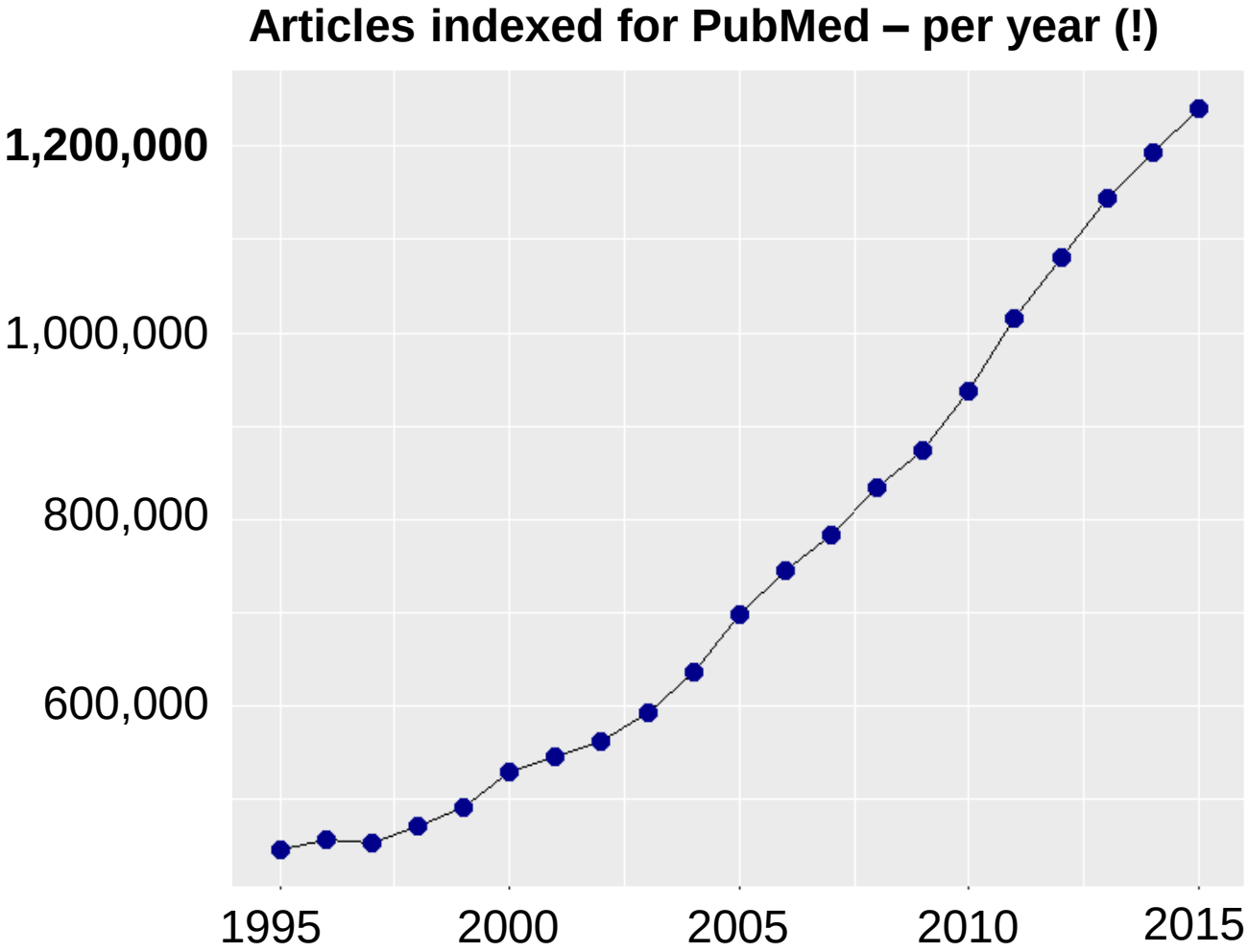
UNIKLINIK RWTH AACHEN
Institut für Medizinische Informatik

How Is My Field Evolving? – Network Based Analysis of Biomedical Scientific Discourse

C. Spreckelsen[1] K. Kabino[1]

[1] Dept. of Medical Informatics, RWTH Aachen University

Overwhelming Scientific Productivity



Spreckelsen C, Kabino K: How is my field evolving? – Network Based Analysis of Biomedical Scientific Discourse. eHealth 2016, Wien

Problem: How to keep track of scientific discourse?

