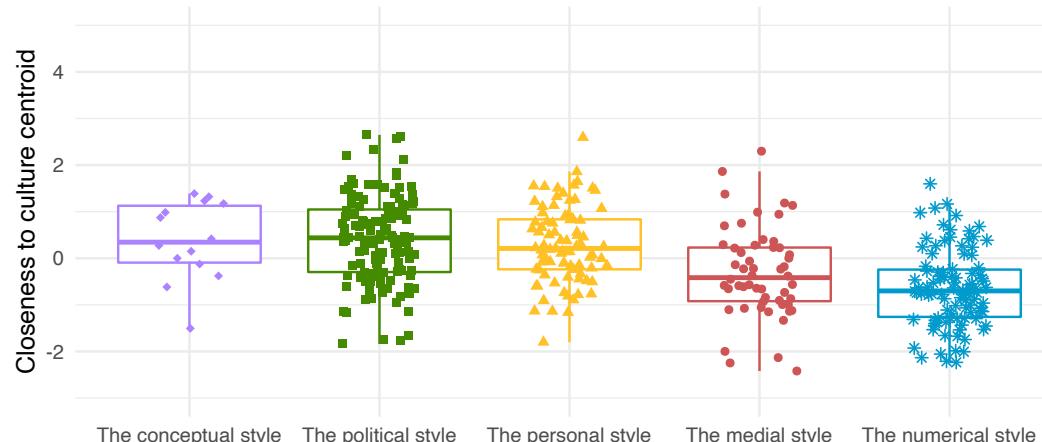
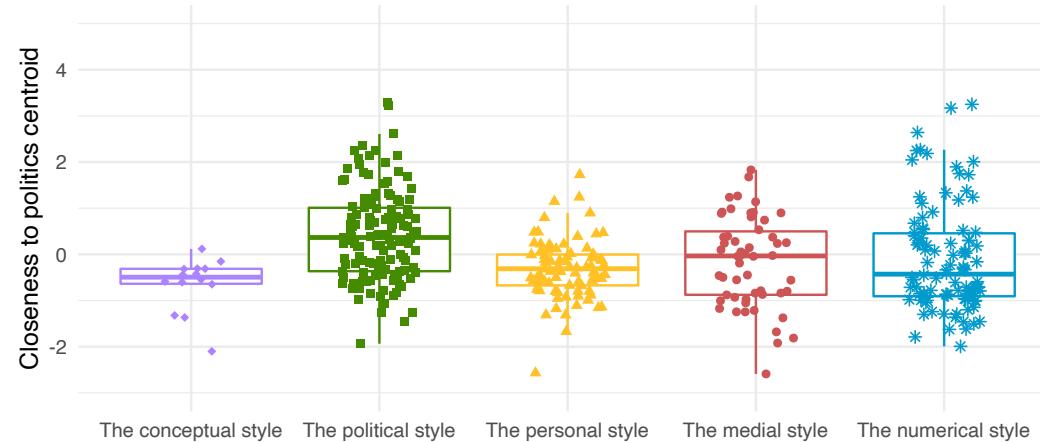
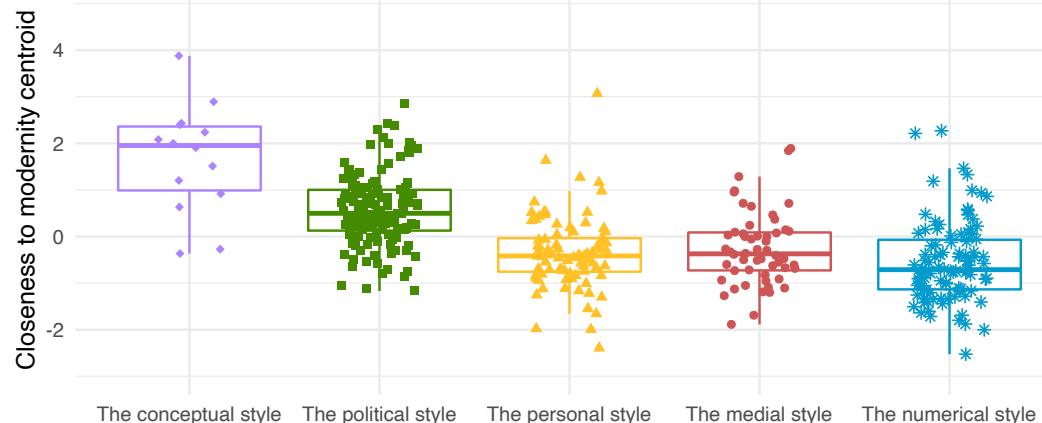
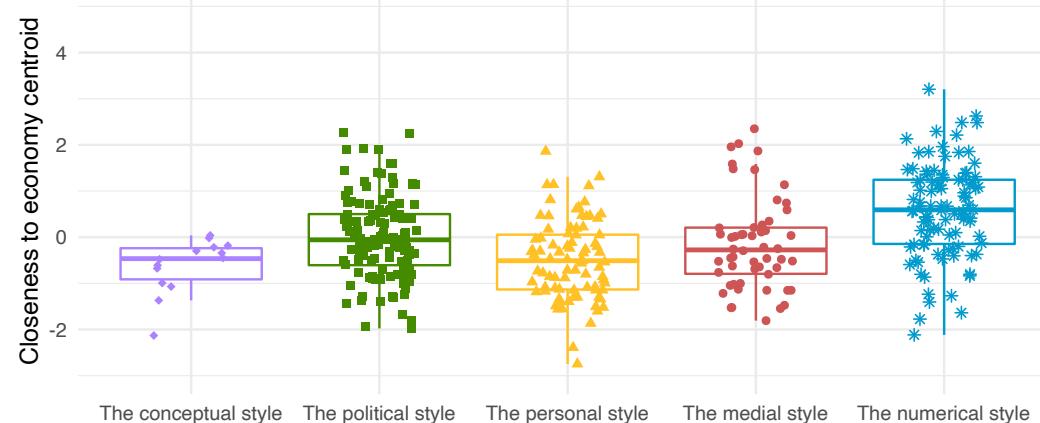
Which styles contribute to the meaning of sociology?

Closeness to centroids (dictionaries)

The meaning of society?

5 closest neighbors of “society”,
their 5 closest neighbors,
and their 5 closest neighbors





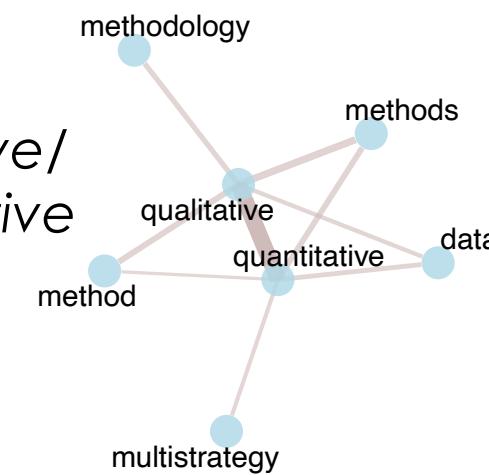
Which styles contribute to the meaning of society?

