Digital humanities

Analýza sociálních sítí a korespondenční analýza

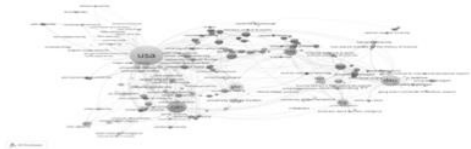
Jindřich Marek

Analýza sociálních sítí

Síťové grafy

- výpočet (matrix)
- vizualizace
- popis

a. Image speaks for itself



"El resultado es una imagen real de lo que son hoy en día las Humanidades Digitales a nivel global"

Pino-Díaz J. and Fiormonte D. (2018). "La Geopólitica De Las Humanidades Digitales: Un Caso De Estudio De DH2017 Montreal", DH2018, Mexico City.

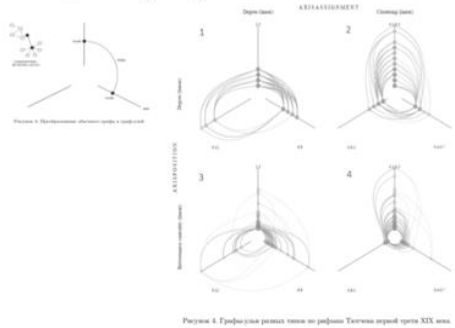
b. Helps to visualise the argument

Colavizza G. et al. (2016). "A Method for Record Linkage with Sparse Historical Data", DH2016, Krakow.

c. Highlights a specific node

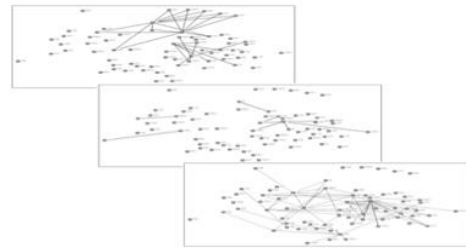
Moretti G. et al. (2016). "Building Large Persons' Networks to Explore Digital Corpora", DH2016, Krakow.

d. Comparing layouts



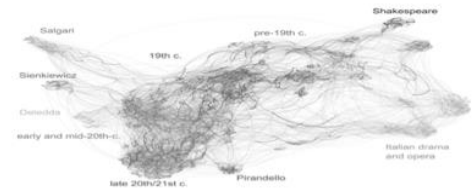
Sozinova O. (2016). "Complex Networks-Based Approach to Russian Rhyme History Description: Linguostatistics and Database", DH2016, Krakow.

e. Same network over time



Wright C. (2016). "The Formation of Australia's Economic History Community, 1950–1970: A Multi-dimensional Network Analysis", DH2016, Krakow.

f. Visualising communities



Rybicki J. et al. (2018). "Polysystem Theory And Macroanalysis. A Case Study Of Sienkiewicz In Italian", DH2018, Mexico City.

g. Image presented as a map

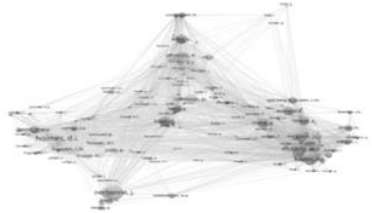
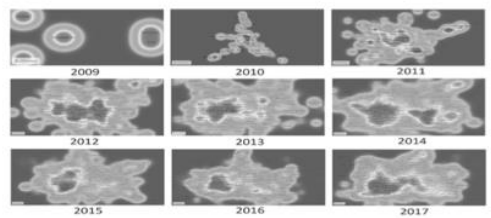


Figure 2. The provisional ACA network map in DH, data from journals *CHum*, *LLC/DSH*, and *DHQ*, 1966-2016, created using VOSviewer

Gao J. et al. (2017). "The Intellectual Structure of Digital Humanities: An Author Co-Citation Analysis", DH2017, Montreal.

h. Comparing densities



Gao J. et al. (2018). "Visualising The Digital Humanities Community: A Comparison Study Between Citation Network And Social Network", DH2018, Mexico.