Definition: Discourse

“a mode of organizing knowledge, ideas, or experiences that is rooted in language and its concrete contexts”

(Merriam-Webster: Dictionary [Internet] [Cited: 2016/02/08])

Goal

- Automatic, **data driven support** for discourse analysis

- **Requirements**:

…Identify **relevance of topics** based on **bibliometric** measures

…Analyze **temporal evolution** of topic relevance

…Identify similarly evolving topics (**trends**)

…**Visualize temporal patterns** of relevance

Related Work: Bibliometric Approaches

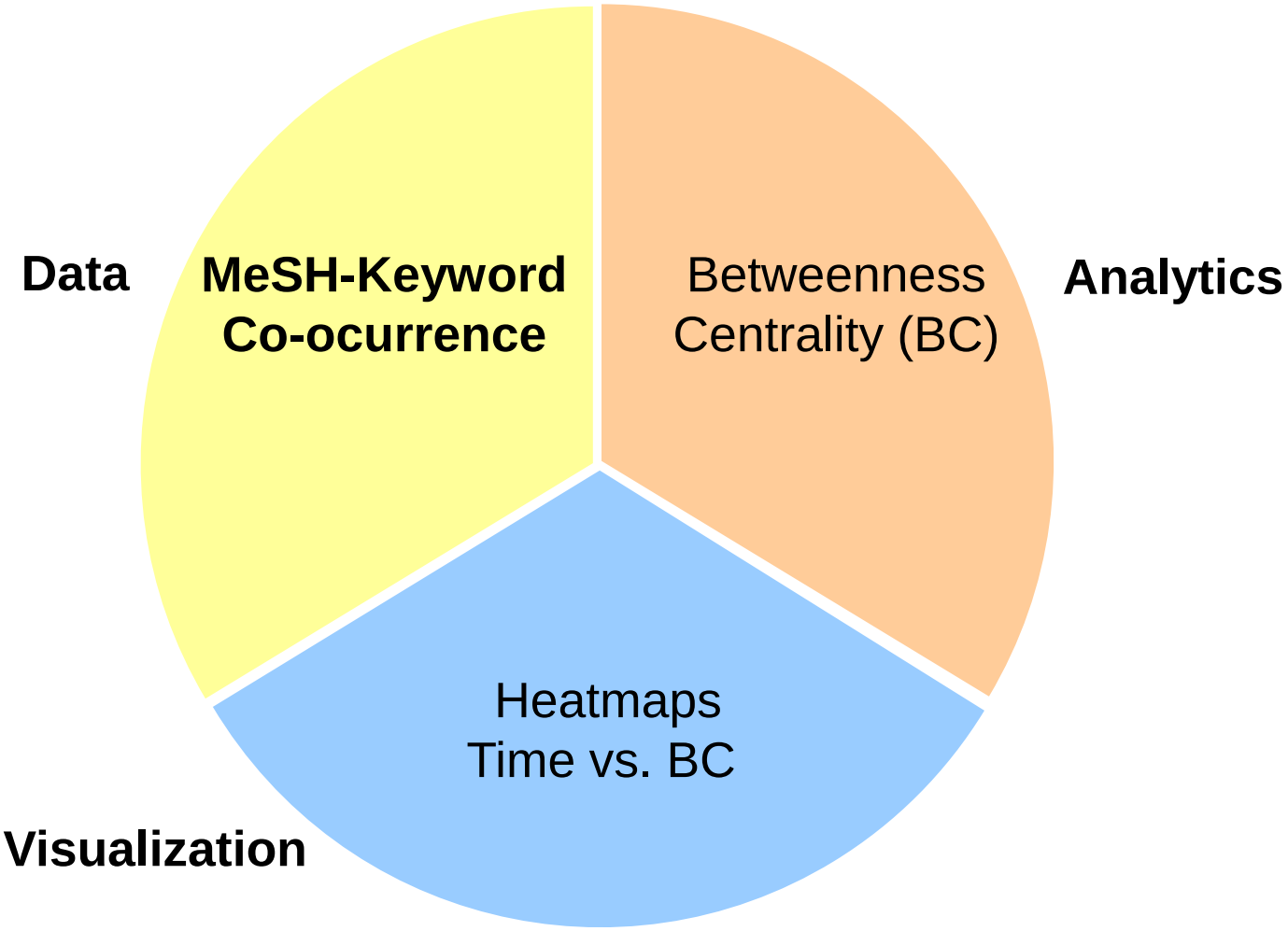
- Citation networks (e.g. citation impact)
- Collaboration diagrams (co-authorship)

- Topic identification based on term statistics
  - Latent Semantic Indexing (LSI)
  - Latent Dirichlet Allocation (LDA)
  - Author Topic Modelling (ATM)
  - Dynamic Topic Modelling (DTM)

