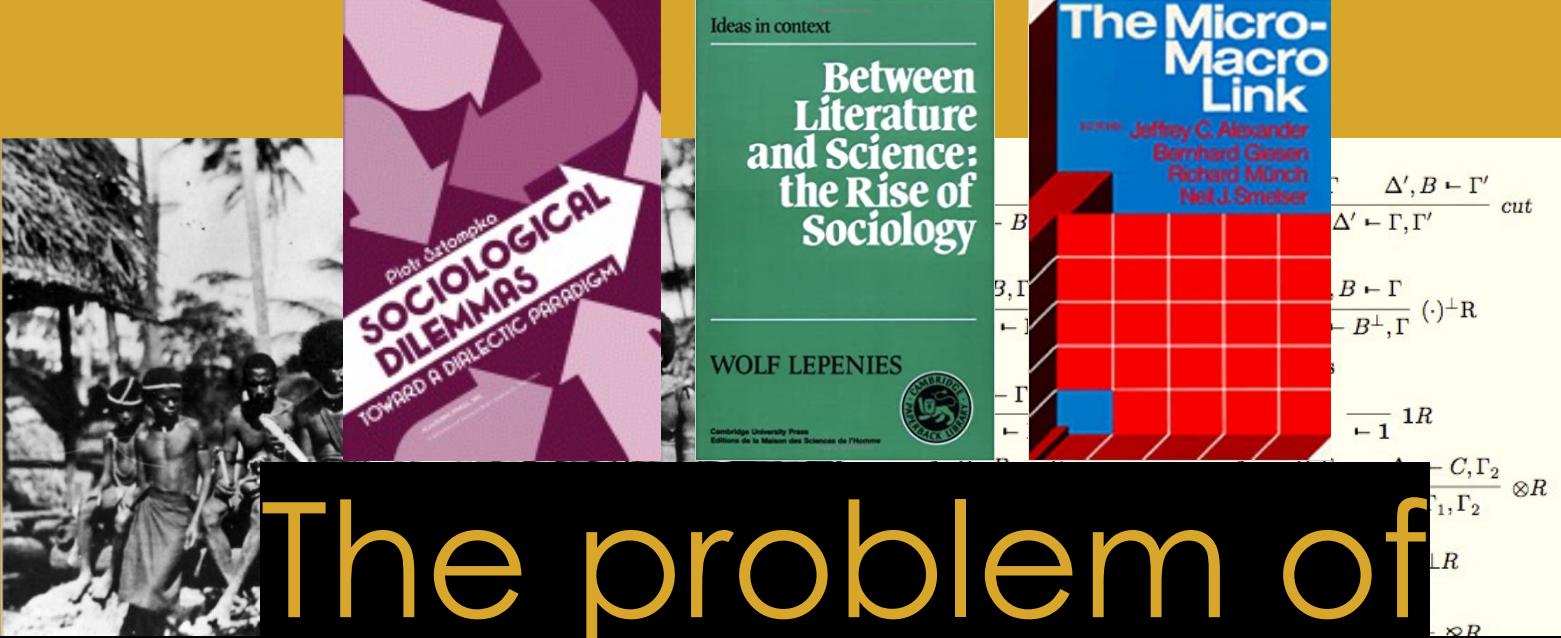


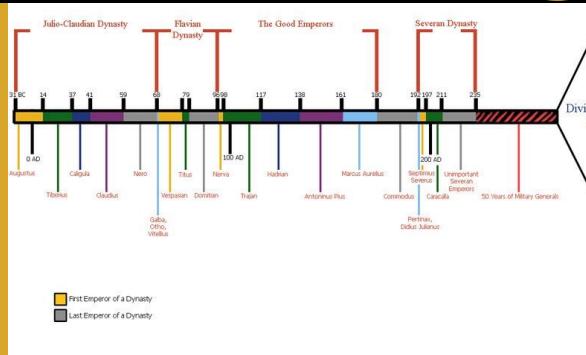
Drawing the tree of sociology in Sweden with computational text analysis

Josef Gimmerskov, doktorand
Sociologiska institutionen
Uppsala universitet

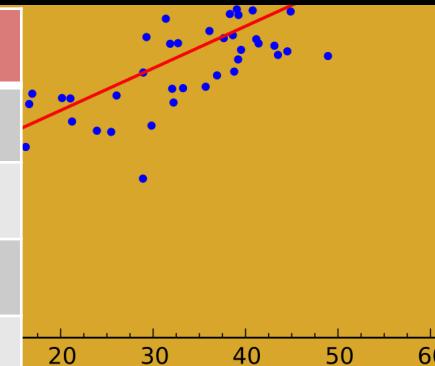
Sociology in crisis!



The problem of sociology's "great divide"

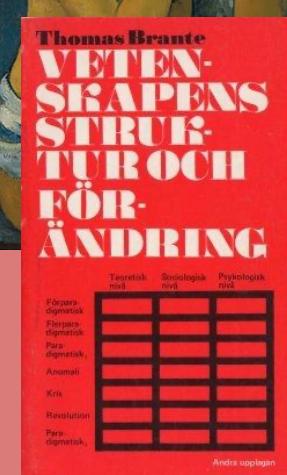
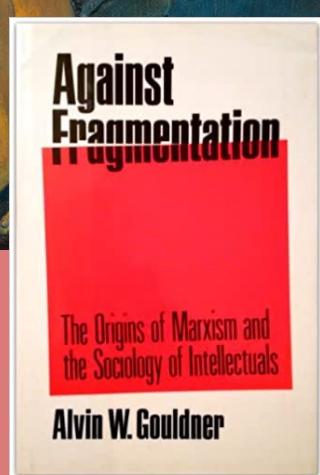
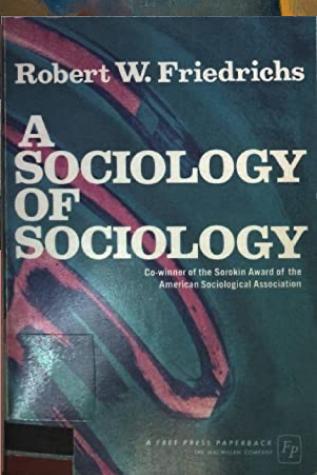
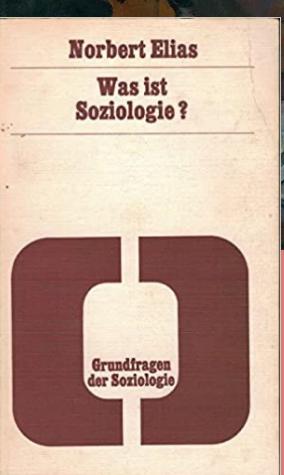
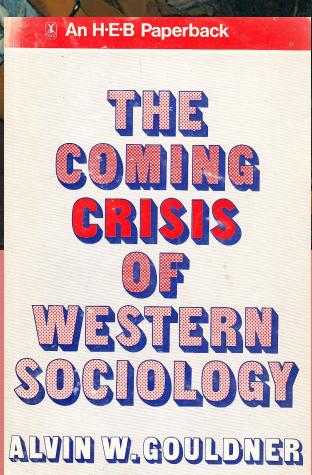


"Science"	"Humanities"
Quantitative	Qualitative
Explanatory	Interpretative
Structure	Action
Objective	Subjective
Positivism	Historicism
Macro	Micro





The problem with sociology's "evolution"



Quantitative
POSITIVISM
ANALYSIS
REALISM
SOCIAL STRUCTURE
INDIVIDUAL LEVEL
TRANSCENDENT KNOWLEDGE

versus

Qualitative
INTERPRETATION
NARRATIVE
CONSTRUCTIONISM
CULTURE
EMERGENT LEVEL
SITUATED KNOWLEDGE



The problem of
sociology's "great
divide"

The Chaos of Disciplines

11

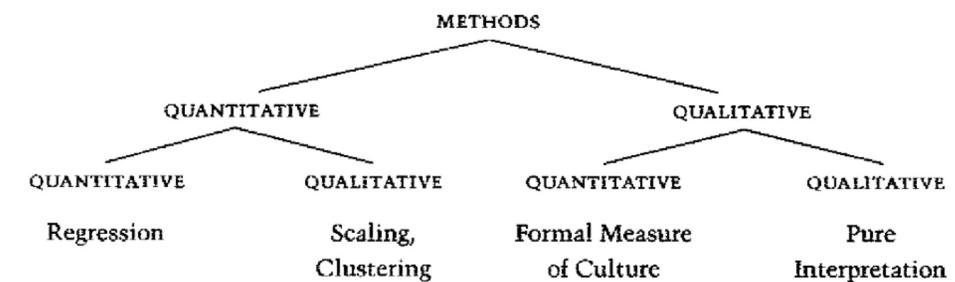


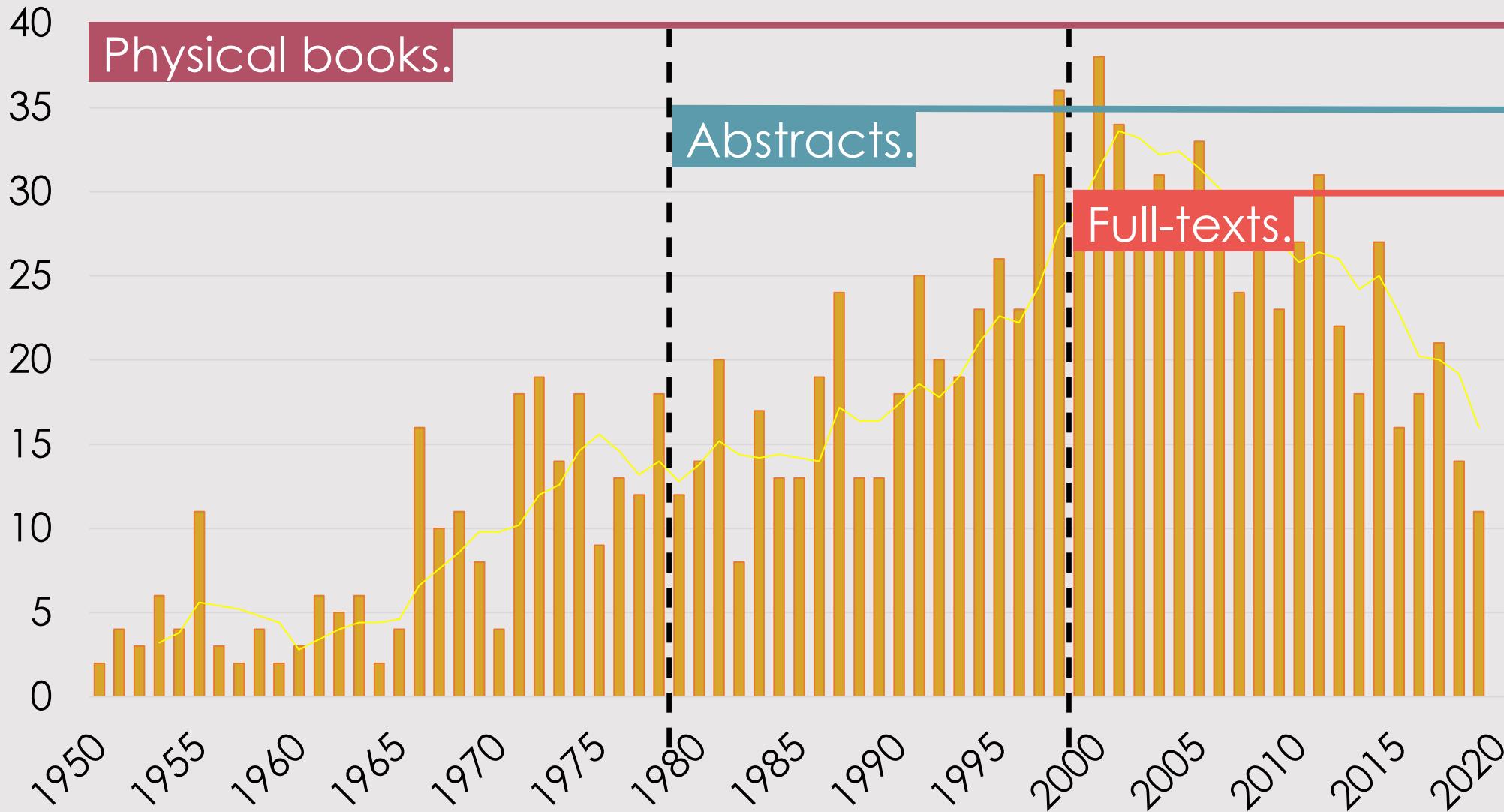
Figure 1.2

The problem with
sociology's
"evolution"