Closeness to centroids (dictionaries)

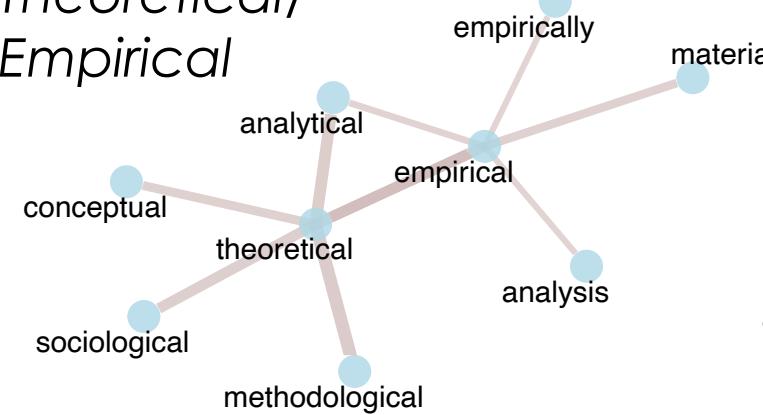
Dichotomies in sociology

5 closest neighbors of famous dichotomies

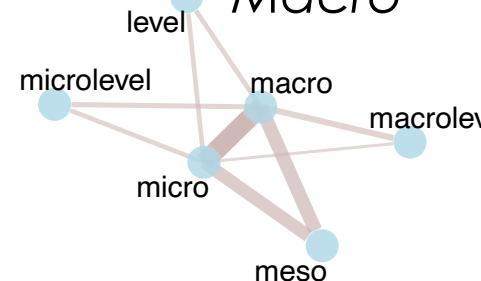
Qualitative/
Quantitative



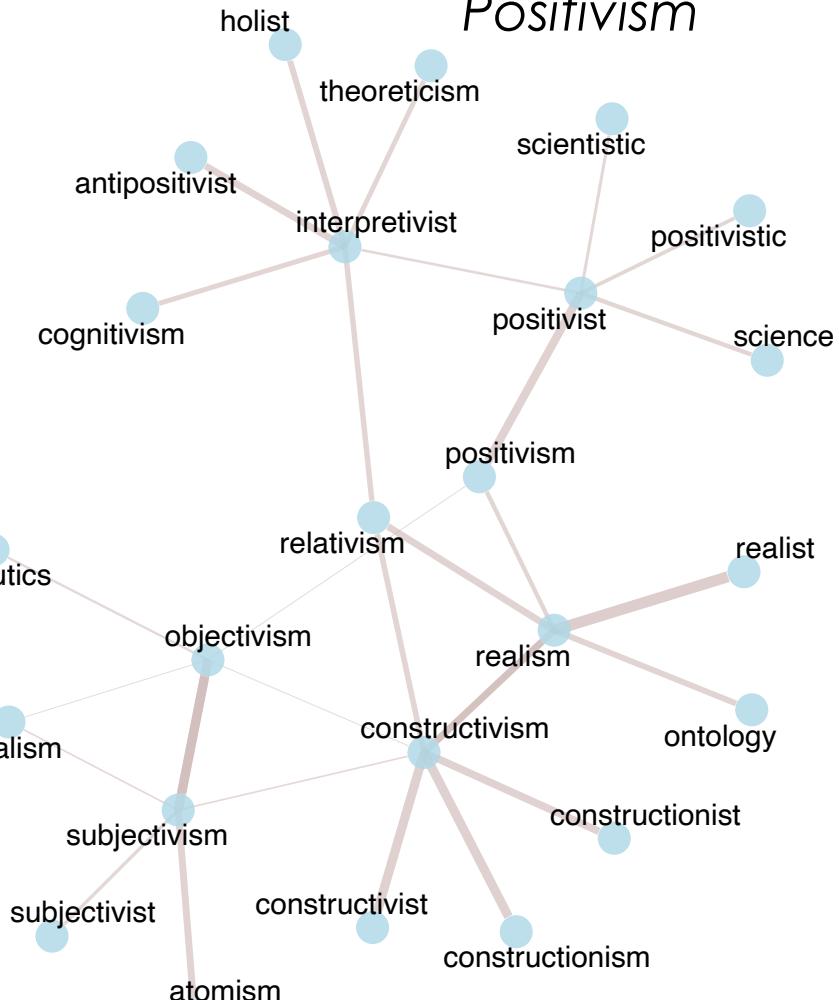
Theoretical/
Empirical

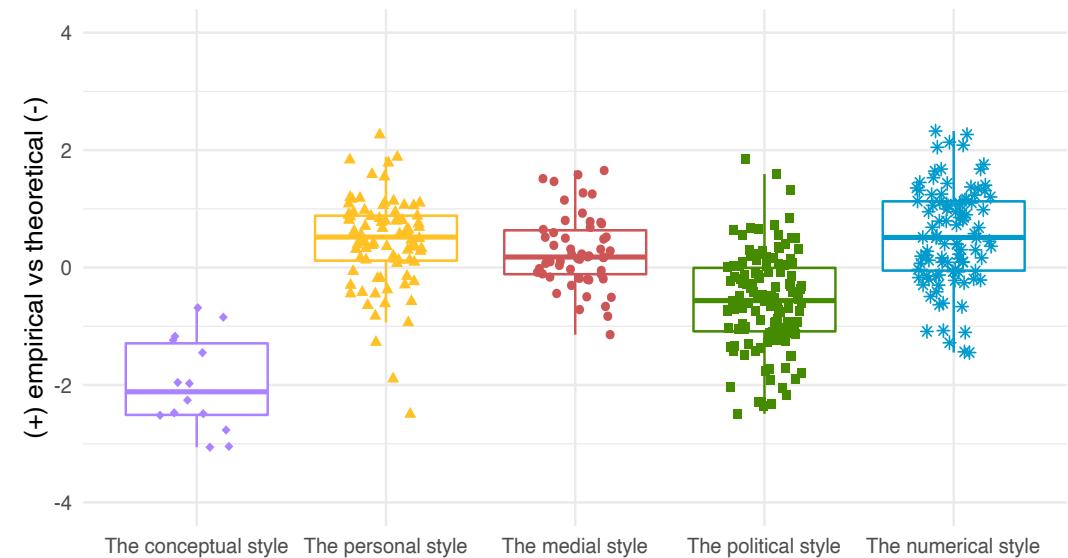
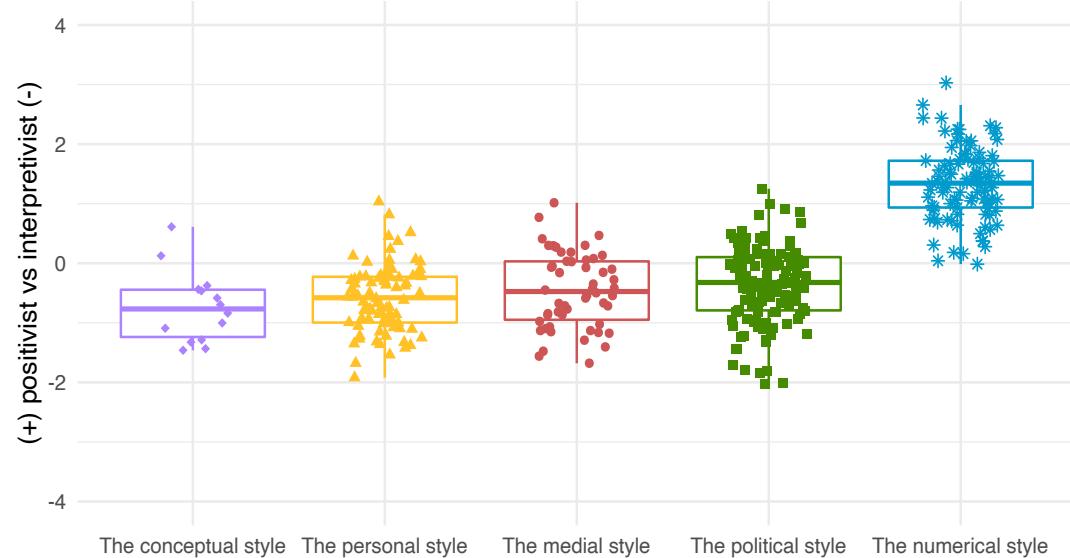
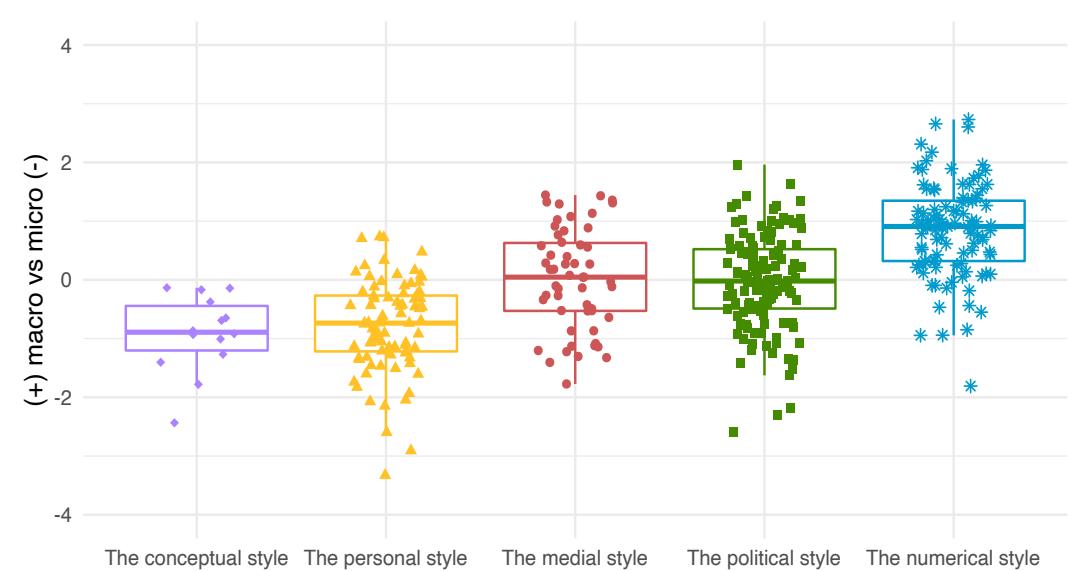
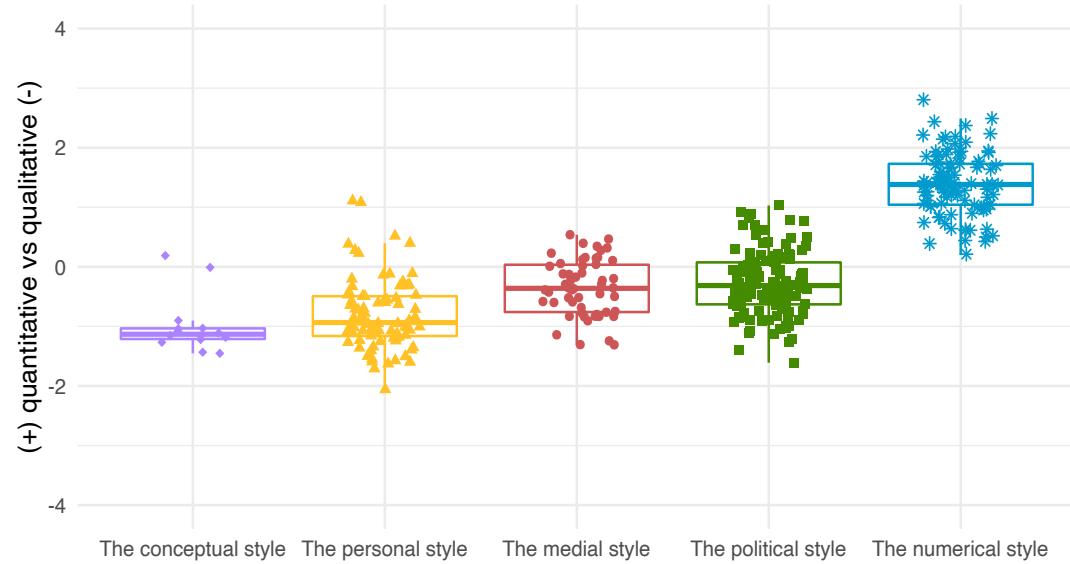


Micro/
Macro



Interpretivism/
Positivism





Dichotomies in sociology