Bibliographic and/or full-text data

Inference of thematic fields

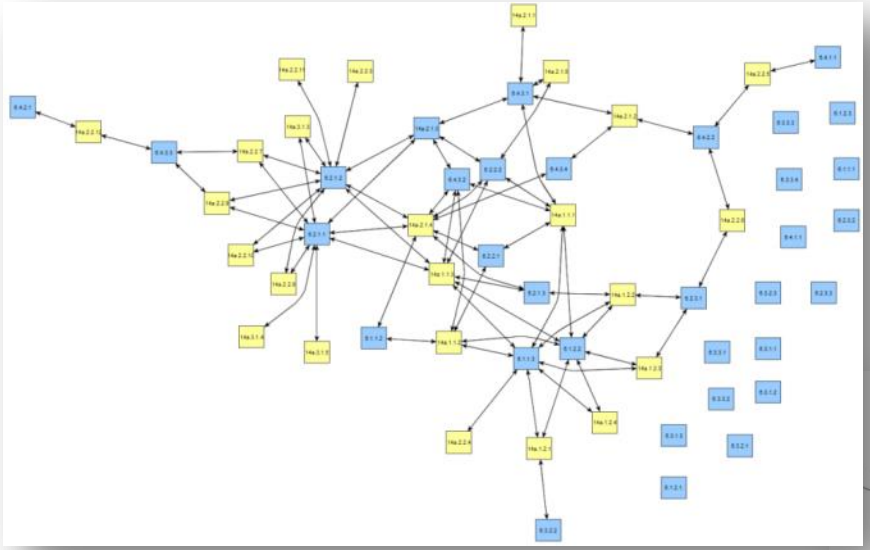
Approach: Bibliometric Analysis of Co-ocurrence Graphs



Spreckelsen C, Kabino K: How is my field evolving? – Network Based Analysis of Biomedical Scientific Discourse. eHealth 2016, Wien

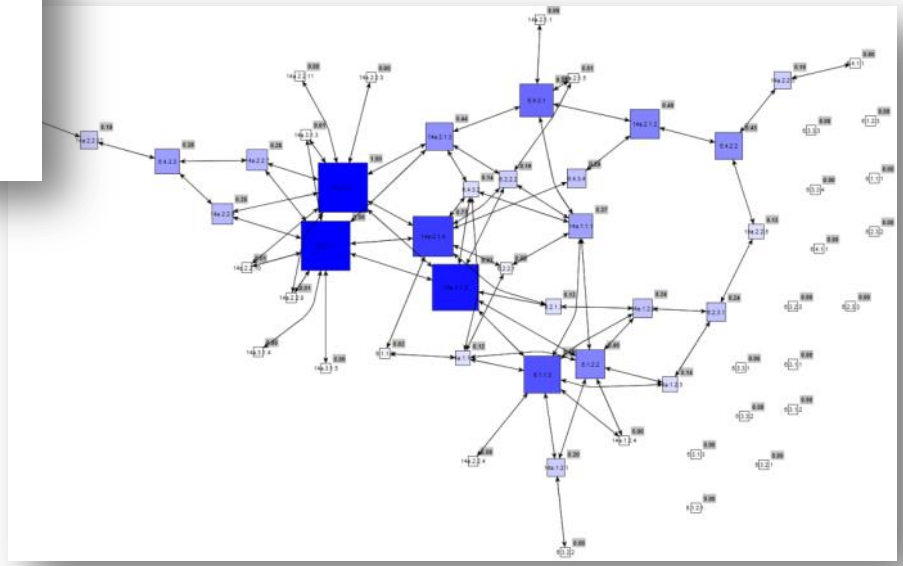
Network Based Analysis: Betweenness Centrality (BC)

Rationale: Keywords **connecting the field of discourse** considered **relevant**



Shortest paths between each pair of nodes

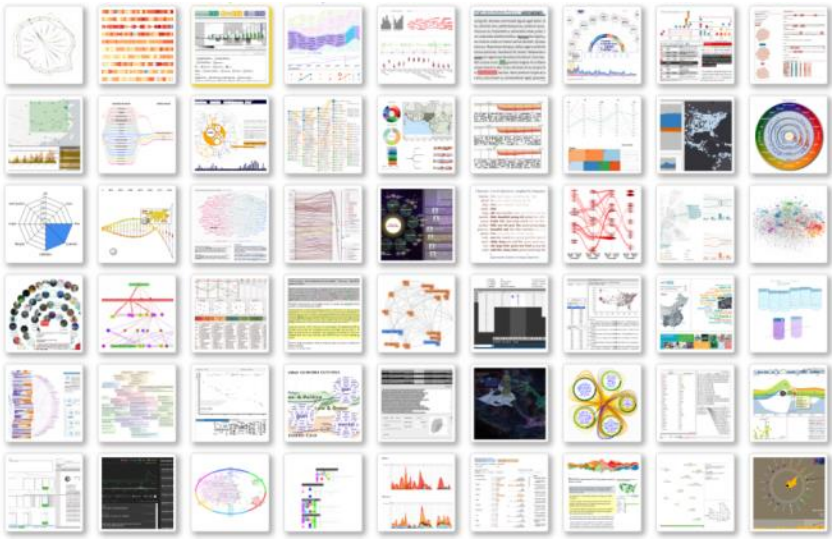
BC: Fraction of paths, which include evaluated node