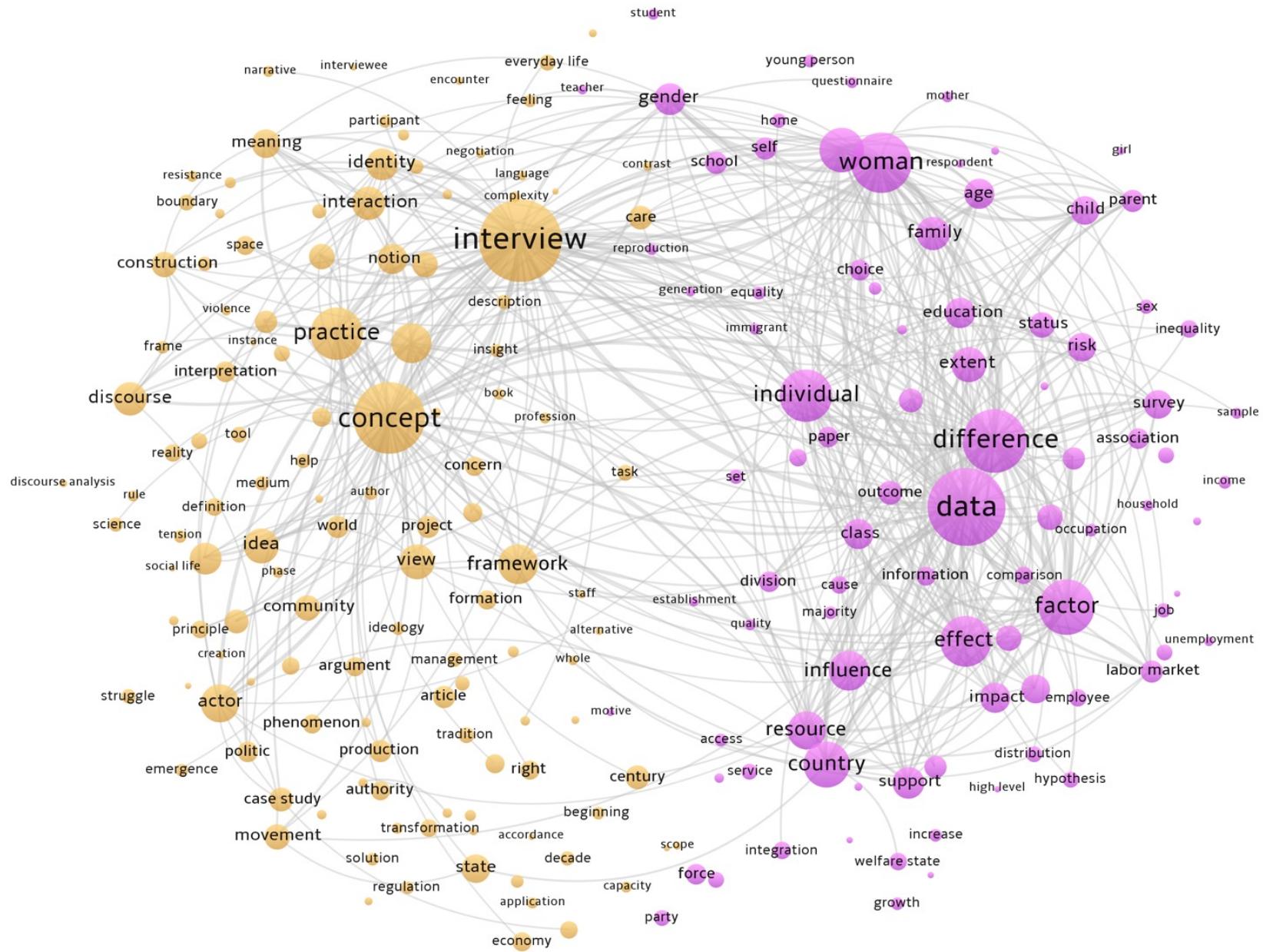
Can computers shed
new light on if
sociology's problems
exist beyond “lore”?