Closeness to manipulated poles (dictionaries)

Some suggested conclusions

- Sociological knowledge does not appear to be overly fragmented
(styles, topics, social phenomenon etc. reoccur over time 20-40 years)
- For full-texts, the mainstream does not follow a ‘master dichotomy’
(the abstract is not a summary, but a genre of positioning research)
- Sociology and its problem is a non-issue for most dissertations
(‘the lore’ is primarily addressed in more conceptual/philosophical works)

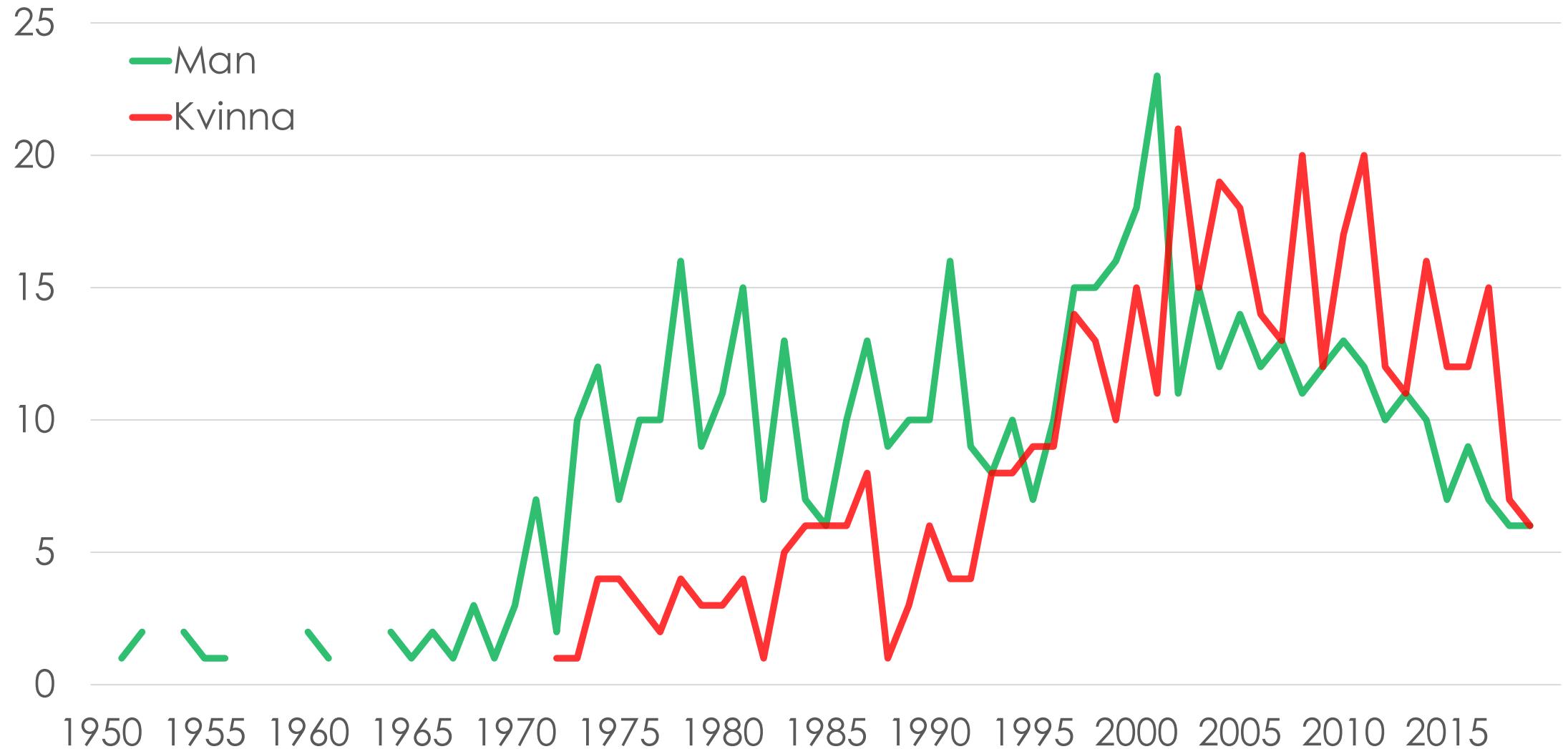
Thank you
for you attention!