i. Visualising modularity clustering

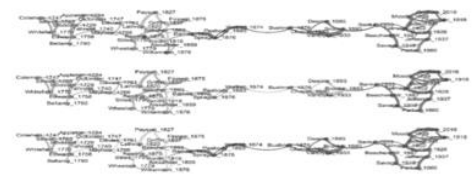
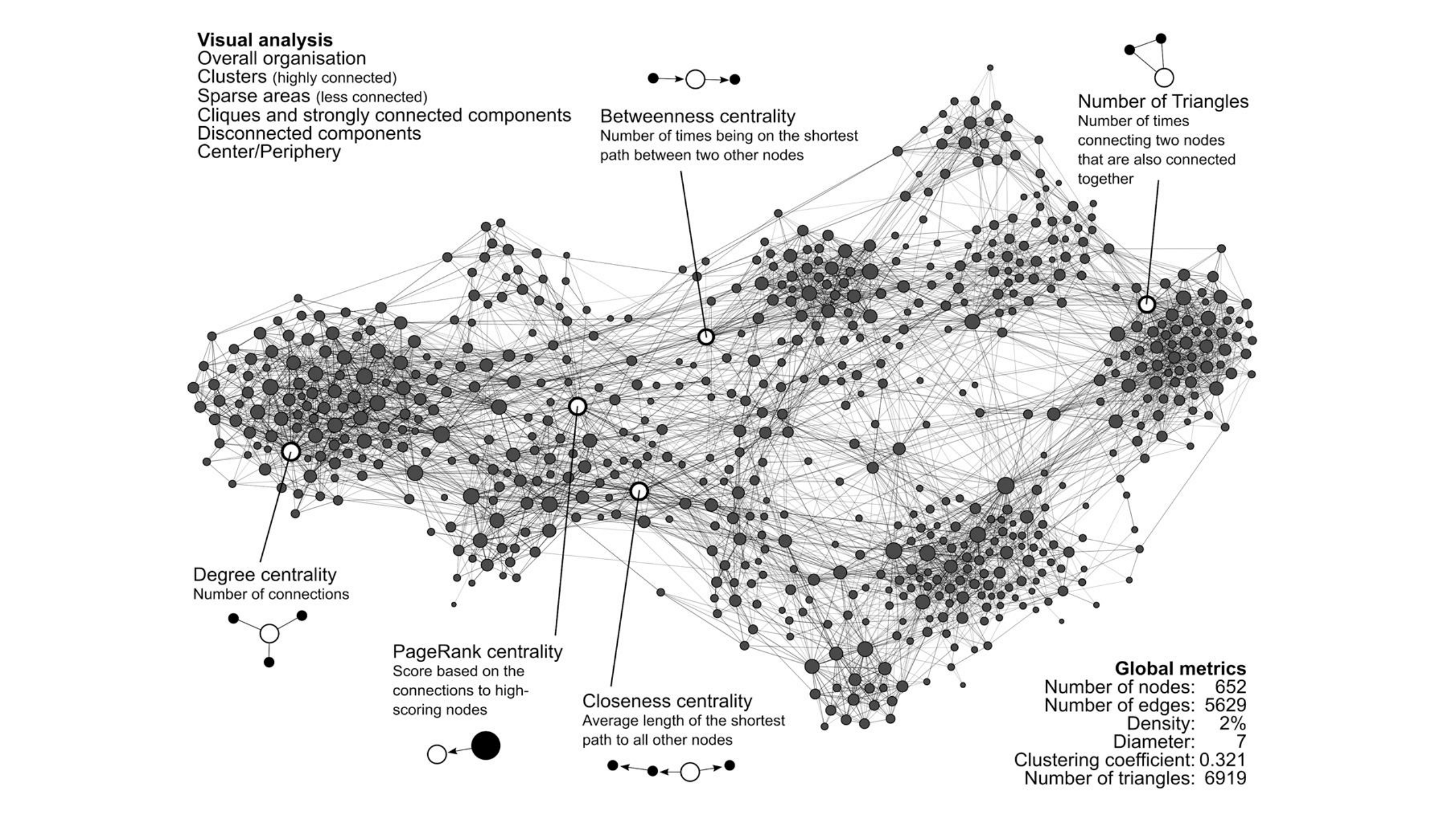
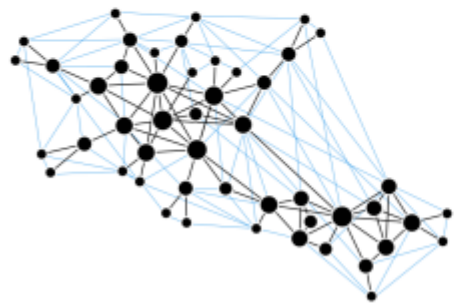


Figure 1. The four Great Awakenings in traditional classification (top); divided by modularity into 3 (center) and 4 (bottom) groups.

Choinski M. and Rybicki I. (2017). "Networks of the Great Awakenings: Classification of Puritan Sermons by Word Usage Statistics", DH2017, Montreal.


Visual analysis
Overall organisation
Clusters (highly connected)
Sparse areas (less connected)
Cliques and strongly connected components
Disconnected components
Center/Periphery
Betweenness centrality
Number of times being on the shortest path between two other nodes
Number of Triangles
Number of times connecting two nodes that are also connected together
Degree centrality
Number of connections
PageRank centrality
Score based on the connections to high-scoring nodes
Closeness centrality
Average length of the shortest path to all other nodes
Global metrics
Number of nodes: 652
Number of edges: 5629
Density: 2%
Diameter: 7
Clustering coefficient: 0.321
Number of triangles: 6919