A “peripheral” case
of assessed
sociological knowledge

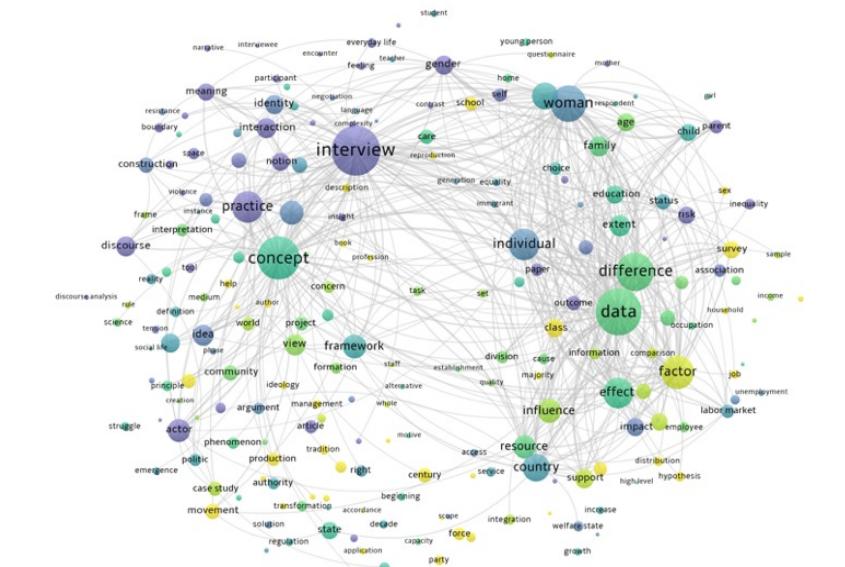
Swedish dissertations in sociology



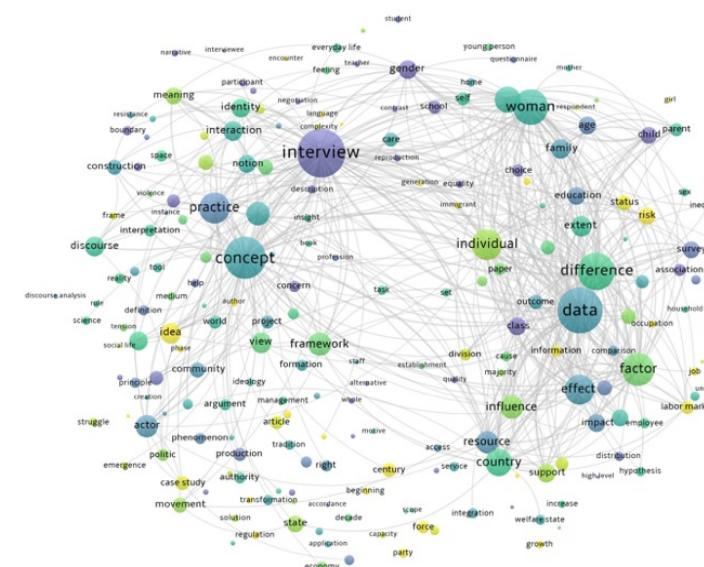
Clustered networks of word-correlations



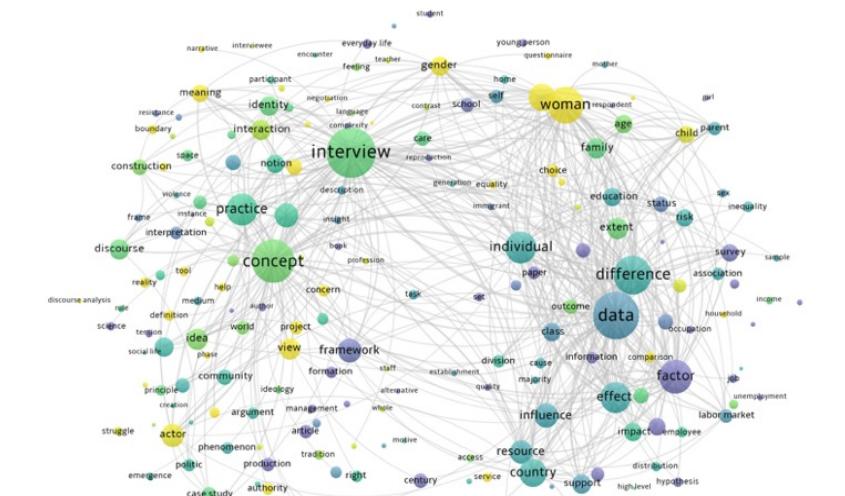
1980-1995



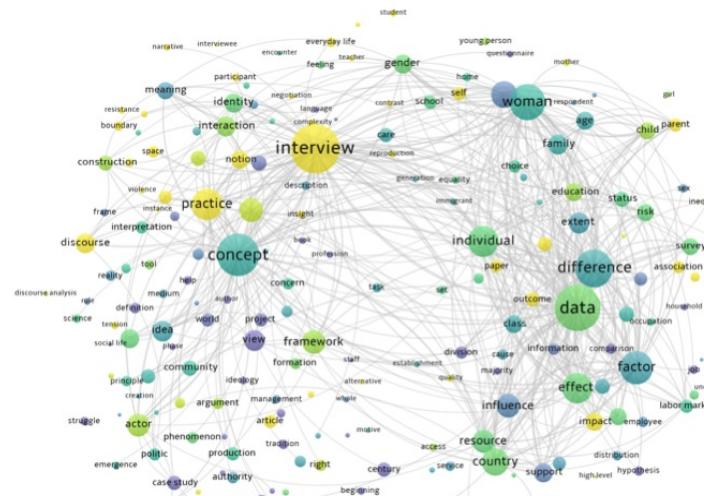
1996-2002

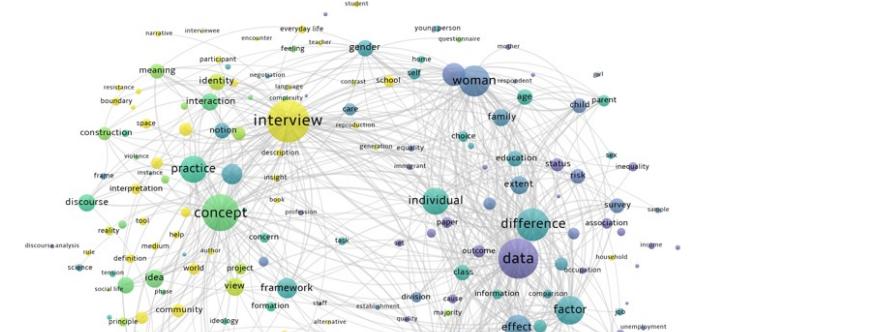


2003-2009

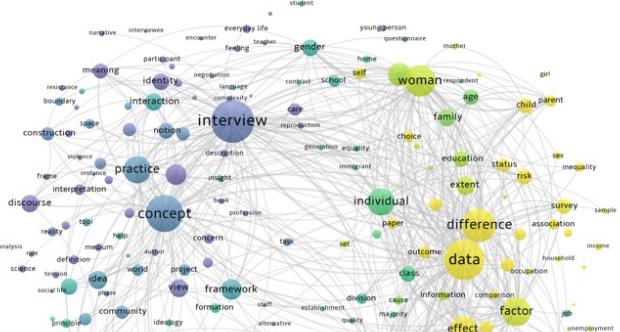


2010-2019

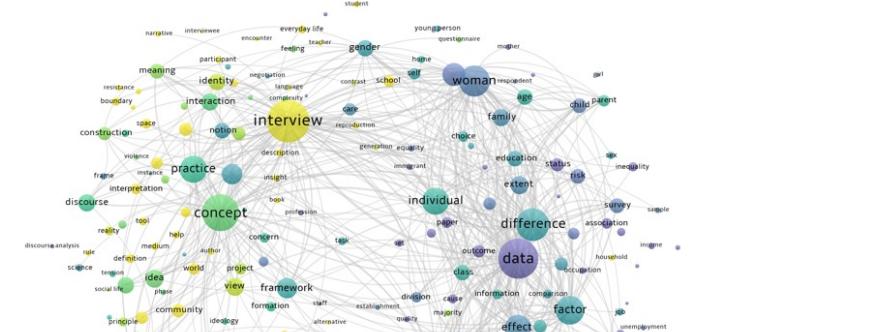