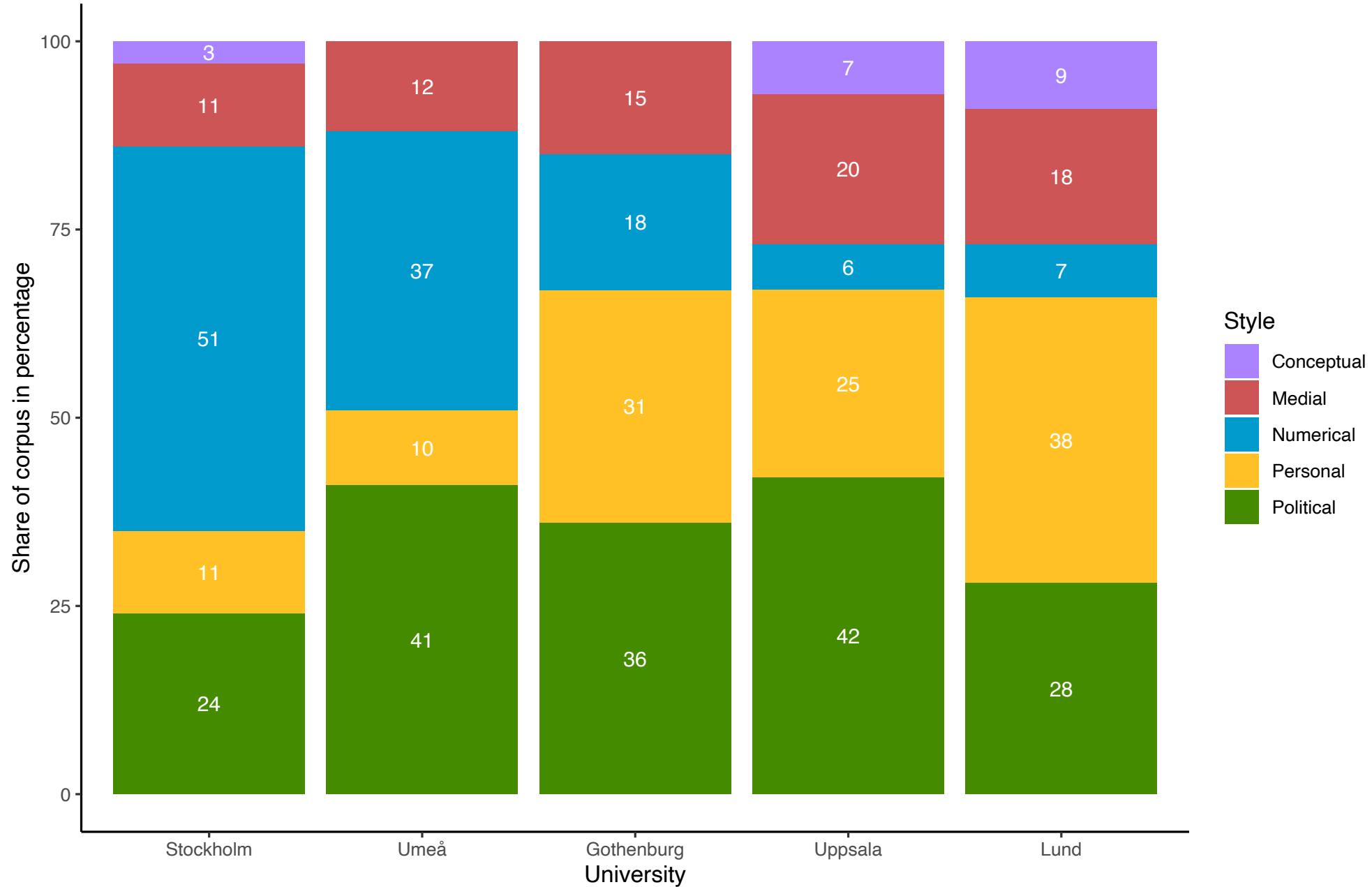
Descriptive stats on the full-text corpus

Variable	Category	n	%
Gender	Man	168	44,1%
	Woman	213	55,9%
University	Gothenburg	49	12,9%
	Lund	67	17,6%
	Stockholm	112	29,4%
	Umeå	63	16,5%
	Uppsala	60	15,7%
	Other	30	7,9%
Language	English	215	56,4%
	Swedish	166	43,6%
Type	Compilation	148	38,8%
	Monograph	233	61,2%

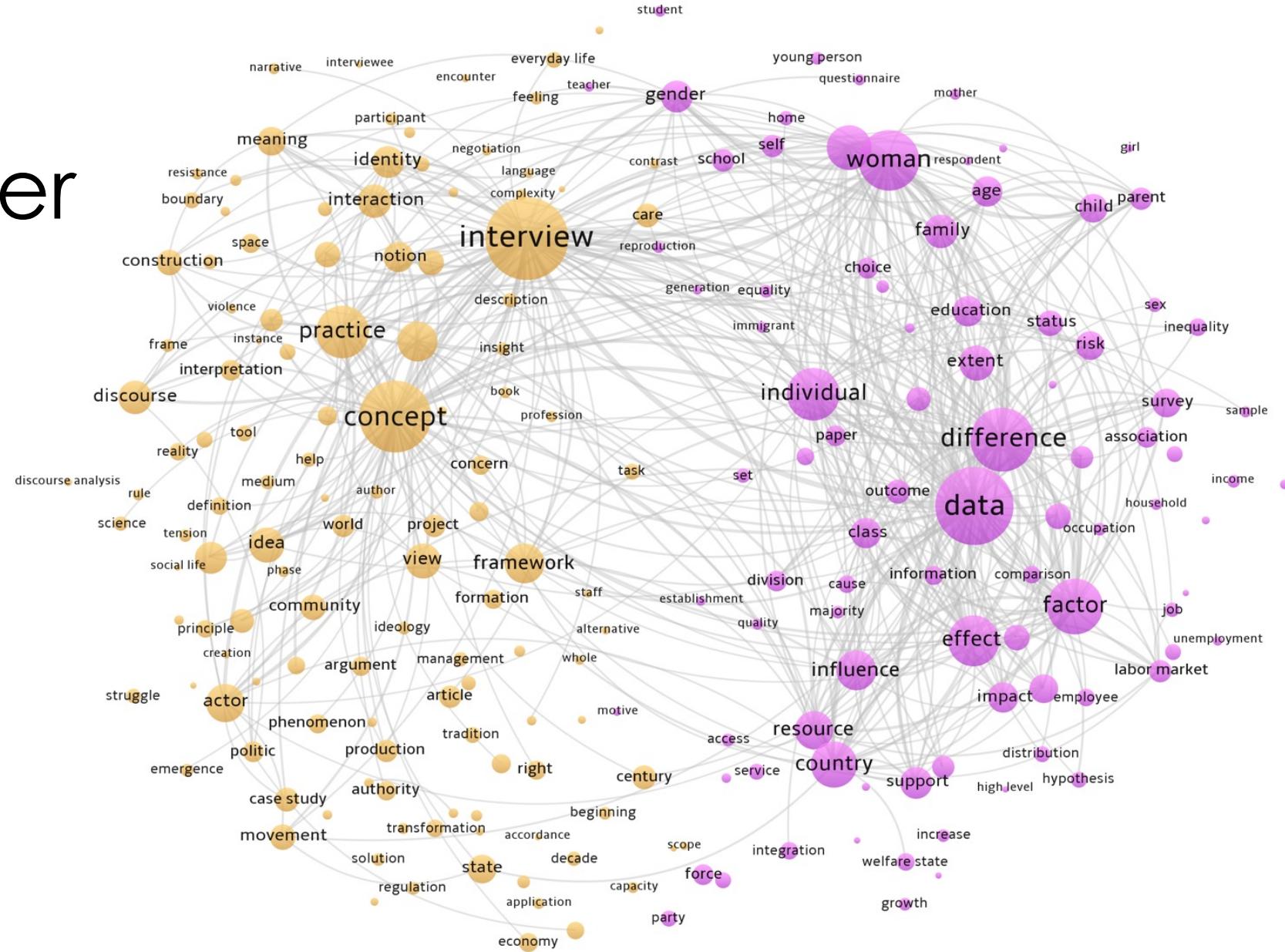
Variable	Category	n	%
Year	2000	18	4,7%
	2001	20	5,3%
	2002	19	5,0%
	2003	16	4,2%
	2004	15	3,9%
	2005	18	4,7%
	2006	18	4,7%
	2007	20	5,3%
	2008	24	6,3%
	2009	21	5,5%
Year	2010	29	7,6%
	2011	26	6,8%
	2012	18	4,7%
	2013	16	4,2%
	2014	22	5,8%
	2015	18	4,7%
	2016	18	4,7%
	2017	20	5,3%
	2018	12	3,2%
	2019	13	3,4%