	NOTION		VISUAL ANALYSIS		COMPUTATIONAL ANALYSIS		INTERPRETATIVE POTENTIAL
Global		**Graph size (nodes)** SIMPLE DEFINITION How many nodes are there in the graph?	VISUAL PATTERN There are few or many nodes. HERMENEUTIC CRITERIA Counting the nodes.	ISSUES Though it is easy to have an estimation of the total number of nodes, visualization decisions (for example, setting node sizes on a large scale) can make nodes with few connections difficult to see.	TOPOLOGICAL PATTERN Nodes count. COMPUTATIONAL CRITERIA Counting the nodes.	ISSUES None. Note that in graph theory the count of nodes is referred to as "graph order" while the "graph size" refers to the count of edges.	USEFULNESS Very basic information but useful when comparing networks. "TRANSLATION" The intuition of the **size** of a network is appropriate.
Global		**Graph size (edges)** SIMPLE DEFINITION How many edges are there in the graph?	VISUAL PATTERN There are few or many edges. HERMENEUTIC CRITERIA Counting the edges.	ISSUES The total number is hard to estimate as soon as the graph is not a simple diagram anymore. The distribution of edges and their weight has an influence on the visual estimation. The difficulty to count edges visually is a known issue, and probably impossible to overcome.	TOPOLOGICAL PATTERN Edges count. COMPUTATIONAL CRITERIA Counting the edges.	ISSUES None. Note that in graph theory the count of edges is referred to as "graph size" while the count of nodes is "graph order".	USEFULNESS Very basic information but useful when comparing networks. "TRANSLATION" Sometimes translated as **size** in natural language, but the number of edges is usually expressed in comparison to the number of nodes to indicate density or complexity, not for itself.
Global		**Density** SIMPLE DEFINITION How connected are the nodes overall?	VISUAL PATTERN The graph is more or less compact. If only certain parts are more compact, see "clusters" below. HERMENEUTIC CRITERIA Accumulation of edges, cluttered groups of nodes, "hairball". Easier to estimate in situation of comparison.	ISSUES The less edges there are, the easier to estimate. High densities are difficult to distinguish because the overall appearance of a graph with 60% edges will looks close to a graph with 100%. The visual aspect also depends on the layout algorithm used: some are more efficient at representing clusters.	TOPOLOGICAL PATTERN Network density. COMPUTATIONAL CRITERIA Divide the actual number of edges by the number of all potential edges.	ISSUES The formula of density slightly changes depending on the type of networks (directed or not, self-loops allowed or not...).	USEFULNESS A very important notion that allows to compare networks of different sizes if they are produced in the same way or on the basis of comparable data sets. "TRANSLATION" **Density**, **complexity**, **completeness**.
			VISUAL PATTERN	ISSUES	TOPOLOGICAL PATTERN	ISSUES	USEFULNESS

Analýza sociálních sítí: pojmy

- řízený / neřízený graf
- modularita
- centralita vlastního vektoru

- (un)directed layout
- modularity
- eigenvector-centrality

Nástroje

- Gephi
  - https://gephi.org
- R: statnet, igraph, tidygraph, ggraph, …
  - https://statnet.org/workshop-intro-sna-tools/

The Open Graph Viz Platform

Gephi is the leading visualization and exploration software for all kinds of graphs and networks. Gephi is open-source and free.

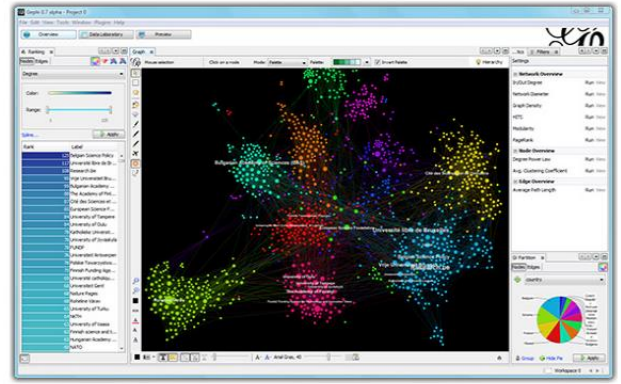
Runs on Windows, Mac OS X and Linux.

Learn More on Gephi Platform »

Download FREE Gephi 0.10

Release Notes | System Requirements

- Features
- Quick start
- Screenshots
- Videos



Support us! We are non-profit. Help us to innovate and empower the community by donating only 8€:

Donate

APPLICATIONS

- **Exploratory Data Analysis**: intuition-oriented analysis by networks manipulations in real time.
- **Link Analysis**: revealing the underlying structures of associations between objects.
- **Social Network Analysis**: easy creation of social data connectors to map community organizations and small-world networks.
- **Biological Network analysis**: representing patterns of biological data.
- **Poster creation**: scientific work promotion with hi-quality printable maps.

Learn More »

METRICS READY

- **Centrality**: used in sociology to indicate how well a node is connected. Available: degree (power-law), betweenness, closeness.
- **And more**: density, path length, diameter, HITS, modularity, clustering coefficient.

Learn More »

TECHNOLOGY

- **Ergonomic interface**: no programming skills needed
- **High-performance**: built-in rendering engine.
- **Native file formats**: GDF (GUESS), GraphML (NodeXL), GML, NET (Pajek), GEXF and more.
- **Customizable by plugins**: layouts, metrics, data sources, manipulation tools, rendering presets and more.

Learn More »

Like Photoshop™ for graphs.

— the Community

LATEST NEWS

- Gephi 0.10.0 released
- Gephi Lite
- Gephi Week 2022: debriefing
- Call for participants: Gephi code sustainability retreat 2022
- Transition to semantic versioning

See All »

Tweets by @Gephi

PAPERS

Bastian M., Heymann S., Jacomy M. (2009). *Gephi: an open source software for exploring and manipulating networks.* International AAAI Conference on Weblogs and Social Media. From AAAI [PDF].