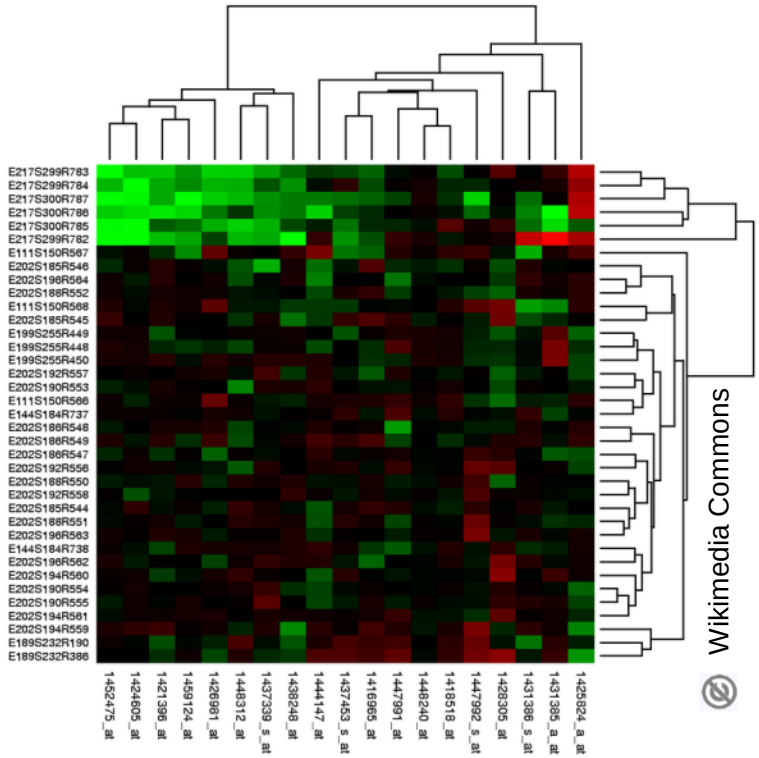
$$BC(v) := \sum_{i,j \in V(g), i \neq j \neq n} \frac{\sigma_{i,j}(v)}{\sigma_{i,j}}$$

Text Visualization Approaches

Overview e.g.: http://textvis.lnu.se/



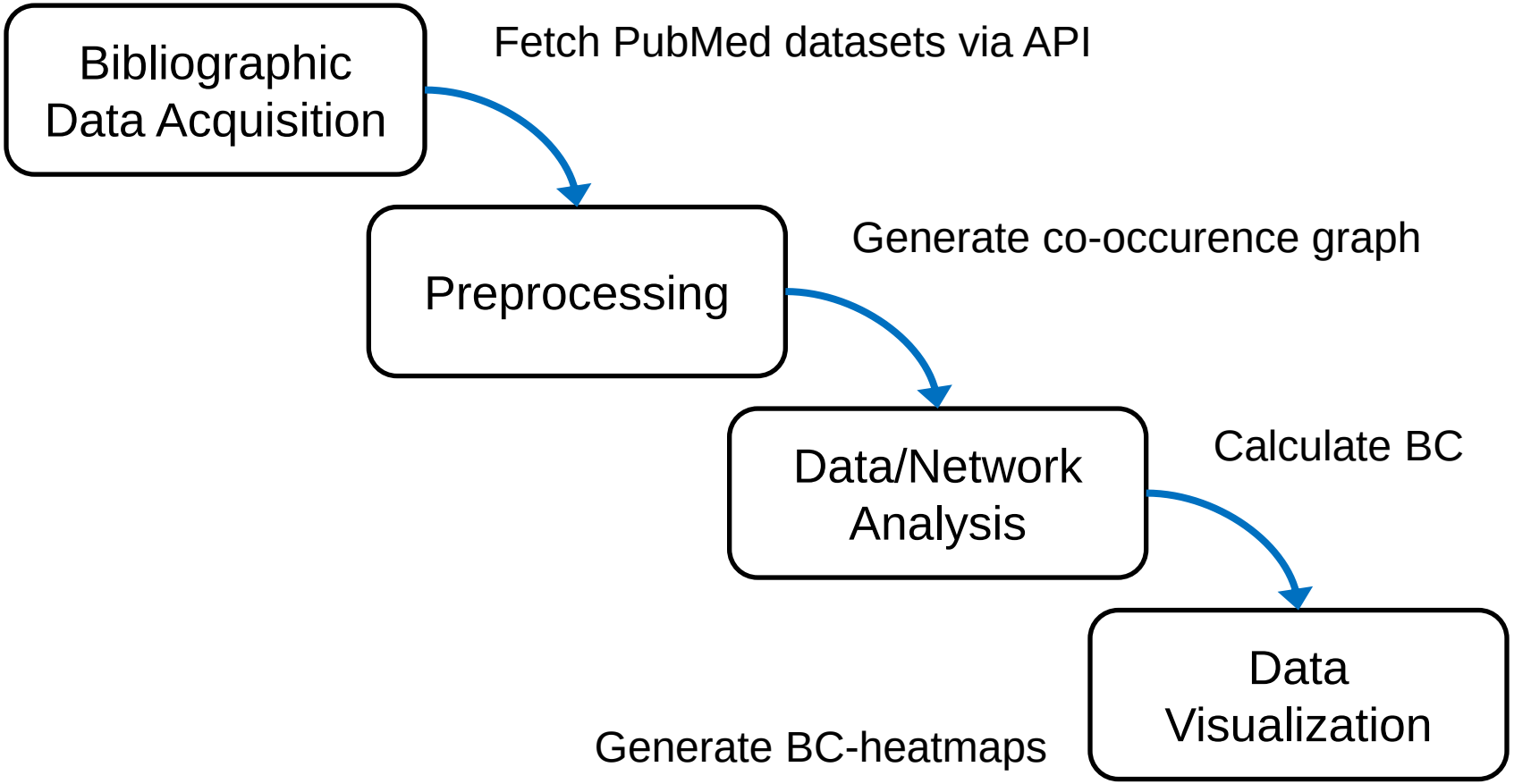
Kucher K, Kerren A. Text visualization techniques: Taxonomy, visual survey, and community insights. In: Visualization Symposium (PacificVis), 2015 IEEE Pacific 2015 Apr 14 (pp. 117-121). IEEE.