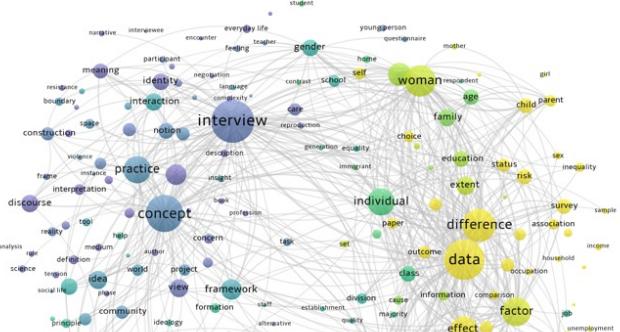
Uppsala University



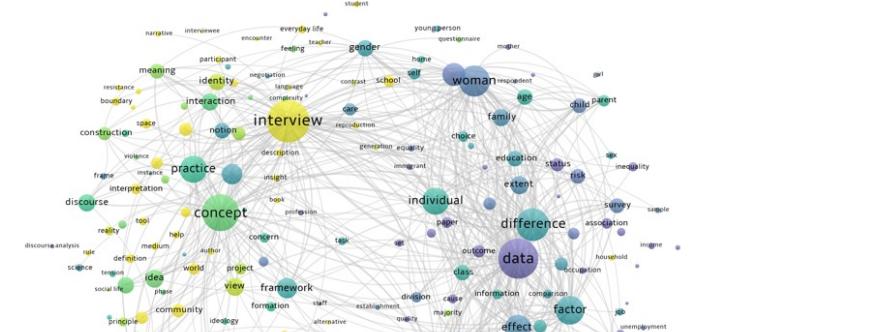
Stockholm University



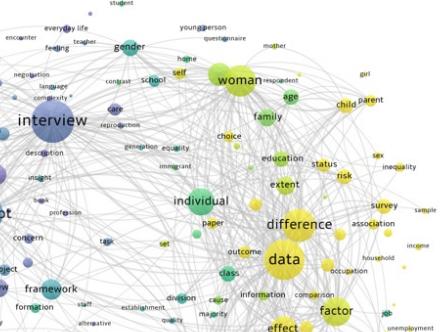
Gothenburg University



Stockholm University



Lund University

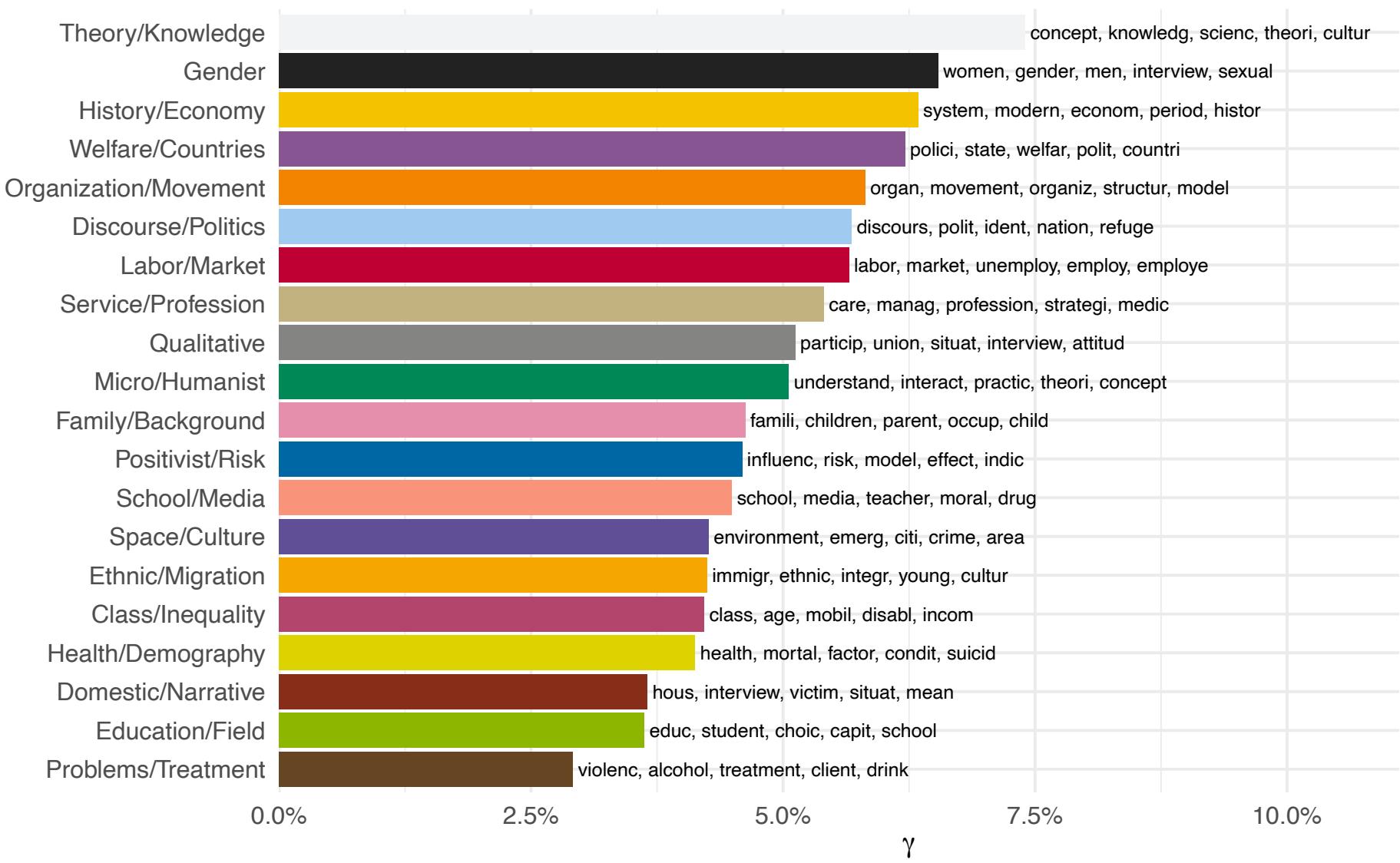


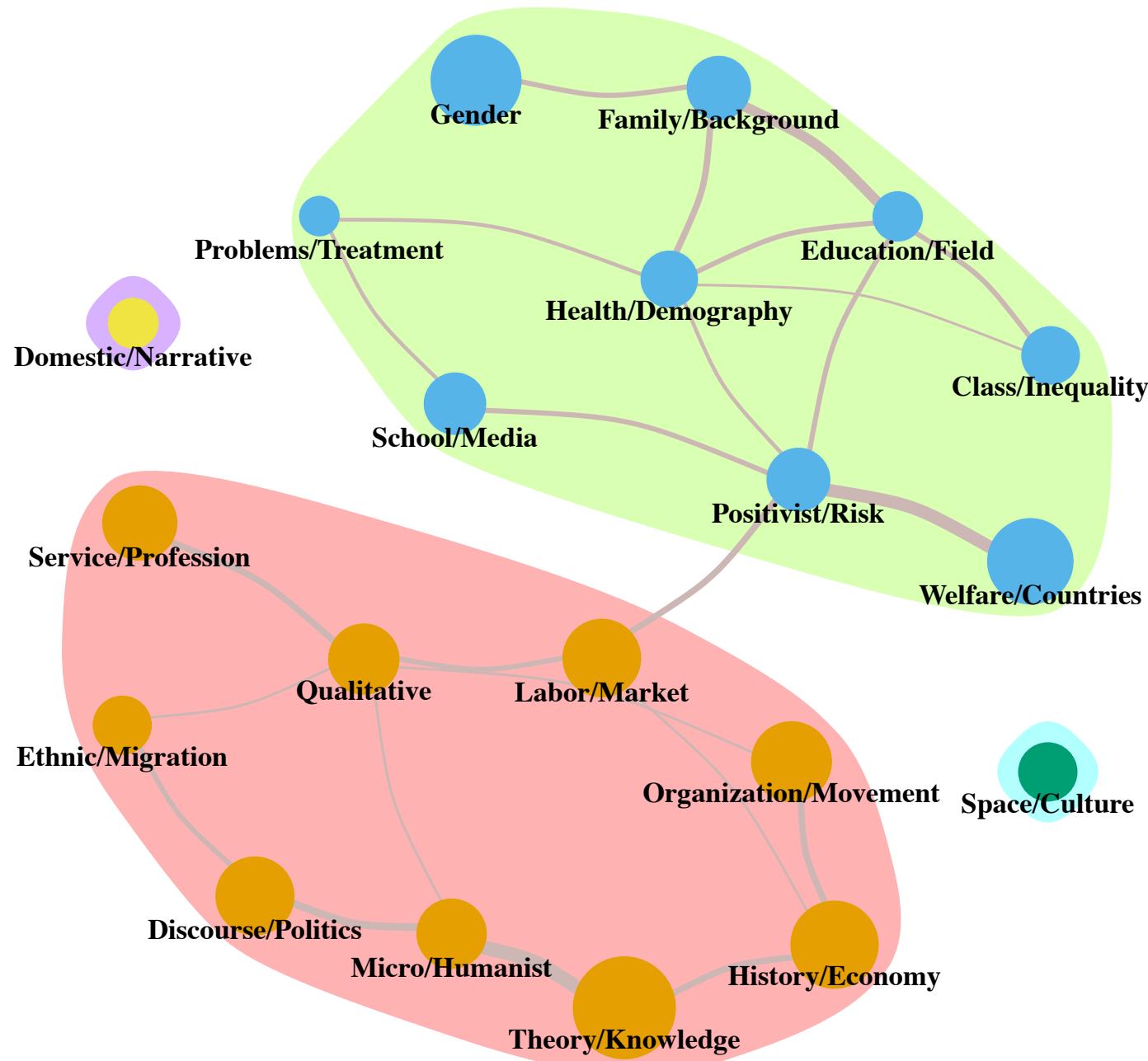
Umeå University

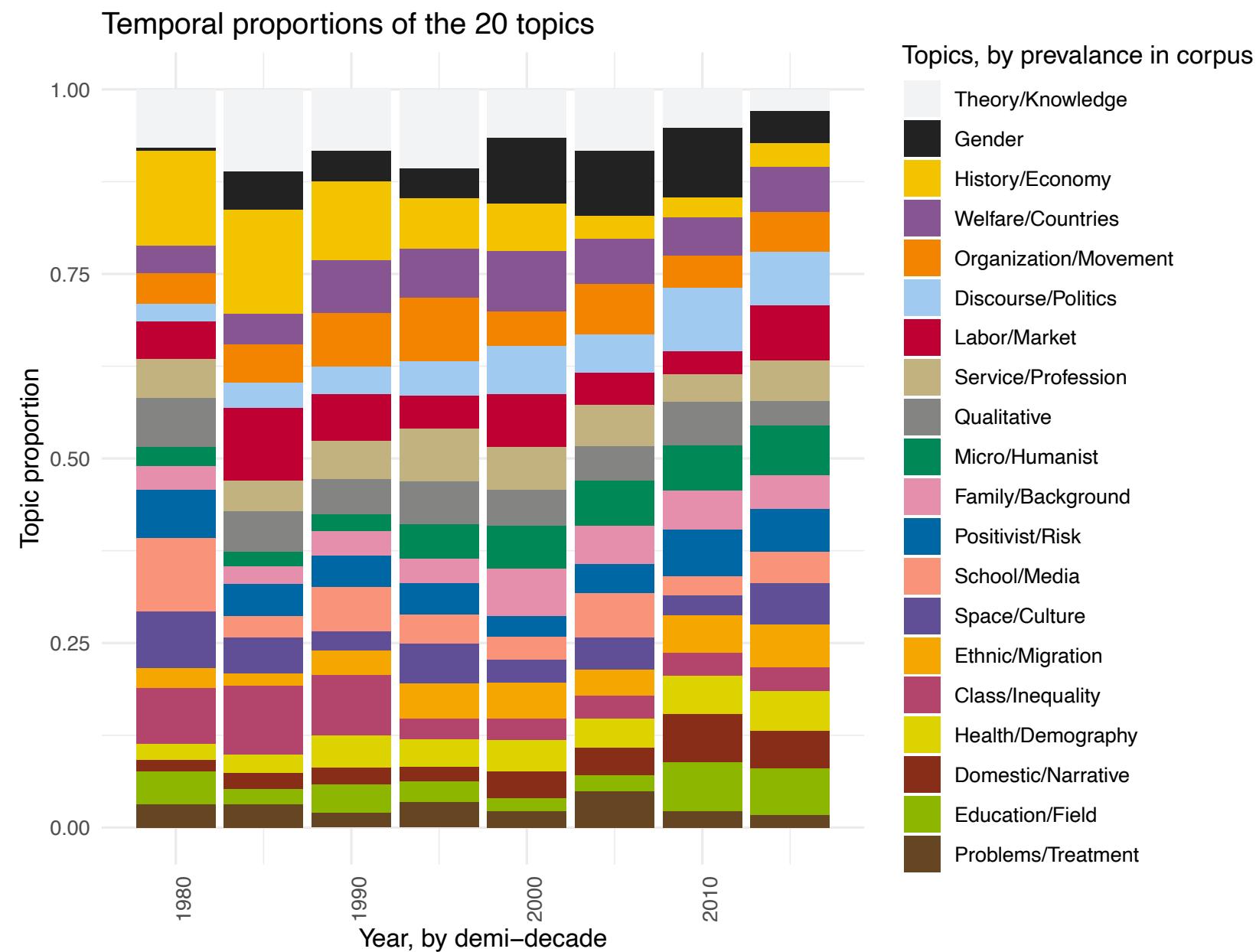
Structural topic modeling

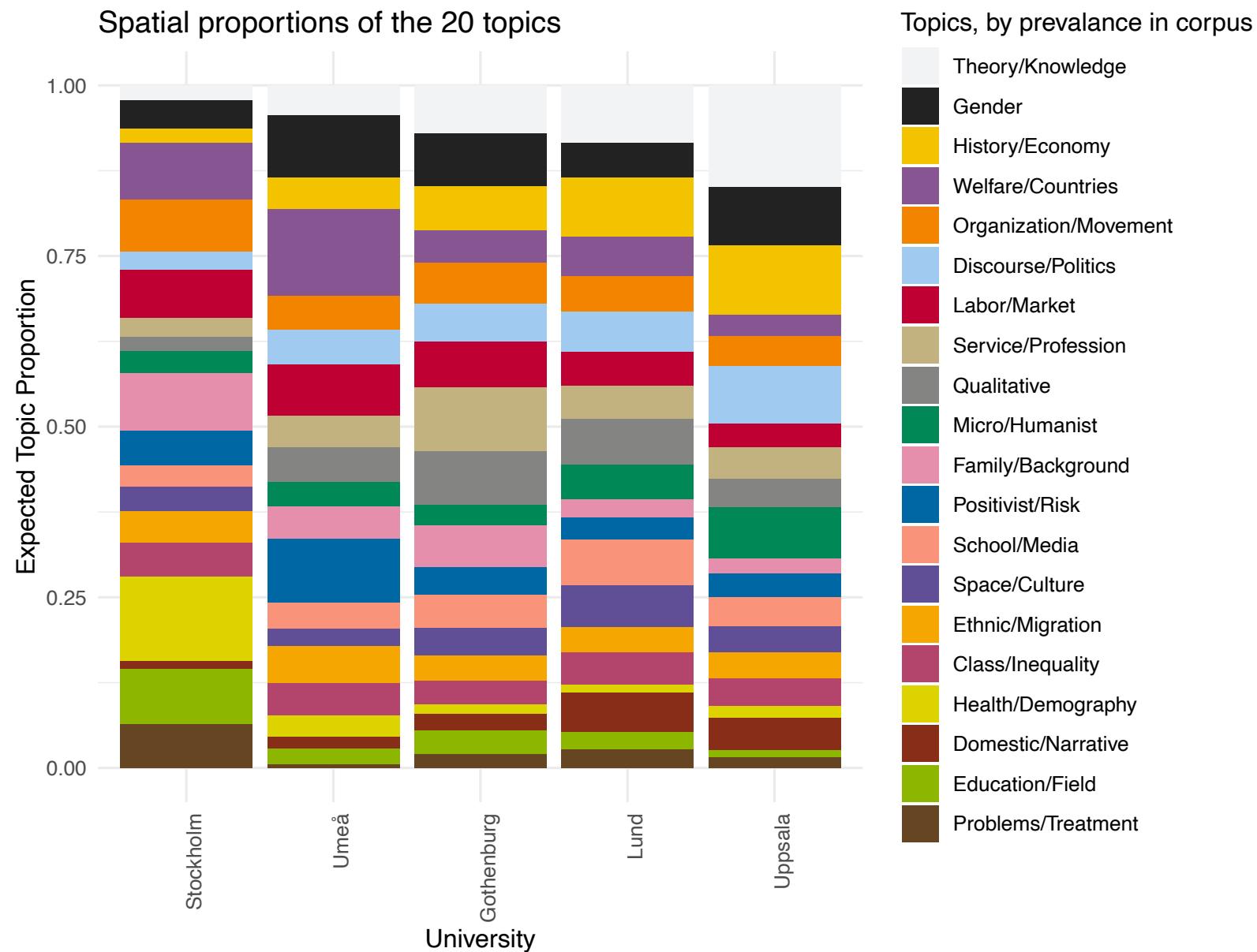
The 20 topics' prevalence in the abstract corpus