Learn More »

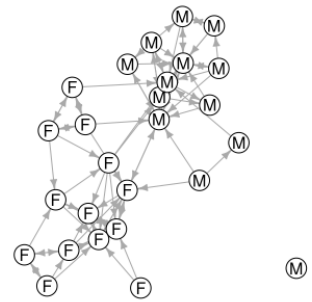
GEPHI 0.10.1 IS HERE!

The latest Gephi version just got released, let us know what you think on Twitter or Facebook.

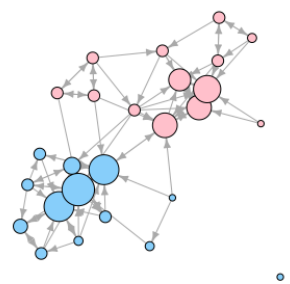
Learn More »

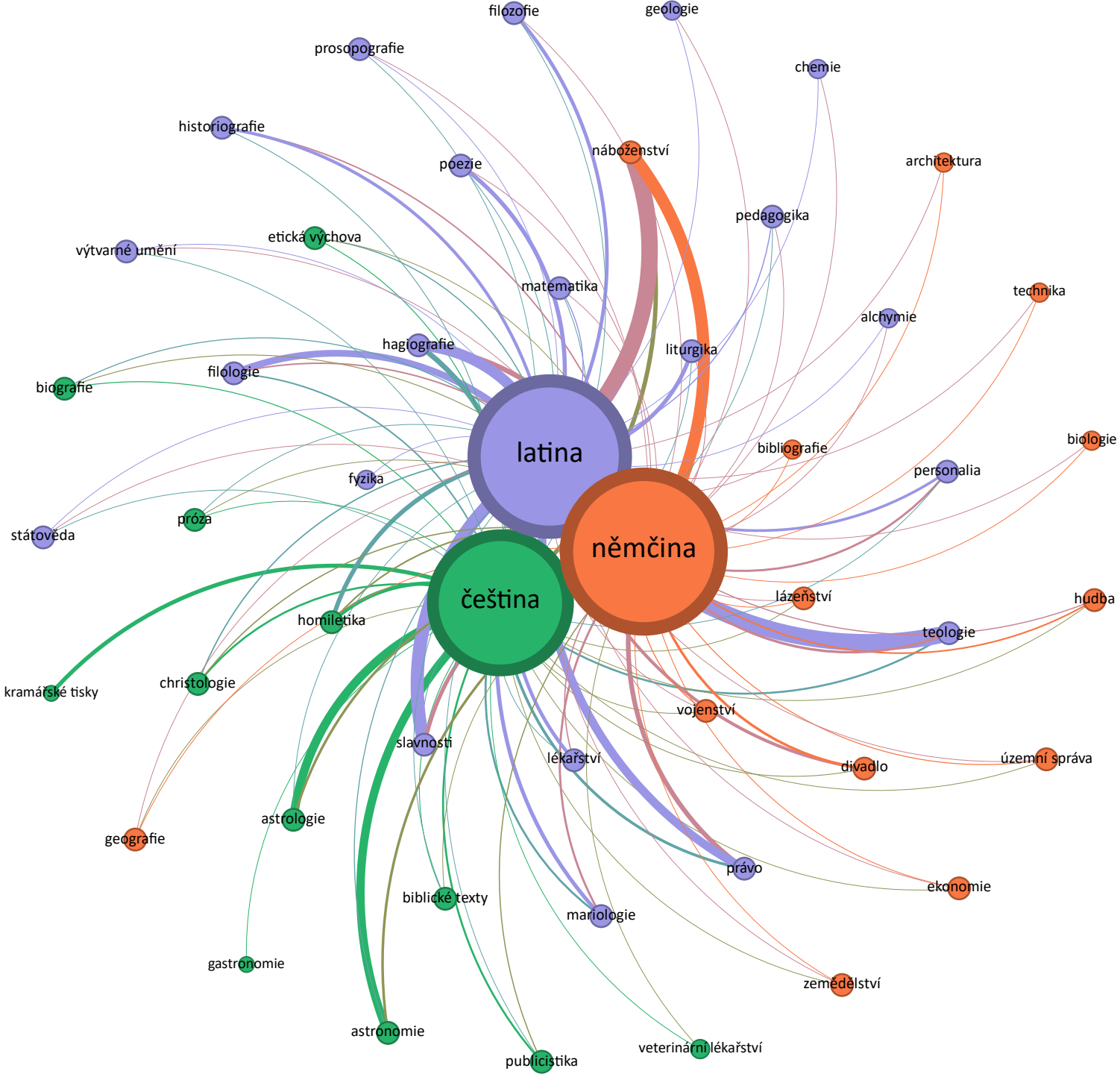
Example plots using the `classroom` data:

```
plot(
  classroom,
  layout=layout_with_fr,
  vertex.color="white",
  vertex.size=15,
  edge.arrow.size=0.5,
  vertex.label.color="black",
  vertex.label.family="sans",
  vertex.label=ifelse(V(classroom)$female, "F", "M")
)
```