Here: **Heatmaps**

- Suitable for **extensive data exploration**
- **Well-understood** in biomedical domain
- Visualization of **quantitative** data

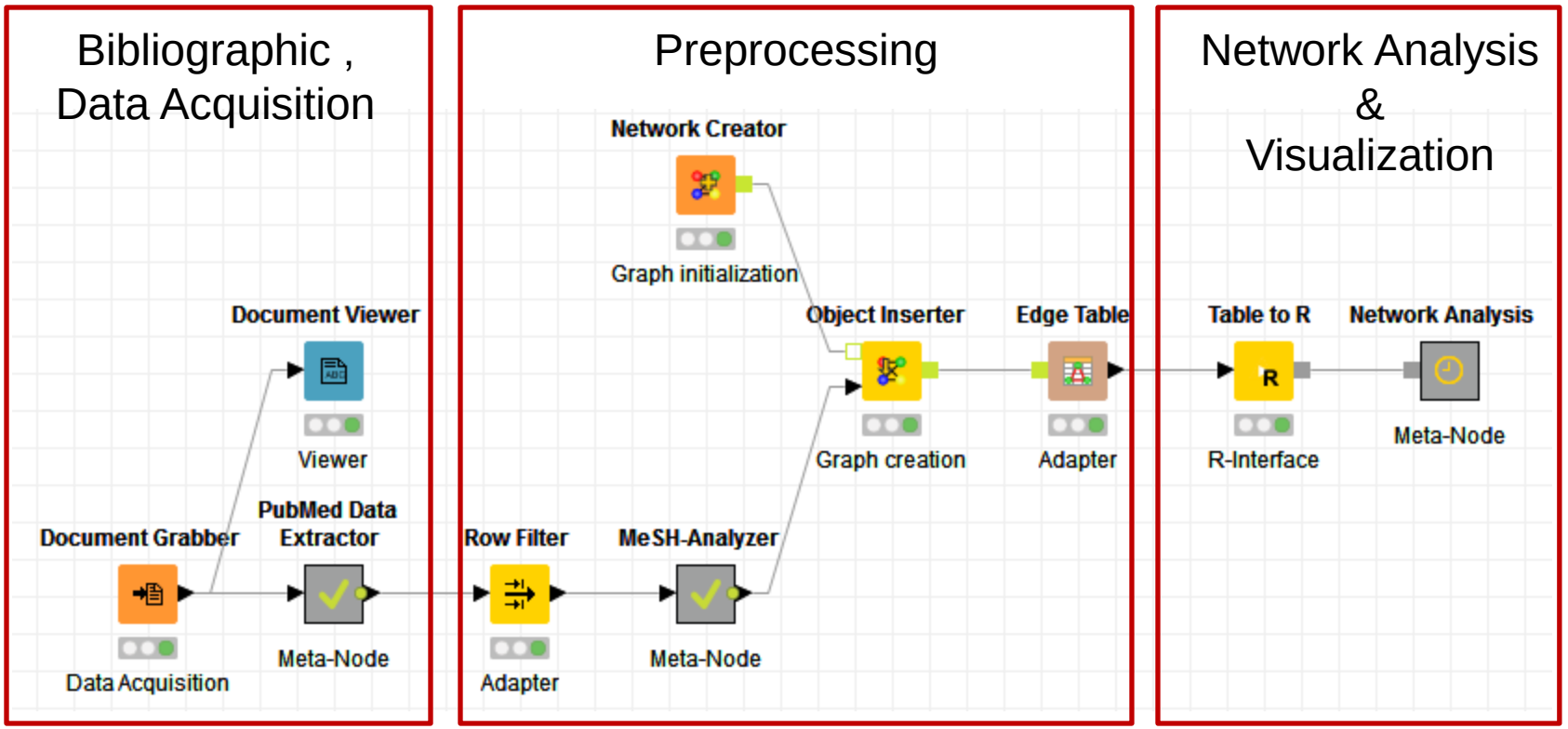
Processing steps of discourse analysis



Implementation: Processing pipeline