Avhandlingarnas genus



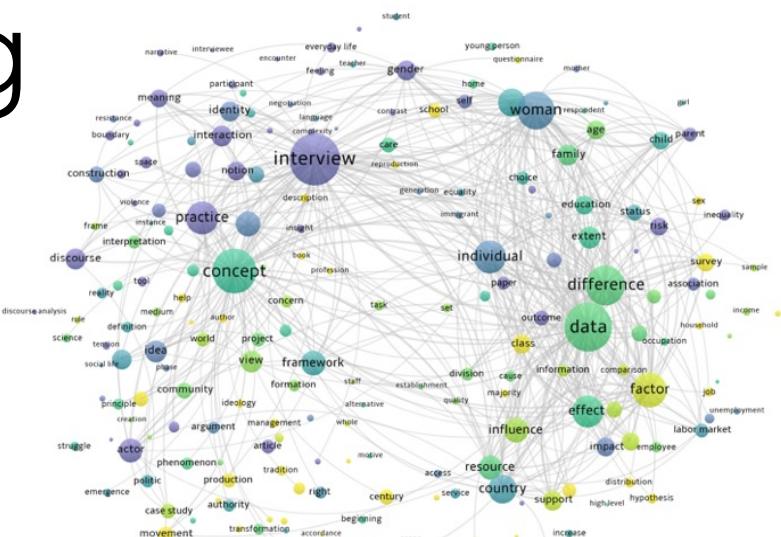


Kluster av ord- korrelationer

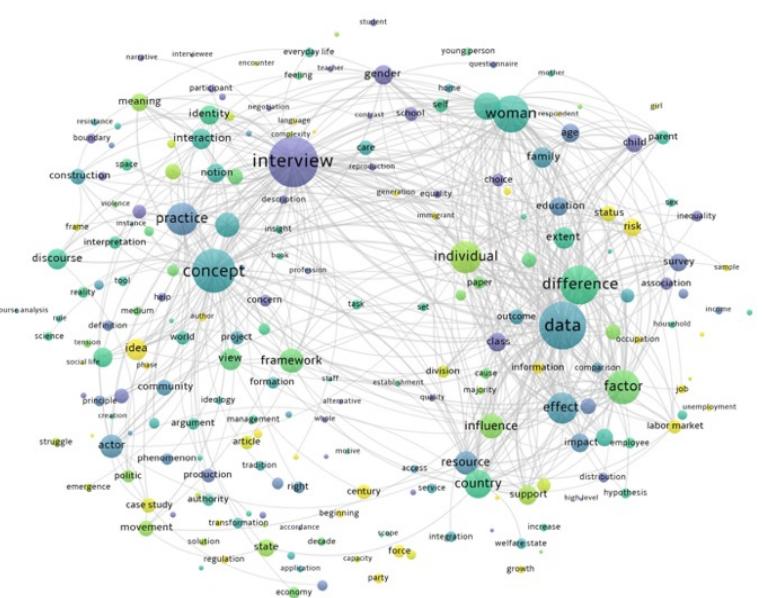


Fördelning över tid

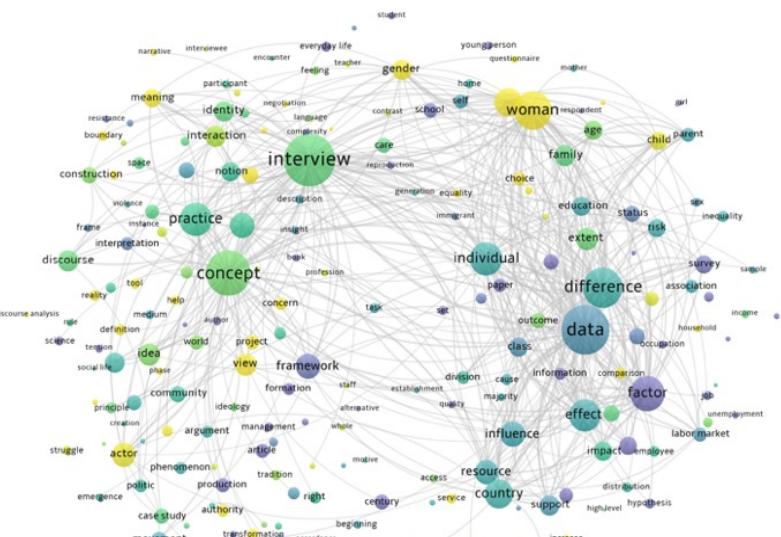
1980-1995



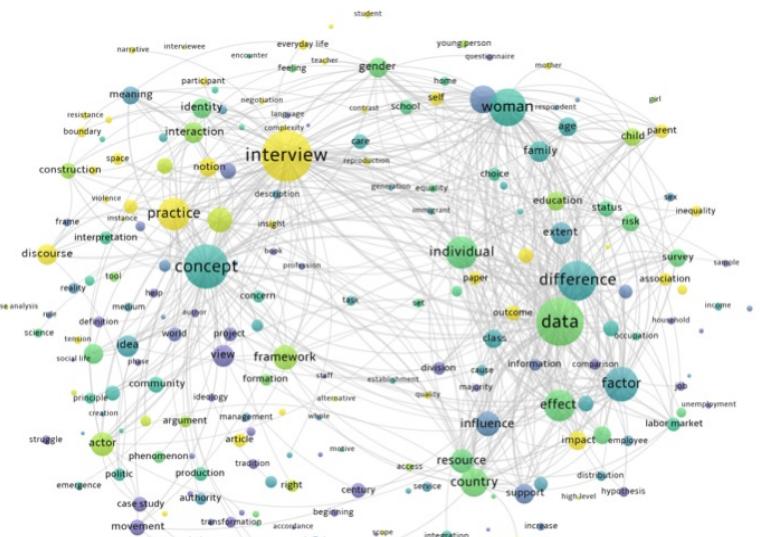
1996-2002



2003-2009