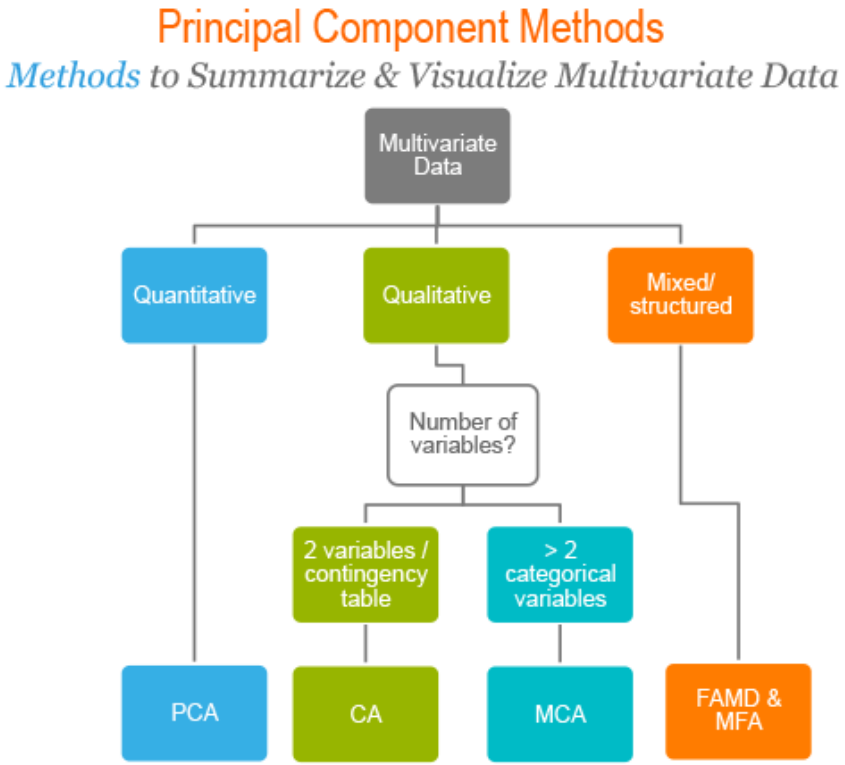

```
plot(
  classroom,
  layout=layout_with_fr,
  vertex.label=NA,
  vertex.size=scales::rescale(degree(classroom, mode="in"), c(5, 25)),
  edge.arrow.size=0.5,
  vertex.color=ifelse(V(classroom)$female, "pink", "lightskyblue")
)
```



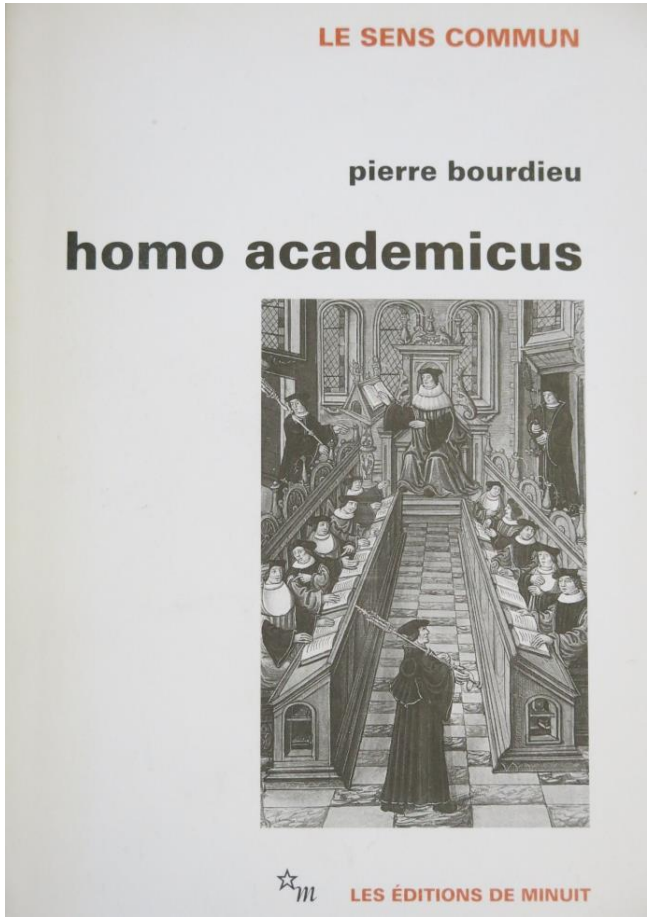

filozofie
geologie
prosopografie
chemie
historiografie
náboženství
poezie
architektura
pedagogika
etická výchova
výtvarné umění
matematika
technika
alchymie
hagiografie
liturgika
filologie
biografie
latina
bibliografie
biologie
fyzika
personalia
němčina
próza
státověda
čeština
lázeňství
hudba
homiletika
teologie
kramářské tisky
christologie
vojenství
slavnosti
lékařství
divadlo
územní správa
astrologie
geografie
právo
ekonomie
biblické texty
mariologie
gastronomie
zemědělství
astronomie
veterinární lékařství
publicistika