Data management: KNIME analytics platform (v3.1.0)

Analytics & Visualization: R Project for Statistical Computing



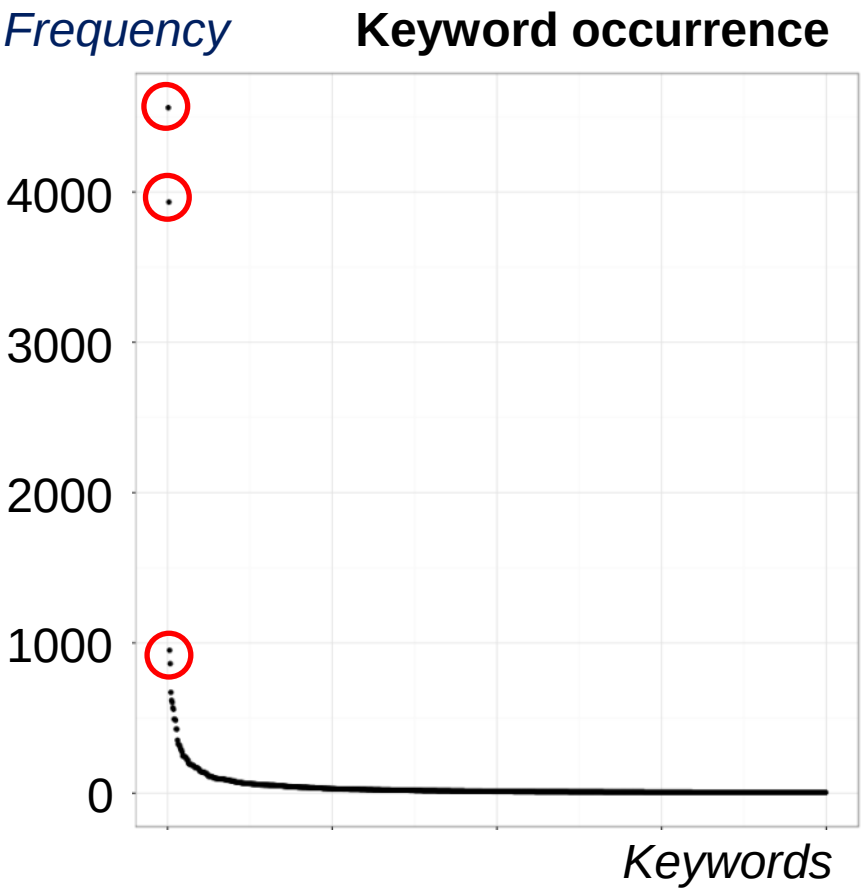
Case study

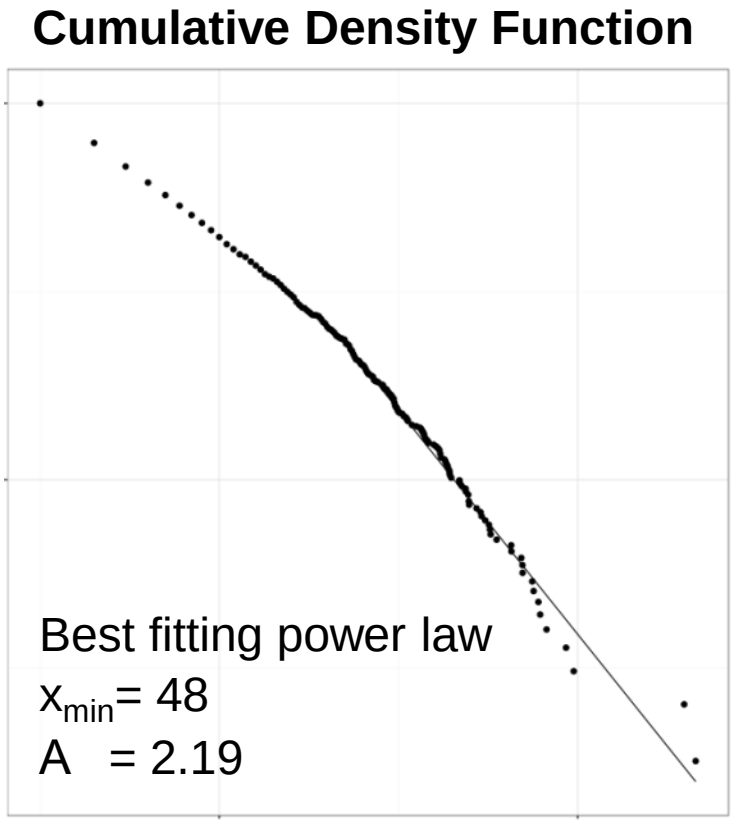
- **Motivation**: Renaissance of AI in medical decision making (?)