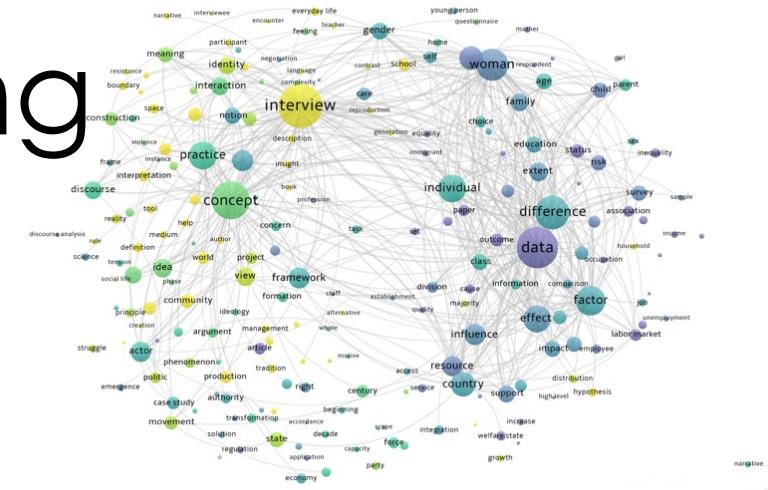
2010-2019



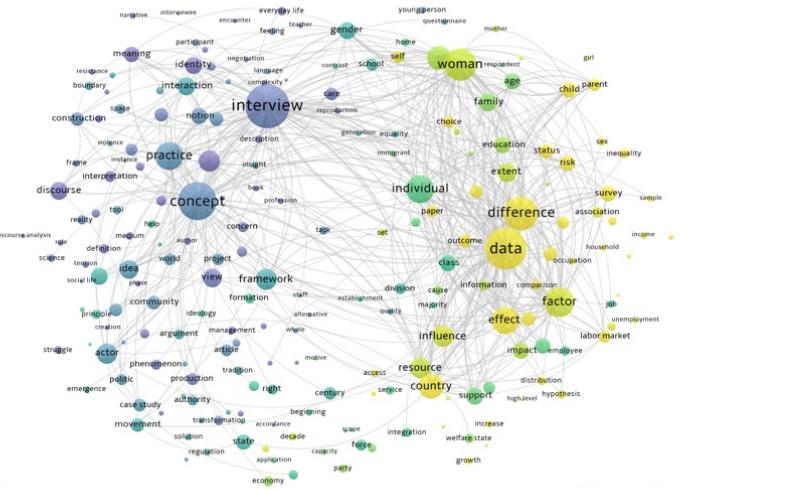
-1.0 -0.5 0.0 0.5 1.0

Fördelning över rum

Uppsala University



Gothenburg University

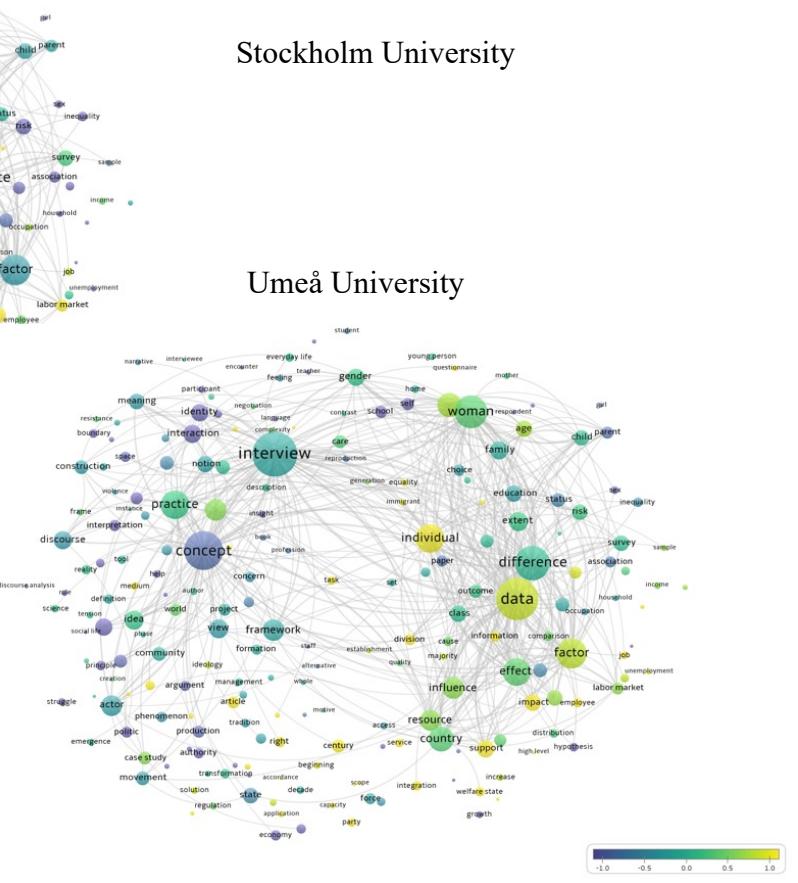


Stockholm University



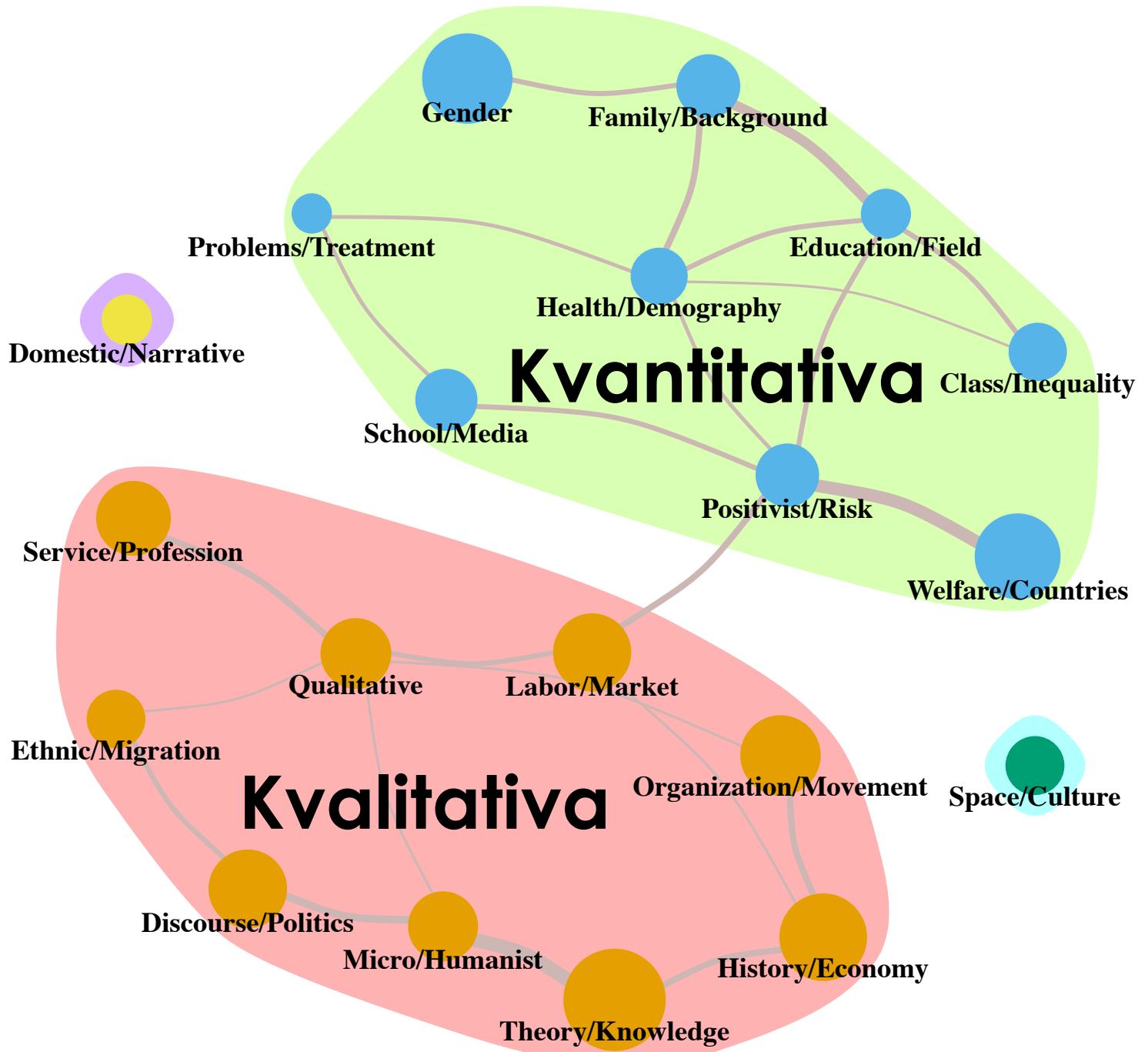