Presented with their assigned labels and most contributing words







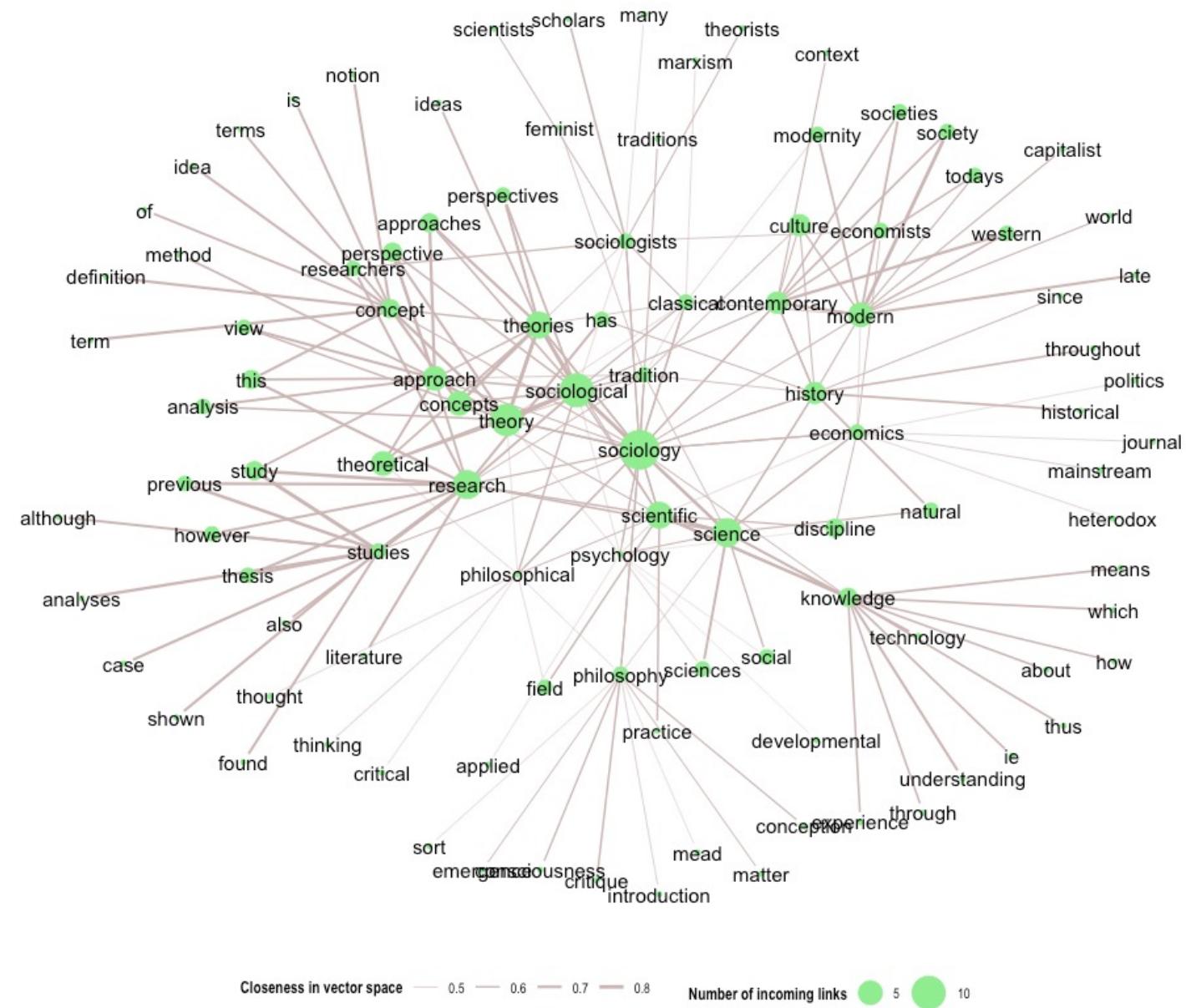


Word embeddings

```
# Create a Term-Count-Matrix
tcm <- create_tcm(it, vectorizer, skip_grams_window = 8L)
tcm

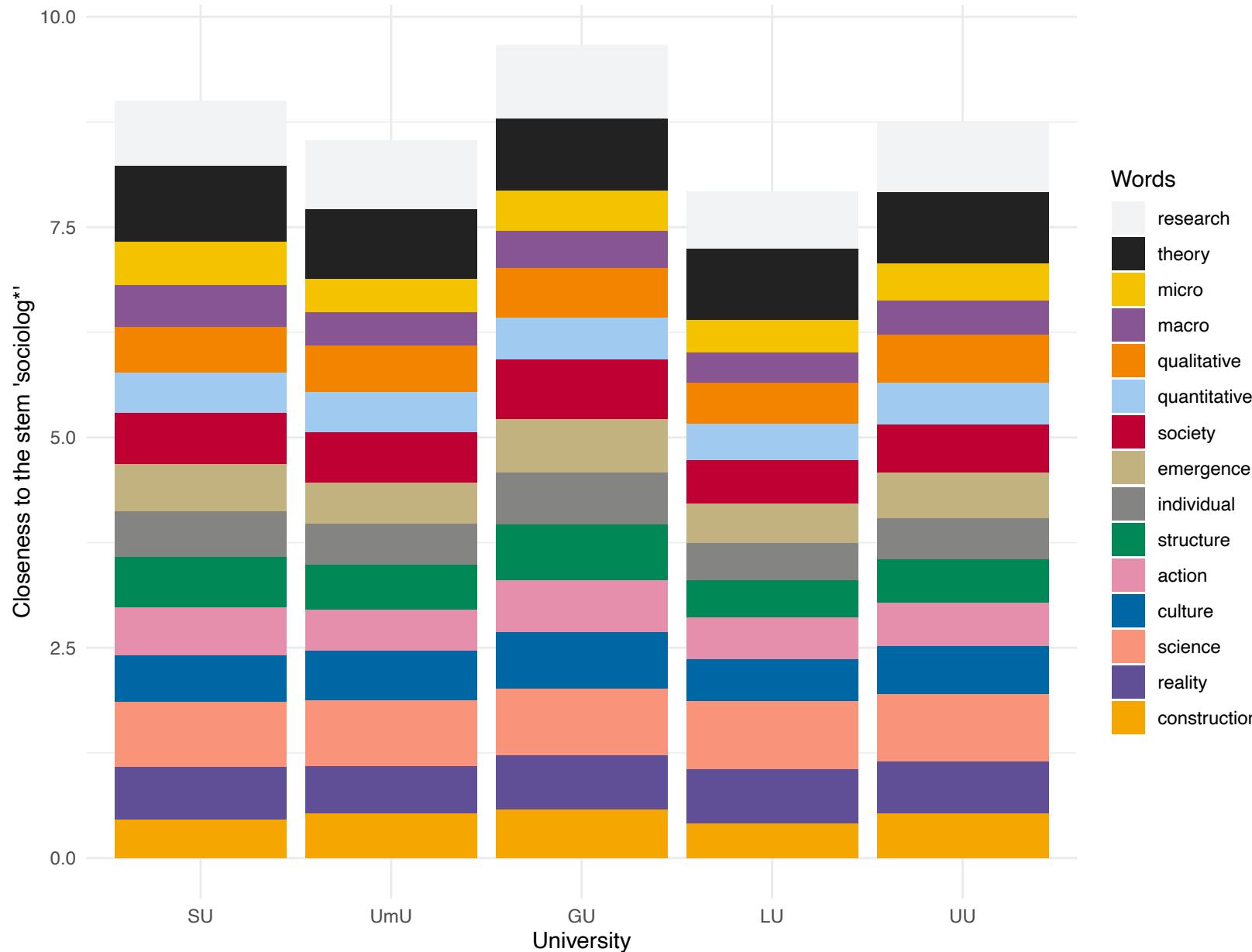
# maximum number of co-occurrences to use in the weighting
# function is the entire token set divided by 100
x_max <- length(vocab$doc_count)/100

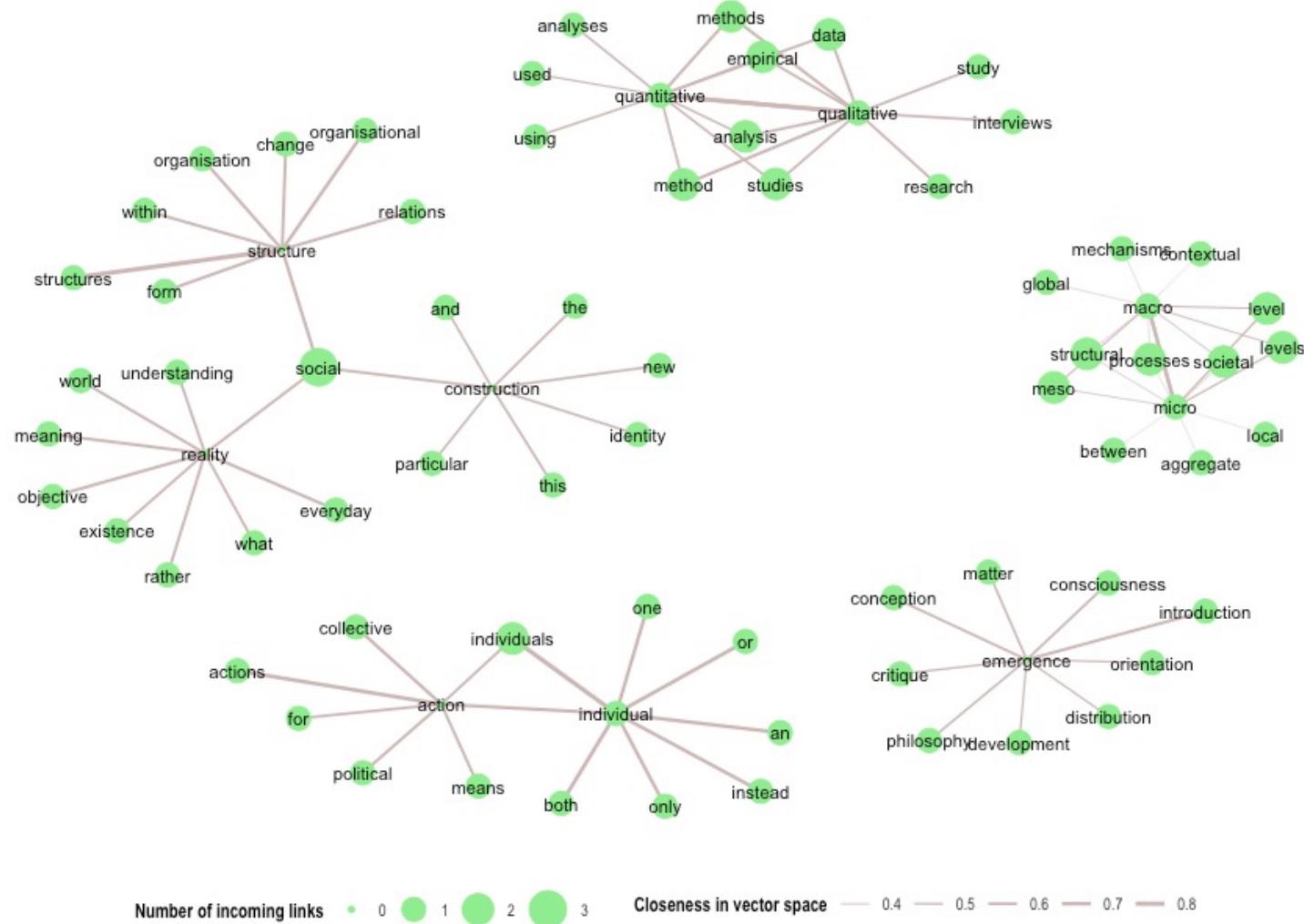
glove_model300 <- GloVe$new(rank = 300, x_max = x_max)
# (rank =dimensions)
```



Cosine similarity between each university's embedding for sociology and related words
The words represent theoretical distinctions commonly addressed within the discipline