- **PubMed-Query**: „Decision Support Systems, Clinical [MeSH Terms]“
  - Filter: Publication date: 2000 to 2016

- **Retrieved** : **5,094** datasets
- **Generated** co-occurence graph: **174,663** inter-keyword links (weighted)

Power law of keyword occurence - revisited

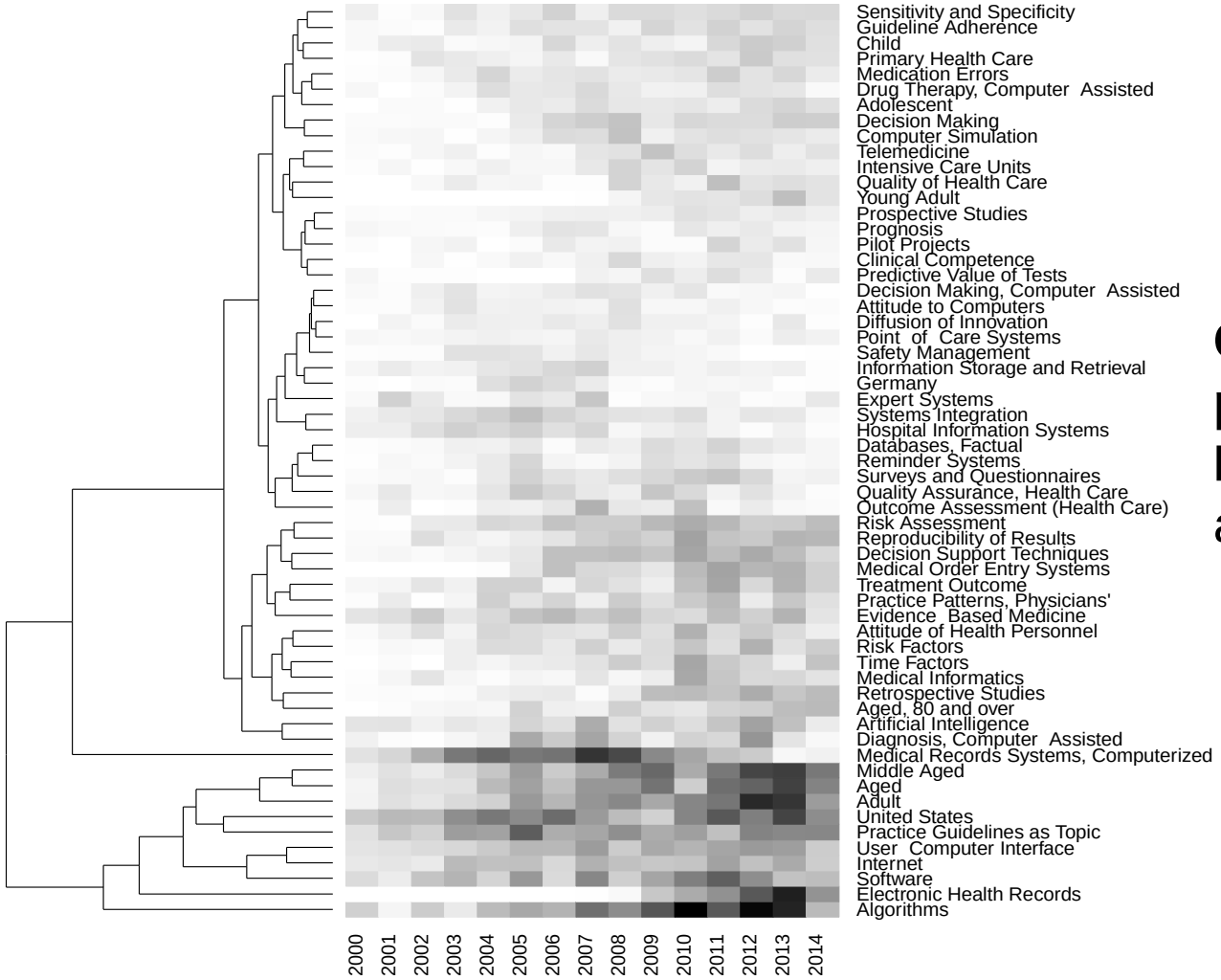




Goodness-of-fit test: **.037**

Power law plausible

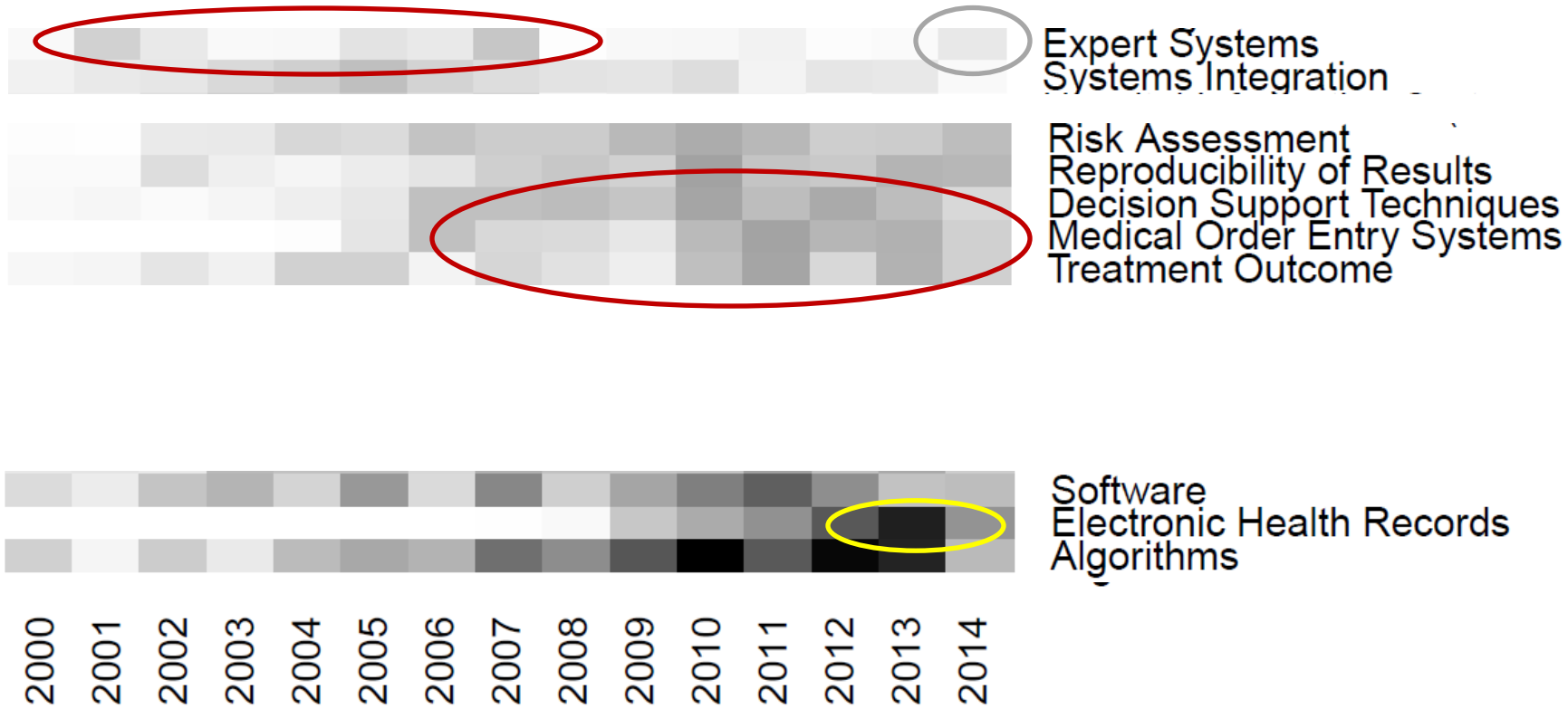
Results for „Decision Support Systems, Clinical“



Compensation of power law distribution: **Non-linear color-scale** applied **after** clustering

Spreckelsen C, Kabino K: How is my field evolving? – Network Based Analysis of Biomedical Scientific Discourse. eHealth 2016, Wien

Results for „Decision Support Systems, Clinical“



Spreckelsen C, Kabino K: How is my field evolving? – Network Based Analysis of Biomedical Scientific Discourse. eHealth 2016, Wien

Limitations

- Performance bottleneck: KNIME-to-R-integration
  - Pure R implementation (finished)
- BC as measure of relevance
  - Plausible, but not exclusive
- Use of MeSH-keywords instead of full text/abstract
  - Strength: Exploits precise semantic aggregation
  - Weakness: Reduced data base compared to text mining

Conclusion

- Data driven and network based discourse analysis feasible
- Efficient processing pipeline
- Reveals expected effects of known causes
- Visualization of trends

Spreckelsen C, Kabino K: How is my field evolving? – Network Based Analysis of Biomedical Scientific Discourse. eHealth 2016, Wien