Lund University

Umeå University



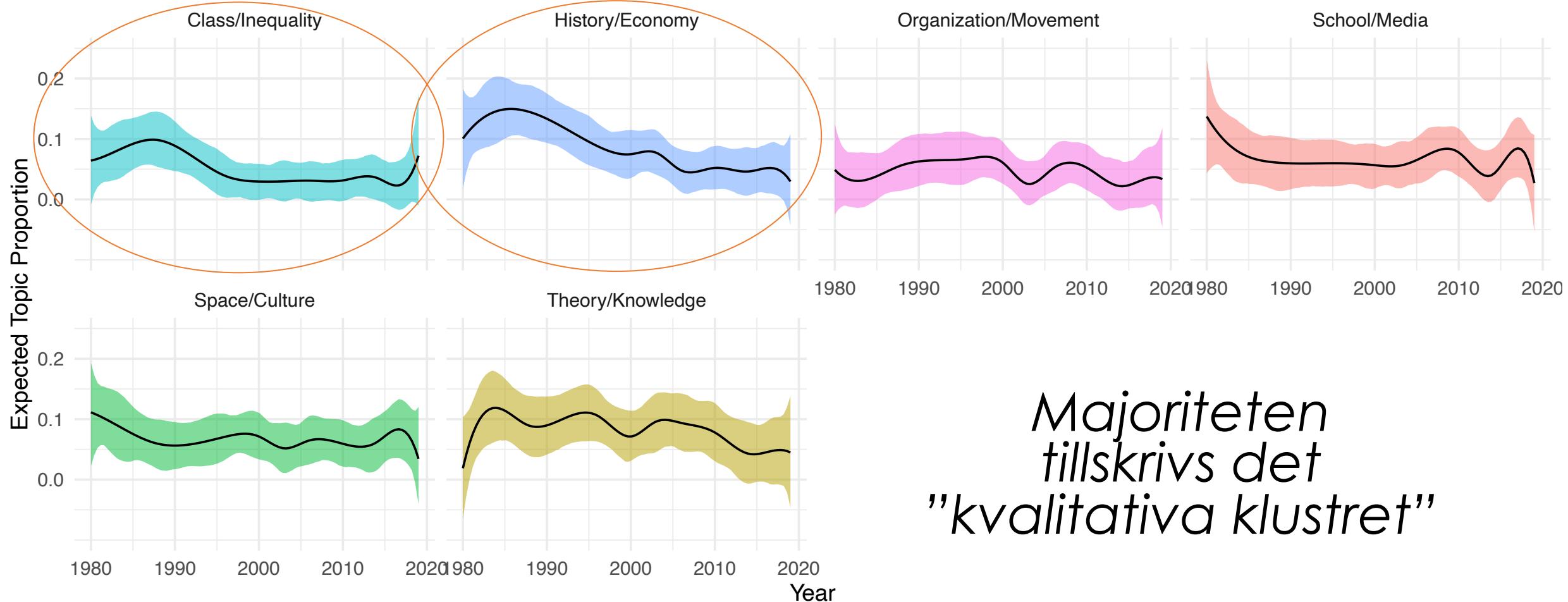
-1.0 -0.5 0.0 0.5 1.0

Tematiska kluster



Tematik över tid

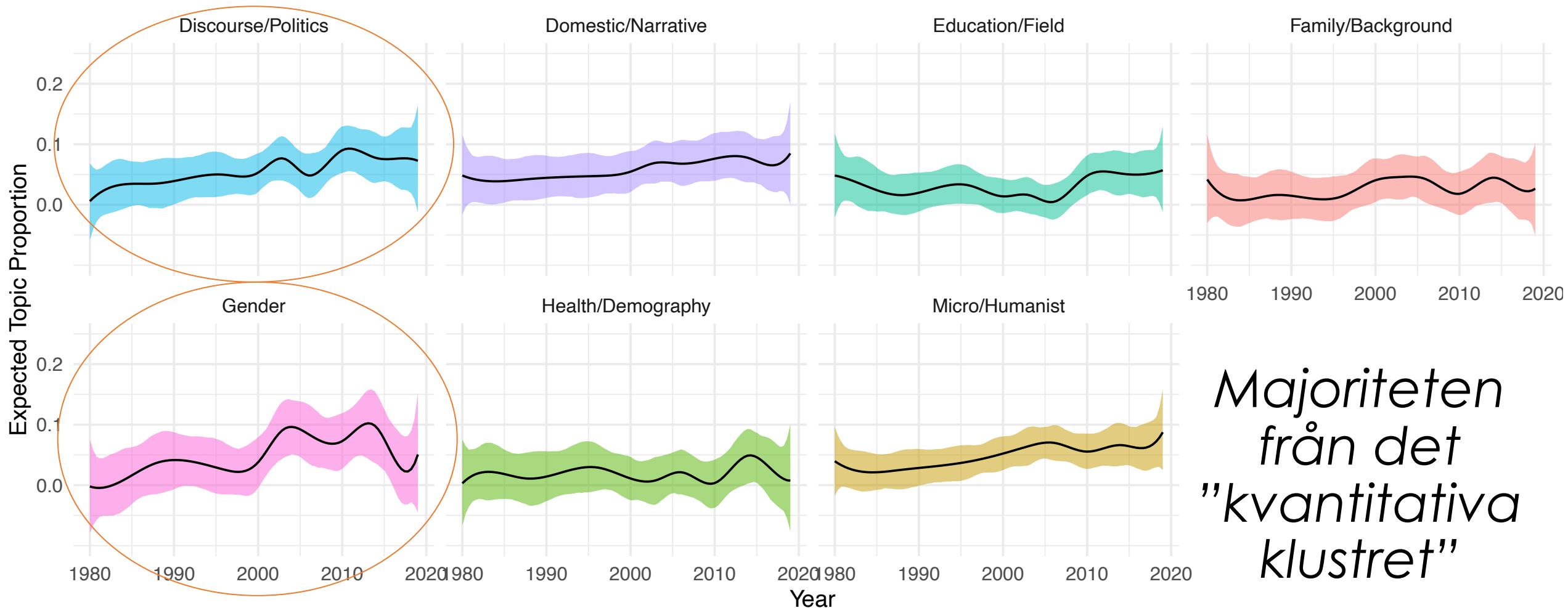
Topics leaning towards the 20th century



Majoriteten
tillskrivs det
"kvalitativa klustret"