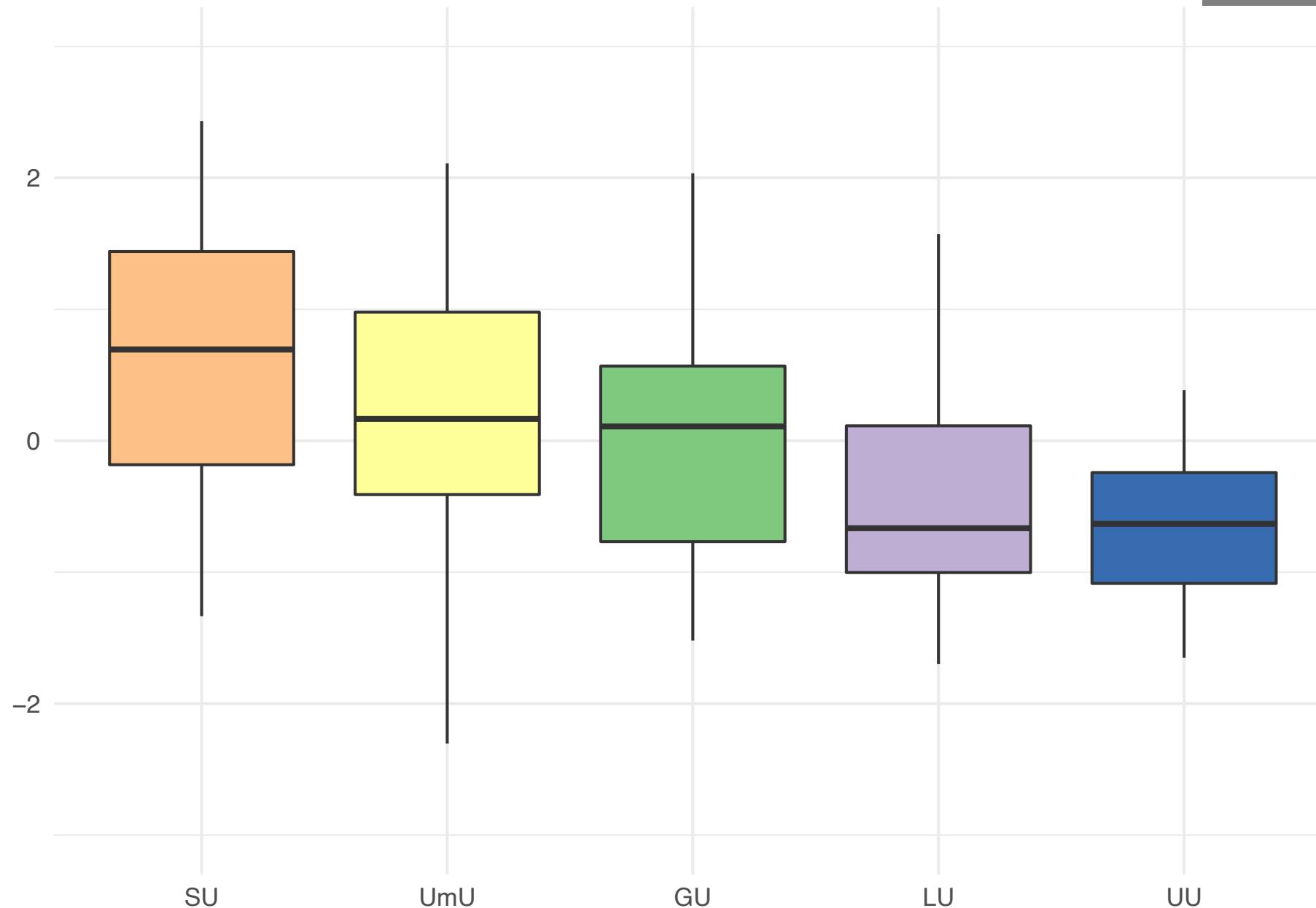
conText R package



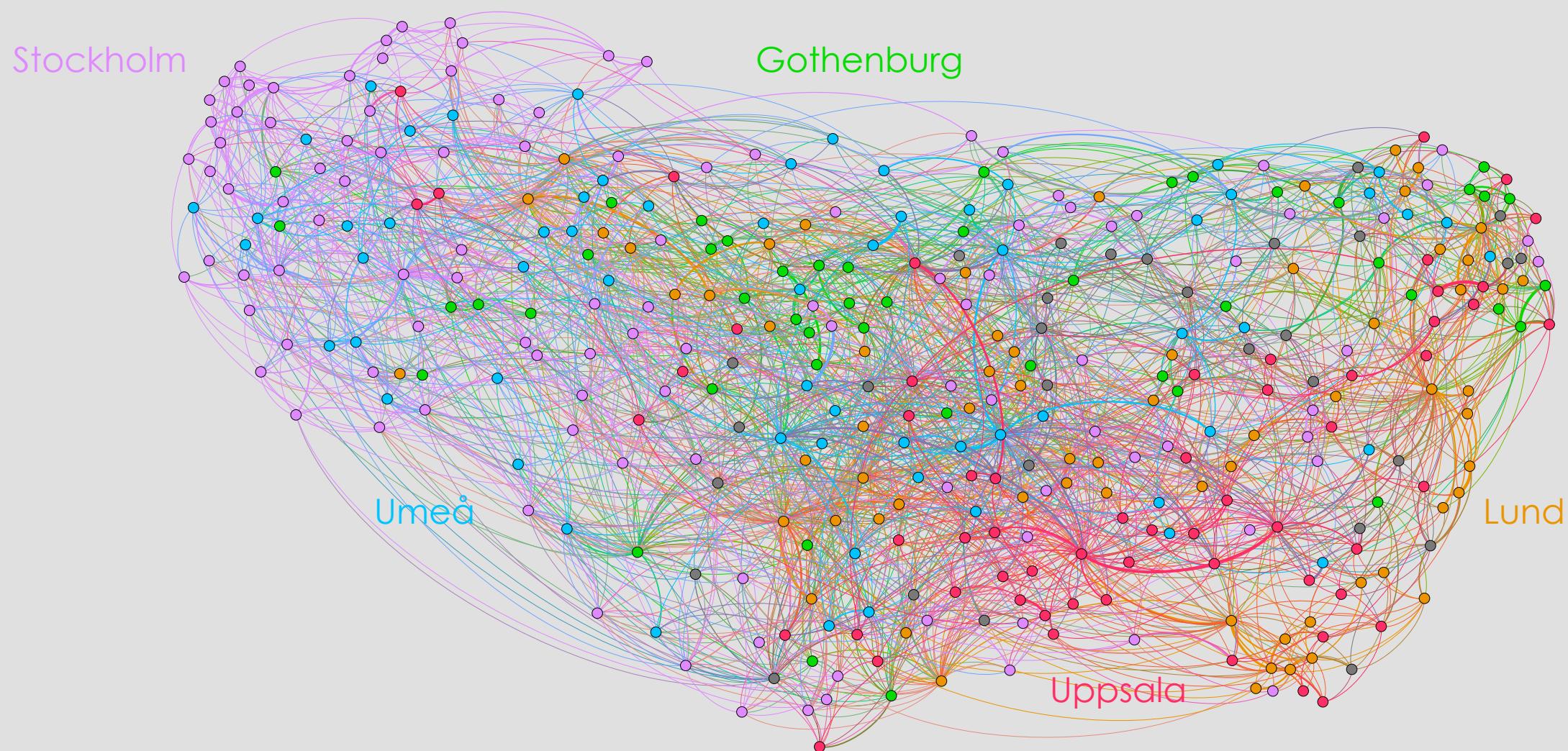


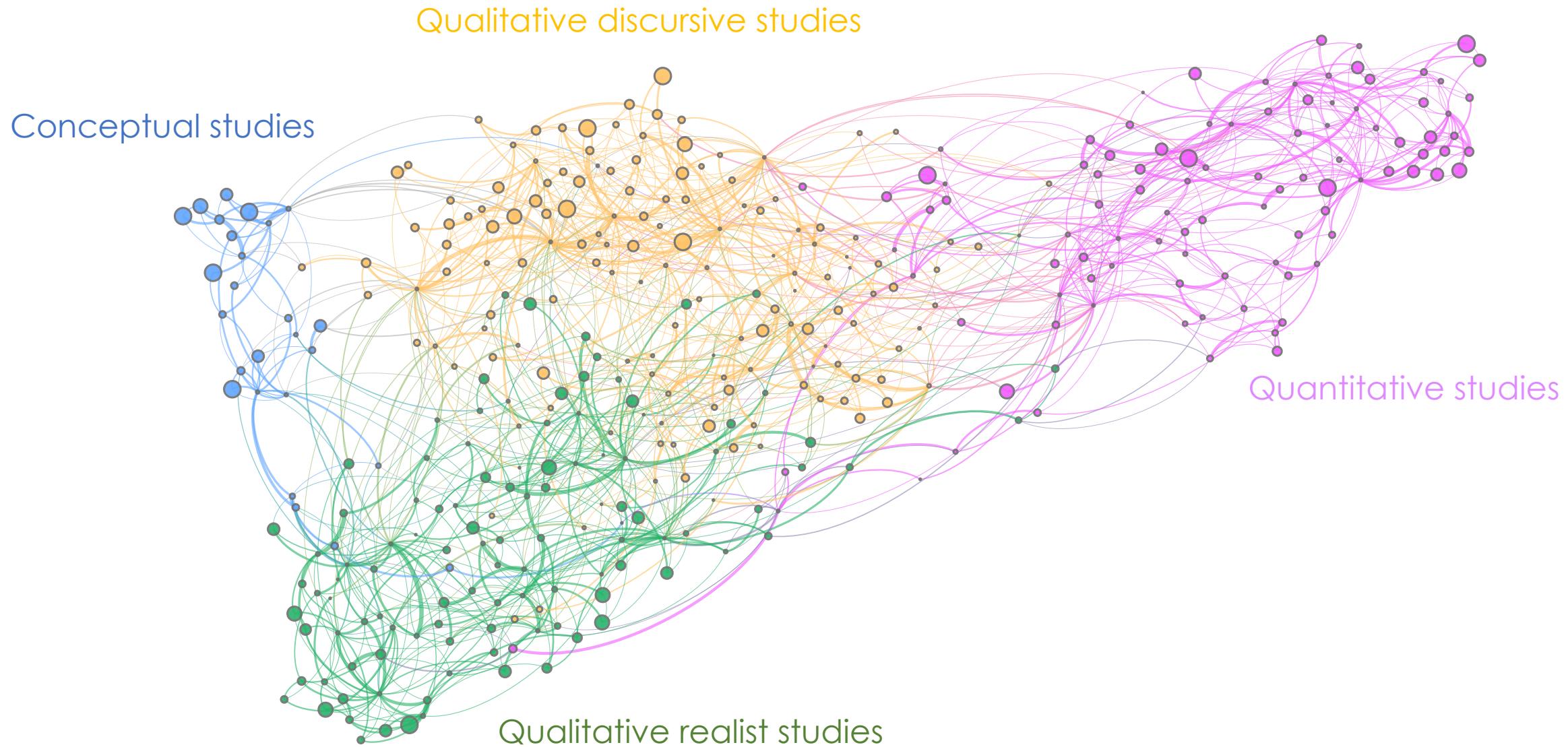
Semantic poles of quantitative and qualitative terms

text2map R package



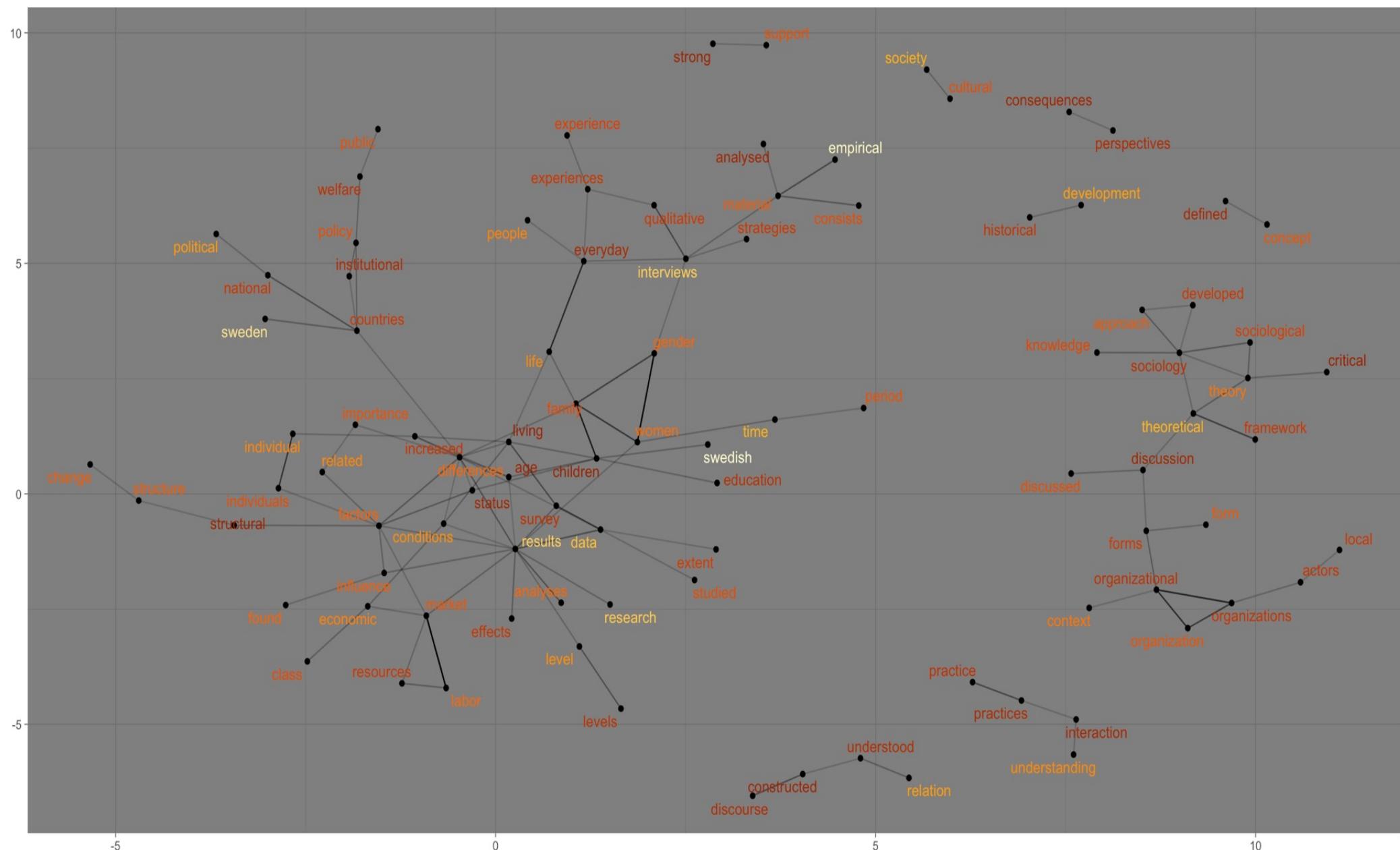
Stylometrics – bootstrap consensus trees





Thank you
for your attention!

(Have I missed some potentially insightful method?)