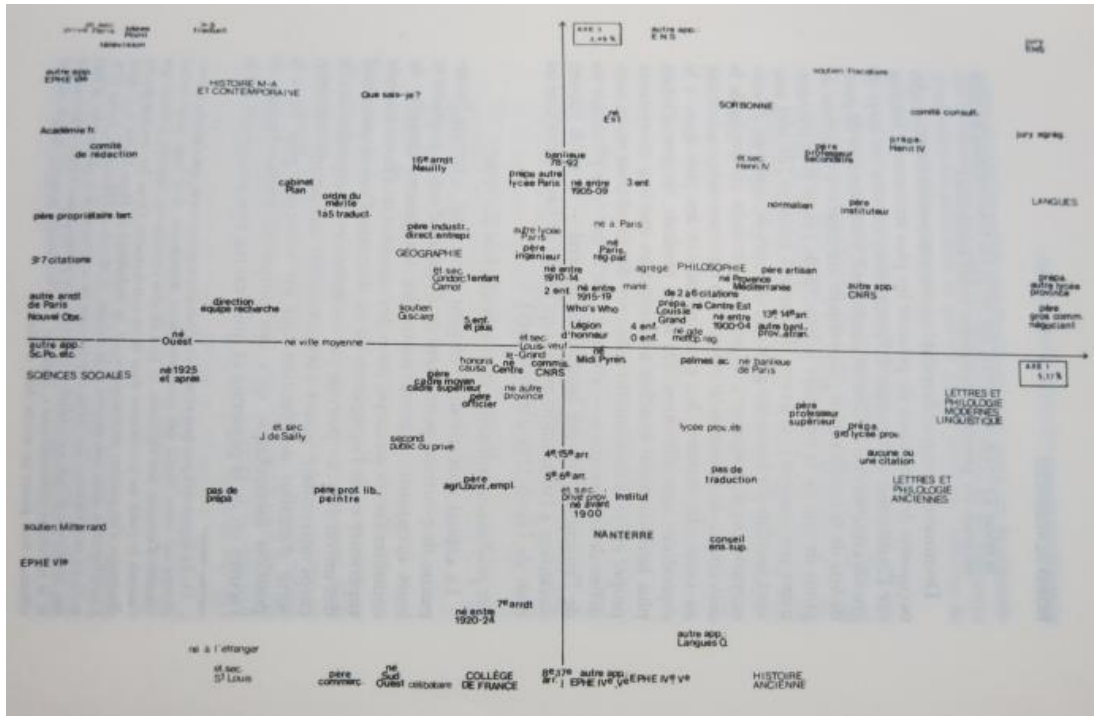
Korespondenční analýza

Analýza hlavních komponent



Korespondenční analýza



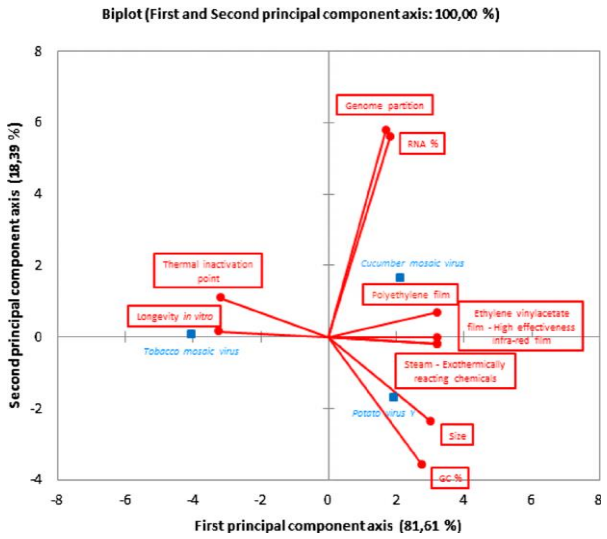


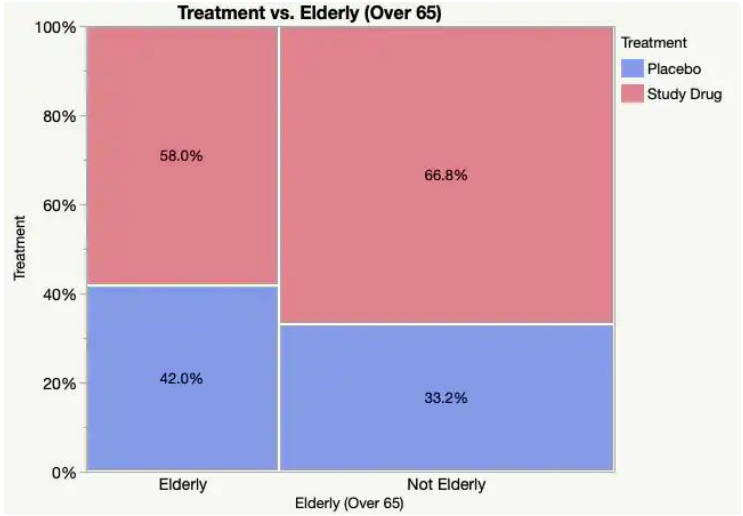
Balíčky v R pro korespondenční analýzu

- ca
- MASS
- FactoMineR, factoextra
- GDATools

Korespondenční analýza: vizualizace

- biplot
  - dvojí koordináty
  - způsoby výpočtu
  - symetrický, asymetrický
- mosaic display