Tematik över tid

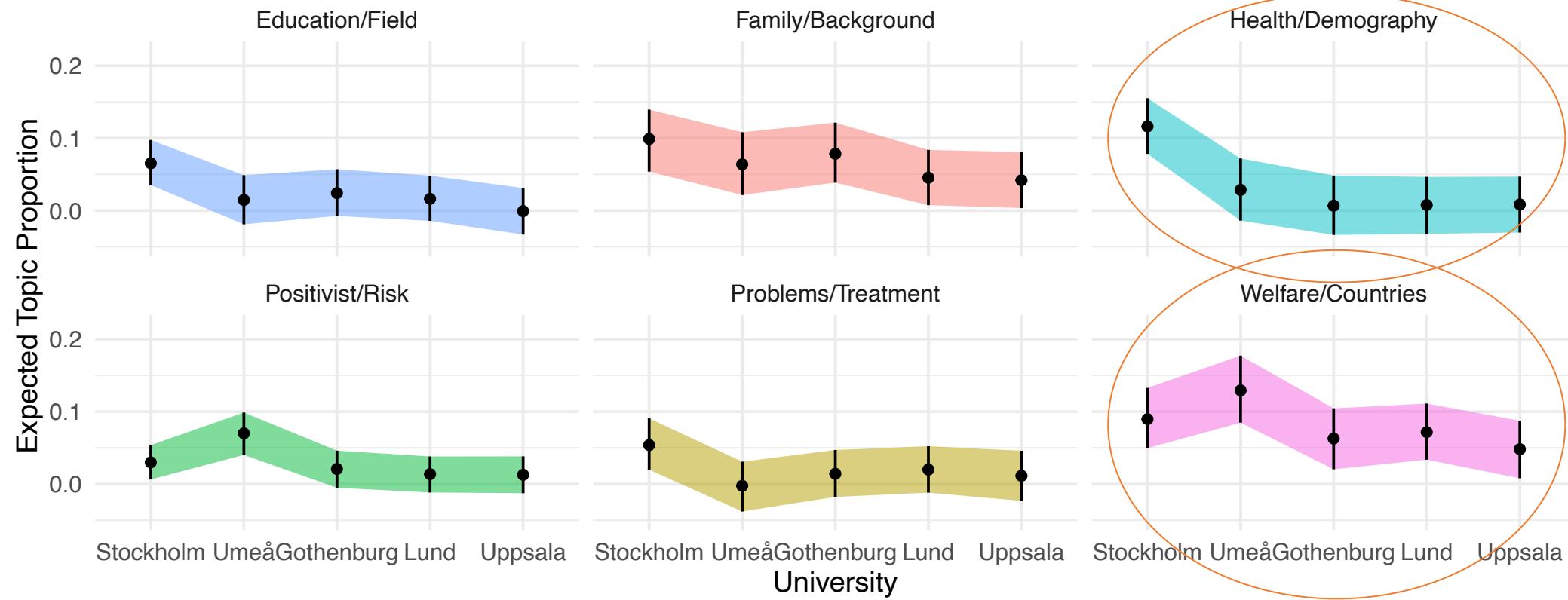
Topics leaning towards the 21st century



Majoriteten
från det
"kvantitativa
klustret"

Tematik över rum

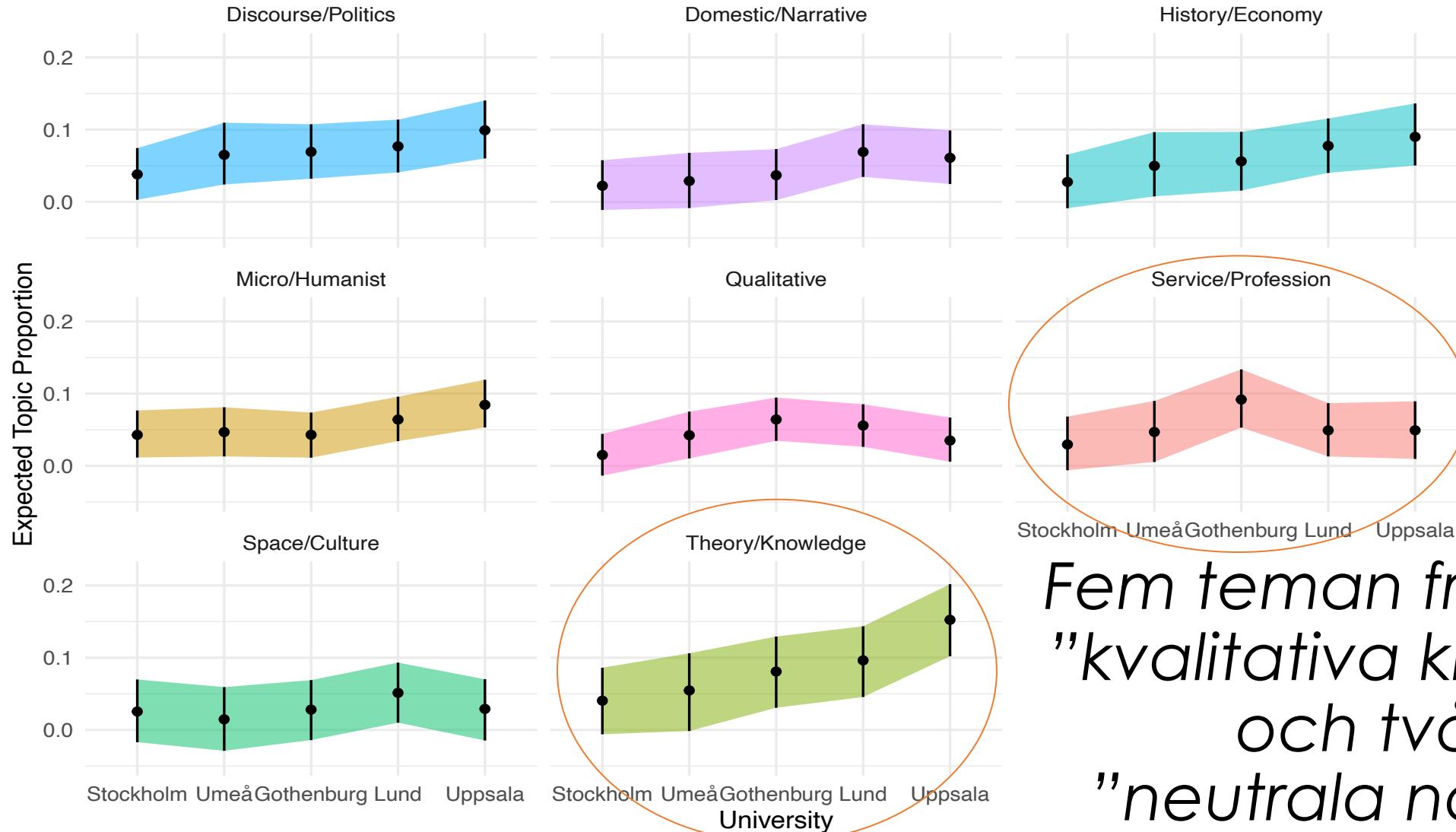
Topics where Stockholm or Umeå has the strongest spatial effect



Alla teman är del av det "kvantitativa klustret"

Tematik över rum

Topics where Lund, Uppsala or Gothenburg has the strongest spatial effect



Fem teman från det
"kvalitativa klustret"
och två
"neutrala noder"