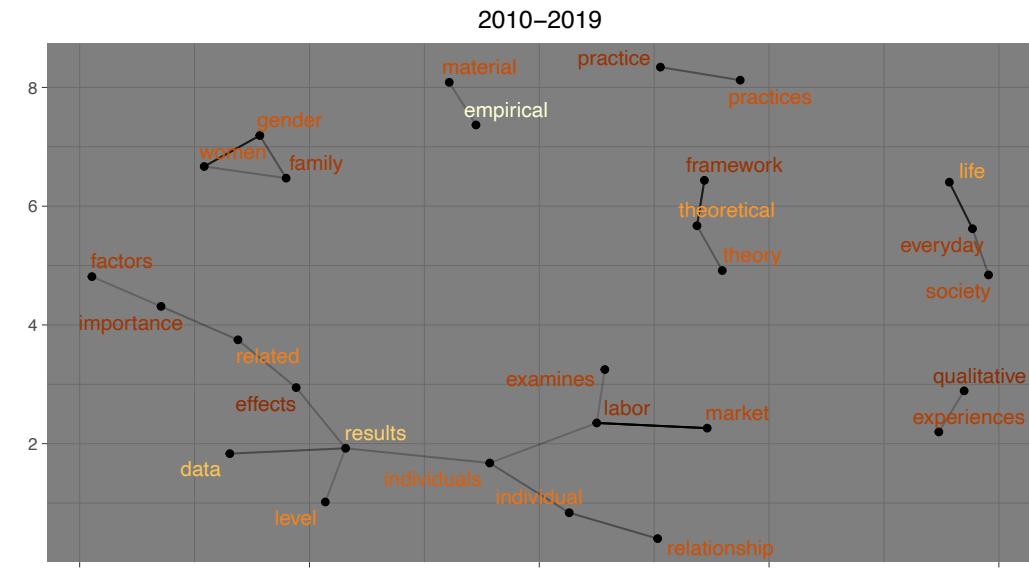
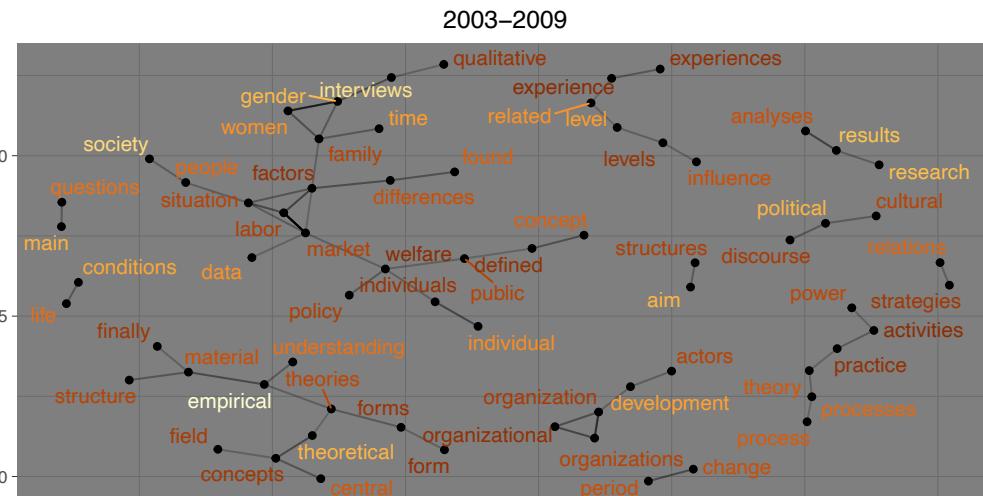
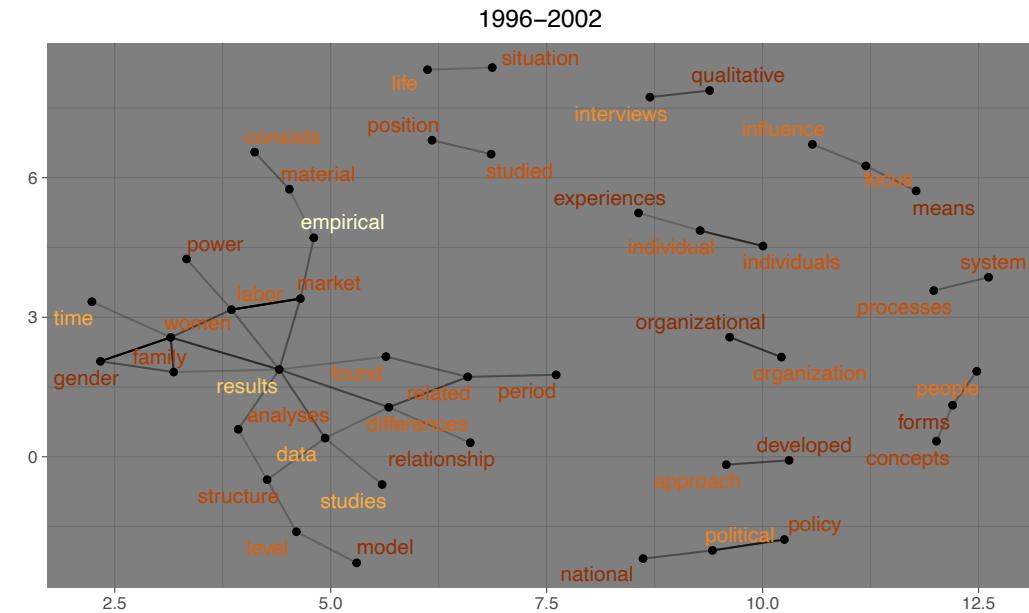
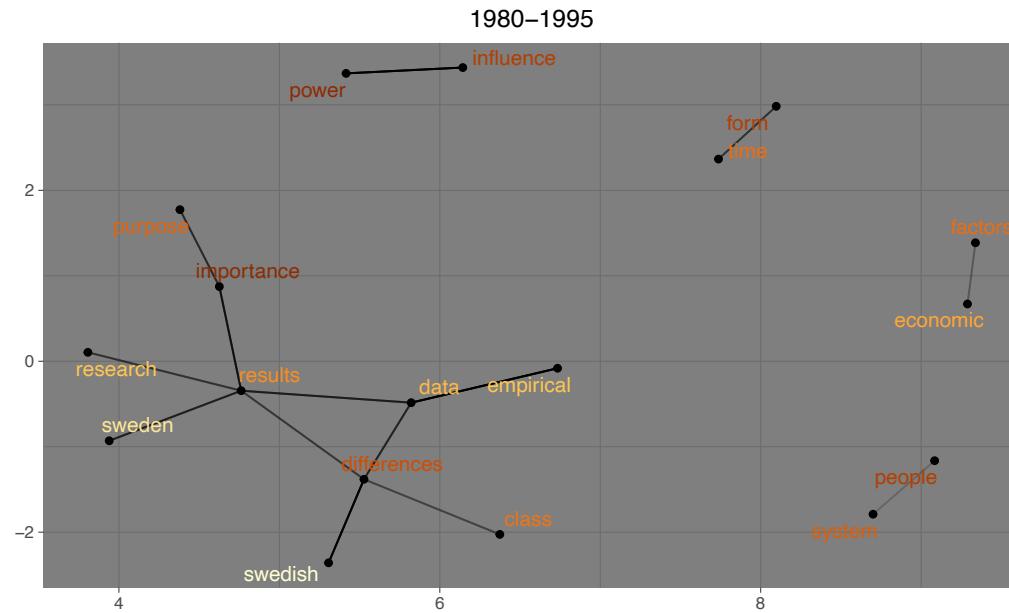


author



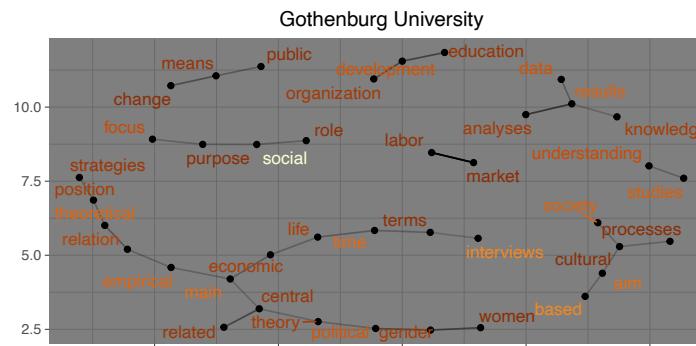
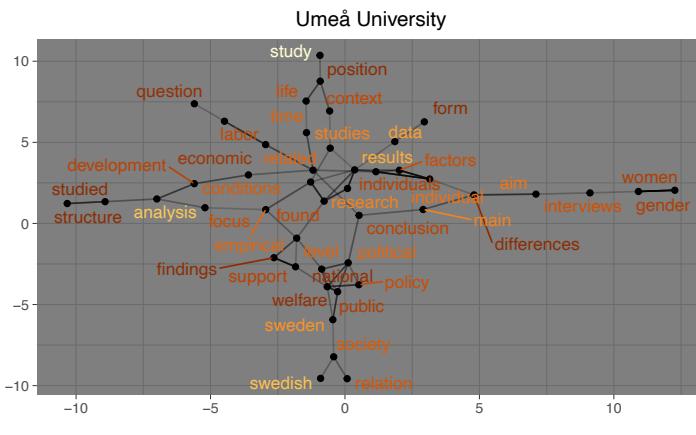
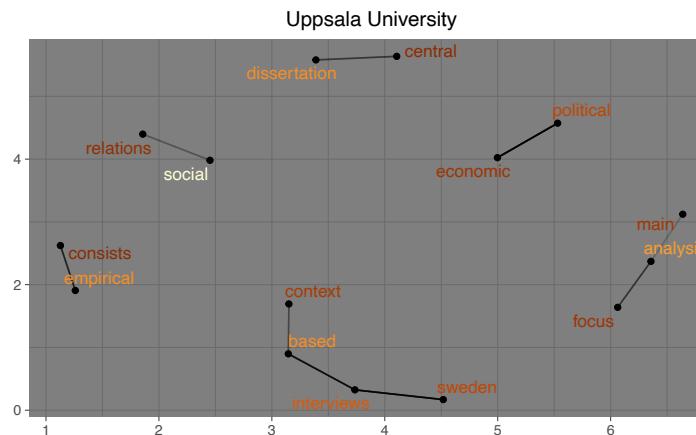
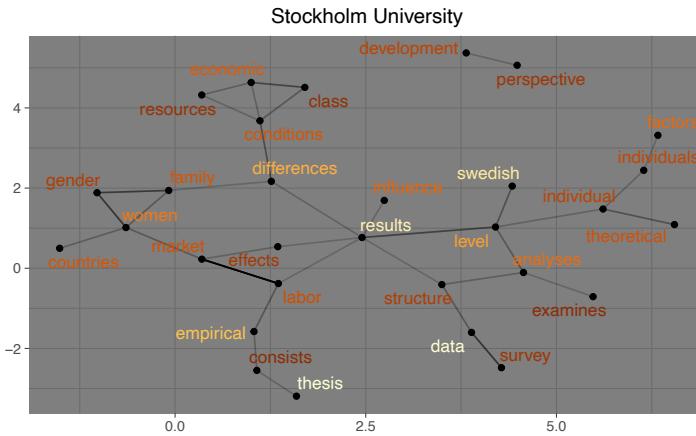
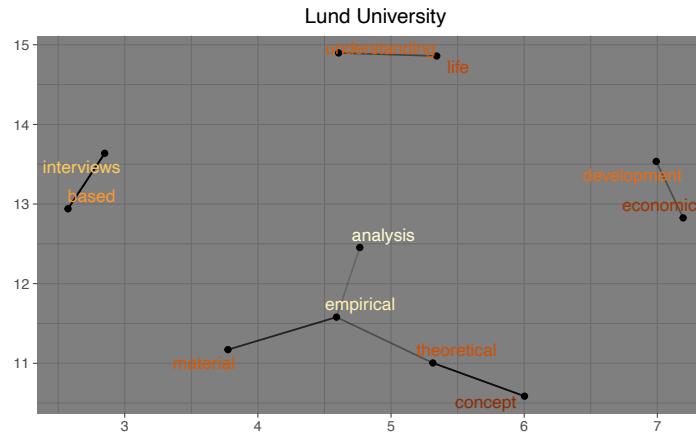
correlation





author	correlation
a	0.3
a	0.4
a	0.5
a	0.6
a	0.7
b	0.4
b	0.5
c	0.5
c	0.6
d	0.6
d	0.7

Words appearing in at least 15% of a time grouping's abstracts with a minimum correlation of 0.25