Vícečetná korespondenční analýza (MCA)

jméno	automobil	oblíbená barva	mobil
František	Škoda	červená	Samsung
Josef	Škoda	červená	Apple
Karel	Volvo	zelená	Apple
Ludvík	Peugeot	bílá	Motorola

- metoda zpracování:
  - vícecestné tabulky
    - zdvojení (nausea_y, nausea_n)
  - indikátorová / Burtova matice

MCA

- explorativní vícerozměrná analýza
  - shluková analýza
  - analýza založená na redukci rozměrů na menší počet

- skupiny jedinců s podobným profilem
  - aktivní jedinci

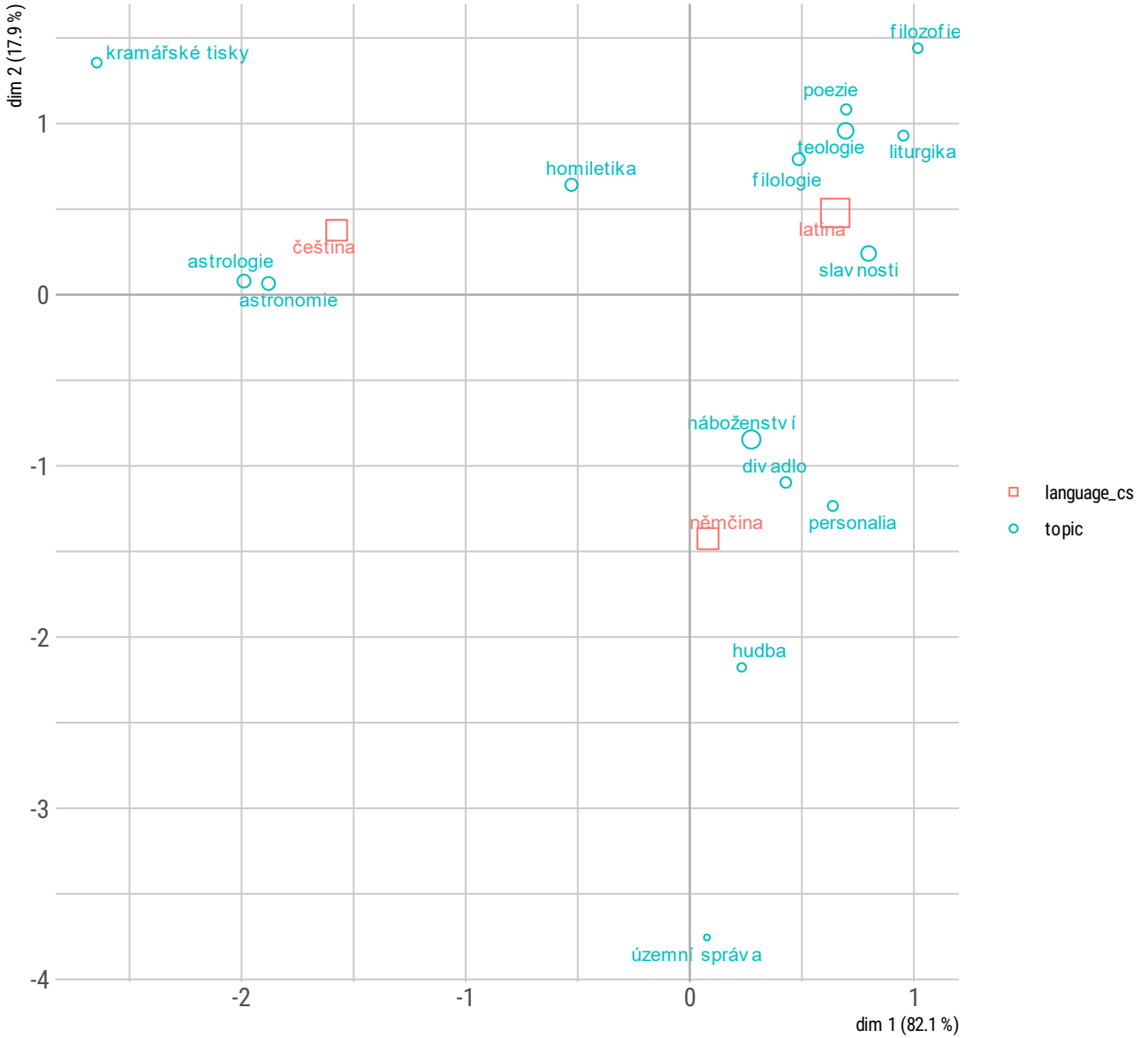
- vztahy mezi kategoriemi proměnných
  - aktivní proměnné
  - doplňkové proměnné: kvantitativní
  - doplňkové proměnné: kvalitativní