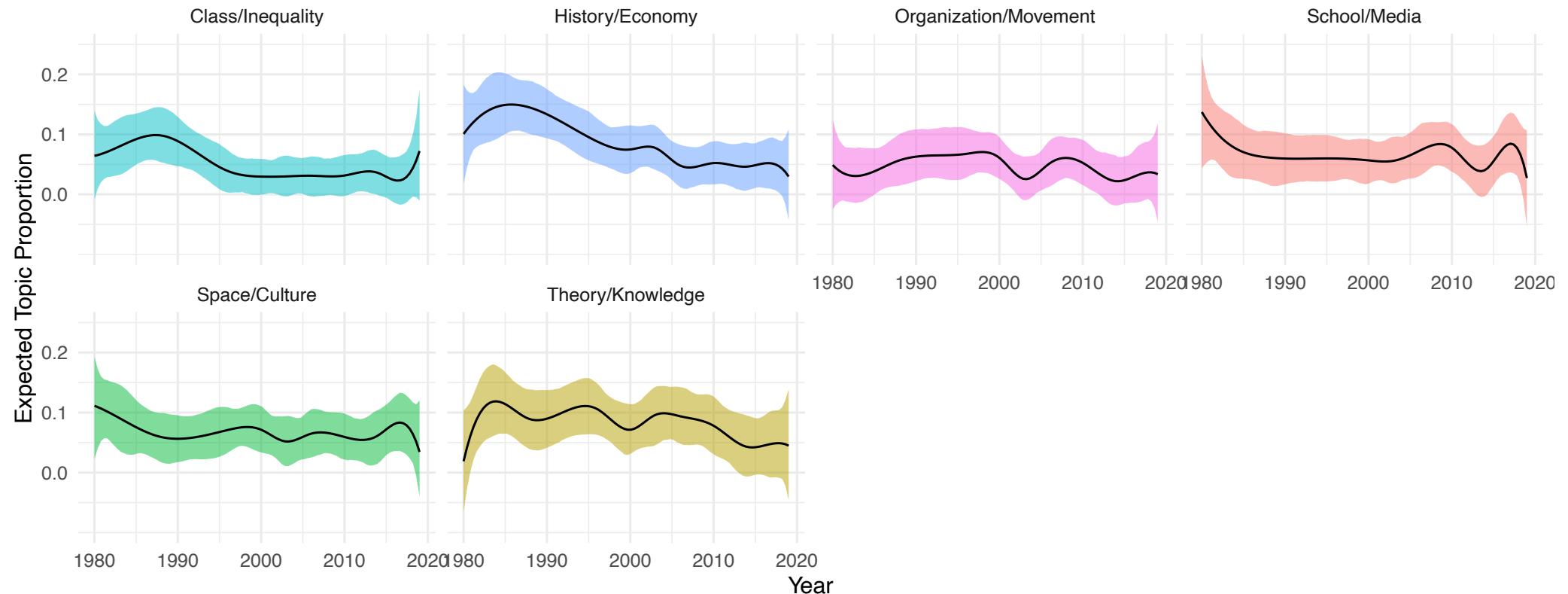


author

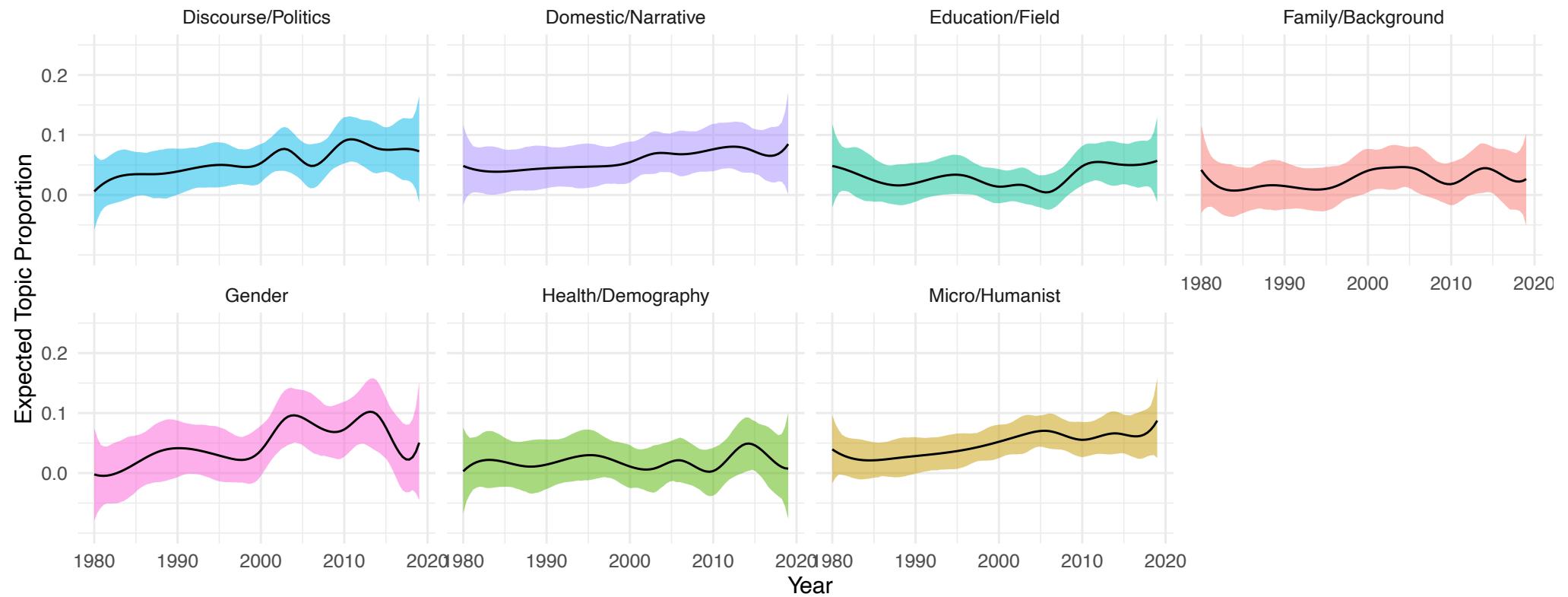
Topics balancing between the centuries



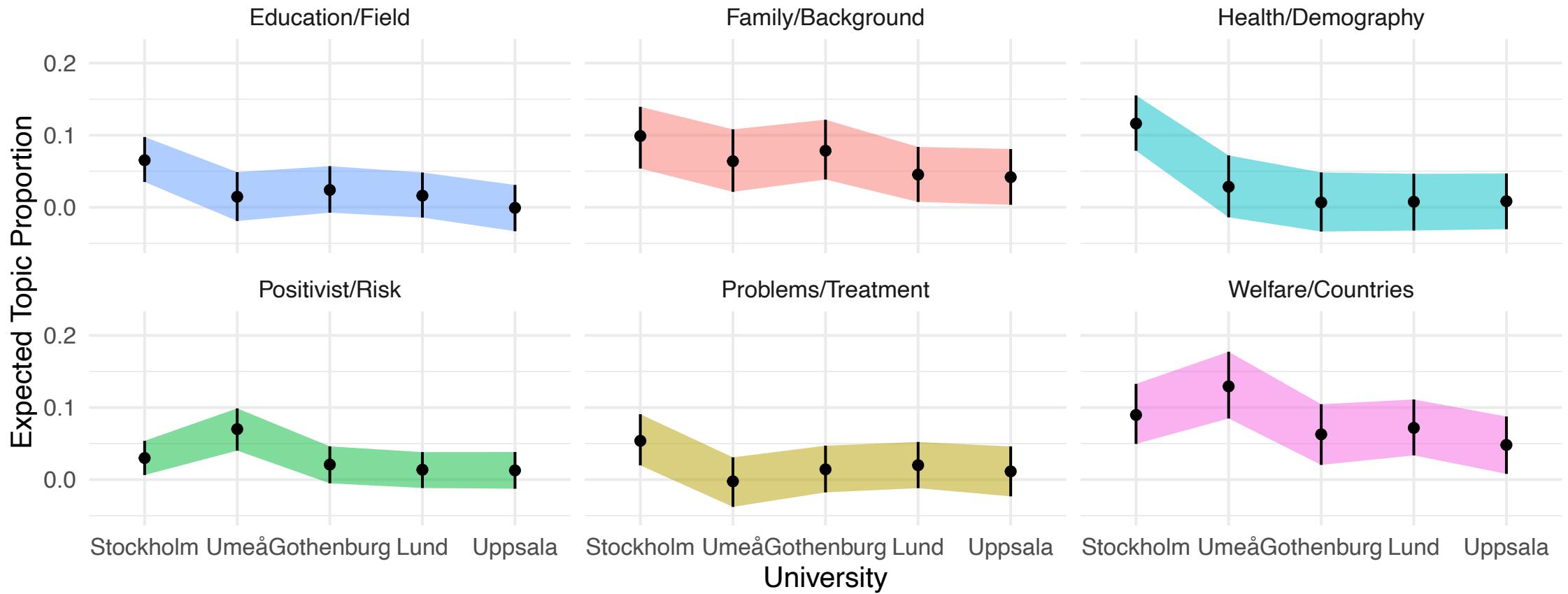
Topics leaning towards the 20th century



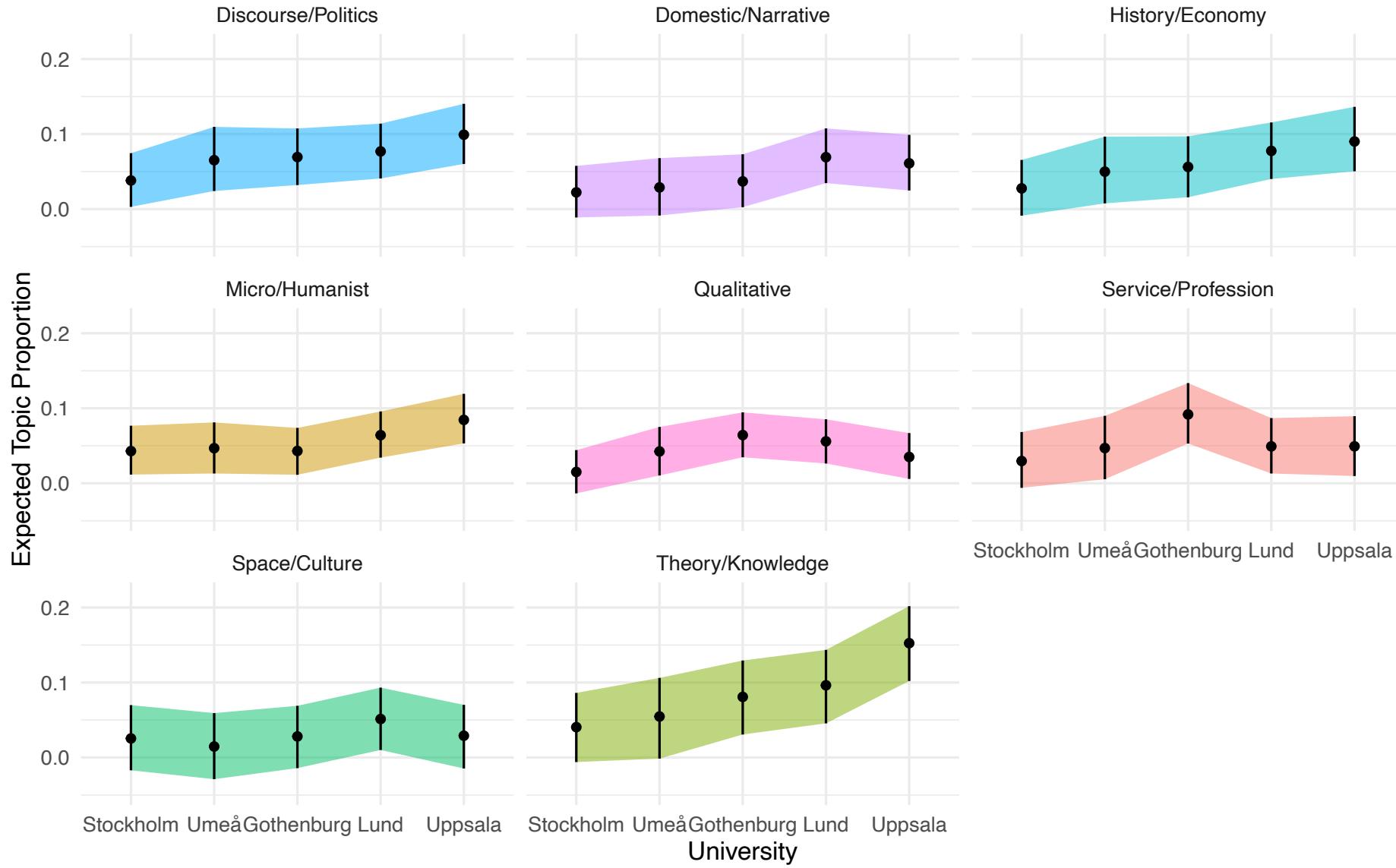
Topics leaning towards the 21st century



Topics where Stockholm or Umeå has the strongest spatial effect



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



Topics where the spatial effect deviates or is marginal