MCA

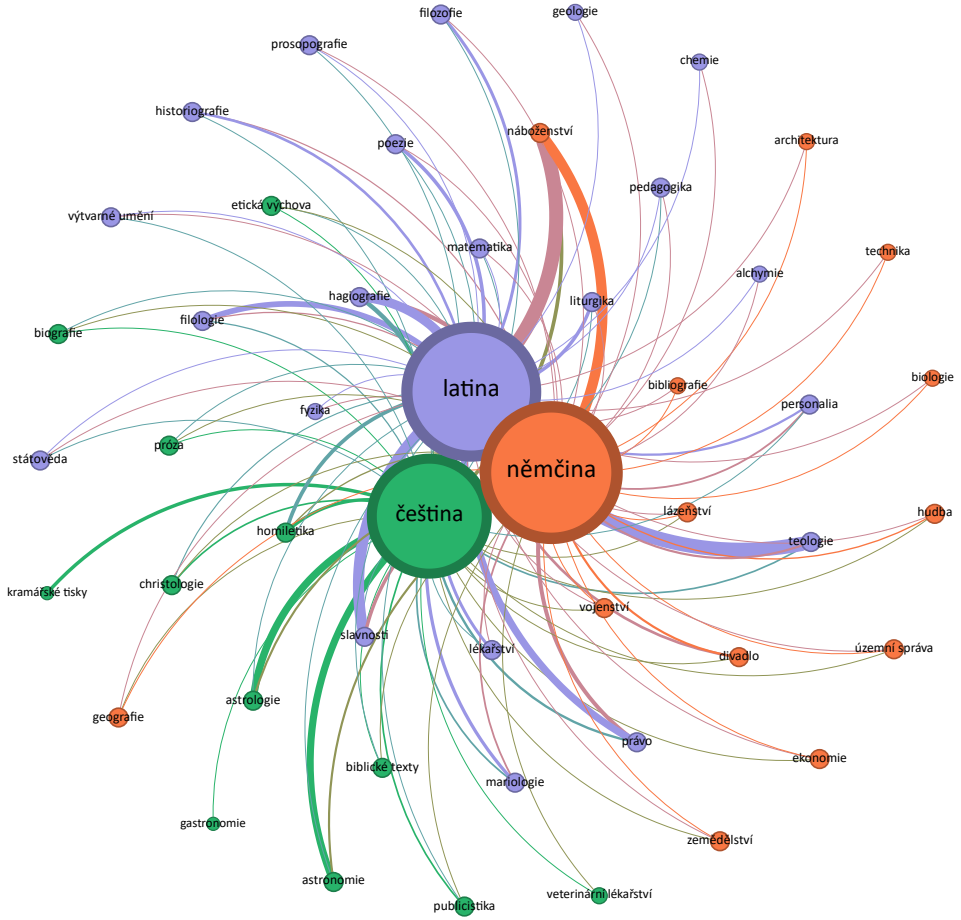
- cos2: kvalita reprezentace
- příspěvek k rozměrům v %

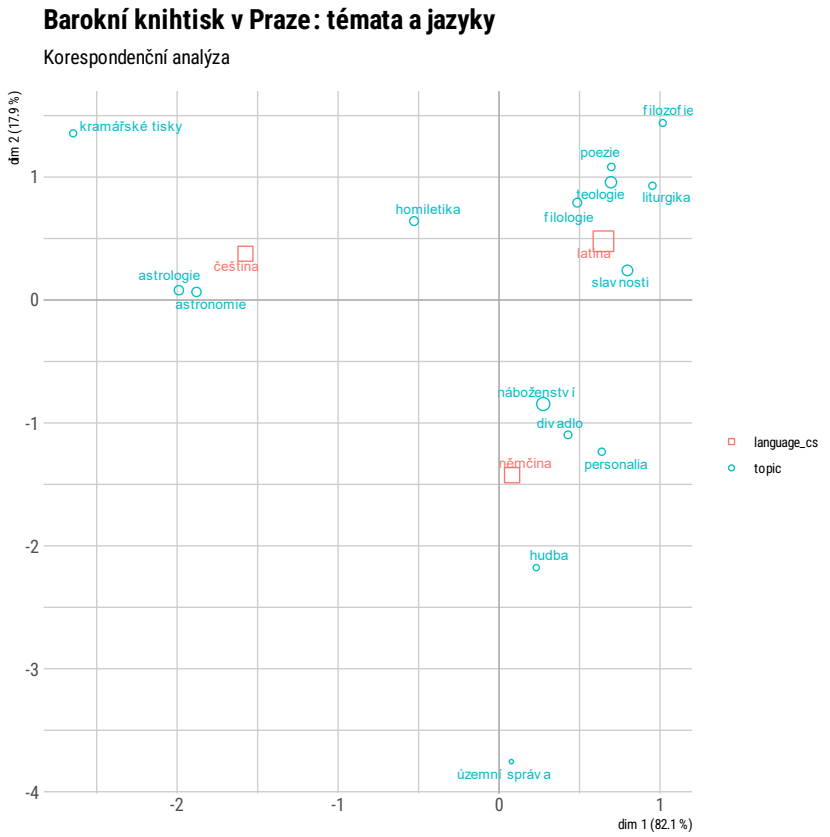
Barokní knihtisk v Praze: témata a jazyky

Korespondenční analýza



Analýza sociálních sítí vs. korespondenční





Analýza sociálních sítí vs. korespondenční

- co mají společného?

- de Nooy 2003: „sociogram je založen na stejných datech a byla použita optimalizační technika, která je založena na víceméně stejných principech jako singulární rozklad v korespondenční analýze“