Definition

The application of statistical methods to the study of bibliographic data.
Literature Review

1. Search for relevant literature
2. Evaluate sources
3. Identify themes, debates and gaps
4. Outline the structure
5. Write your literature review
<table>
<thead>
<tr>
<th>Sources</th>
<th>Authors</th>
<th>Documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most relevant</td>
<td>Most relevant</td>
<td>Most cited</td>
</tr>
<tr>
<td>Dynamics</td>
<td>Papers/Year</td>
<td>References</td>
</tr>
<tr>
<td>H-Index</td>
<td>H-Index</td>
<td>Words</td>
</tr>
<tr>
<td>Bradford's Law</td>
<td>Countries</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lotka's Law</td>
</tr>
</tbody>
</table>
Analyze Search Results

Documents per year by source

Compare the document counts for up to 10 sources.

Compare sources and view CiteScore, SJR, and SNIP data
Analyze Search Results

Documents by author

Documents by affiliation
Analyze Search Results

Still a paper-to-paper analysis
Citation analysis is the examination of the frequency, patterns, and graphs of citations in articles and books. It uses citations in scholarly works to establish links to other works or other researchers.

Tools

- MetaKnowledge in Python
- Bibliometrix in R
- VOSviewer
- CiteSpace
Bibliometric Tools

Highly developed and isolated (ltd. Importance)

Weakly developed and marginal

Well developed and important structures

General Topics transverse different research areas

Still a paper-to-paper analysis

- R’s Bibliometrix

- Highly developed and isolated (ltd. Importance)

- Weakly developed and marginal

- Well developed and important structures

- General Topics transverse different research areas

- Still a paper-to-paper analysis

- R’s Bibliometrix

Bibliometric Tools

Highly developed and isolated (ltd. Importance)

Weakly developed and marginal

Well developed and important structures

General Topics transverse different research areas

Still a paper-to-paper analysis

R’s Bibliometrix

Bibliometric Tools

Highly developed and isolated (ltd. Importance)

Weakly developed and marginal

Well developed and important structures

General Topics transverse different research areas

Still a paper-to-paper analysis

R’s Bibliometrix

Bibliometric Tools

Highly developed and isolated (ltd. Importance)

Weakly developed and marginal

Well developed and important structures

General Topics transverse different research areas

Still a paper-to-paper analysis

R’s Bibliometrix
<table>
<thead>
<tr>
<th><strong>Sources</strong></th>
<th><strong>Authors</strong></th>
<th><strong>Documents</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Most relevant</td>
<td>Most relevant</td>
<td>Most cited</td>
</tr>
<tr>
<td>Dynamics</td>
<td>Papers/Year</td>
<td>References</td>
</tr>
<tr>
<td>H-Index</td>
<td>H-Index</td>
<td>Words</td>
</tr>
<tr>
<td>Bradford's Law</td>
<td>Countries</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lotka's Law</td>
</tr>
</tbody>
</table>
### Bibliometric Search

**Focus on Relevance**

<table>
<thead>
<tr>
<th>Sources</th>
<th>Authors</th>
<th>Documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most relevant</td>
<td>Most relevant</td>
<td>Most cited</td>
</tr>
<tr>
<td>Dynamics</td>
<td>Papers/Year</td>
<td>References</td>
</tr>
<tr>
<td>H-Index</td>
<td>H-Index</td>
<td>Words</td>
</tr>
<tr>
<td>Bradford's Law</td>
<td>Countries</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lotka's Law</td>
<td></td>
</tr>
</tbody>
</table>

### Mapping the Science

<table>
<thead>
<tr>
<th>Structures in Lit</th>
<th>Bibliometric Technique</th>
<th>Unit</th>
<th>Statistical Technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concepts</td>
<td>Co-words</td>
<td>Keywords</td>
<td>Network Analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Author Keywords</td>
<td>PCA, MDS, MCA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Title</td>
<td>Theme Mapping</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract</td>
<td>Theme Evolution</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Full Text</td>
<td>Topic Modeling</td>
</tr>
<tr>
<td>Knowledge</td>
<td>Co-citation</td>
<td>References</td>
<td>Network Analysis</td>
</tr>
<tr>
<td></td>
<td>Citation</td>
<td>Authors</td>
<td>Historiographs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Journals</td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td>Collaboration</td>
<td>Authors</td>
<td>Collaboration Network</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Institution</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Journal</td>
<td></td>
</tr>
</tbody>
</table>
Workflow

Data acquisition

Citation Networks

Establish citation relationship

Creating citation network

Clustering

Citation network clusters
Profiles knowledge landscape by topical aggregation of academic articles

Semantic Analysis

Term retrieval and cleaning

Co-term network creation

Identification of core terms and topics

Hierarchy detection based on network properties

Hierarchical Topic Tree
Profiles knowledge landscape in a hierarchical term structure

Relation identification

Extract topic features

Time-based topic weighting

Evolutionary relationship identification

Scientific Evolutionary Pathways
Tracks evolutionary relationships among scientific topics over time
Natural Language Processing (NLP)

Bibliometric Results

Tokenization
Stop Word
Lemmatize
Word Clouds
Summary Stats

15
NLP -> Feature Engineering

- **Shallow Parsing**
  - Part-of-Speech Tagging

- **Named Entity Recognition**
  - identifying authors, journal names, affiliations, publication dates, and other key information

- **Word Sense Disambiguation**
  - ability to determine the meaning of a word by its context
NLP -> Topic Extraction and Modeling

Extraction
- Identify main topics in a document

Algorithms
- Latent Dirichlet Algorithm (LDA)
- Large Language Models (LLMs)

Modeling
- recognizes topics by detecting patterns like clustering algorithms
  - TF-IDF
  - BERT
  - t-SNE
Network Science

- Bibliometrics uses some form, fit for their own purpose.
- Degree distribution
- Size
- Density
- Components
- Centrality
- Communities

Focuses on the study of patterns of connection in a wide range of physical and social phenomena.

Descriptive Statistics to describe network structure.
Network Characteristics

A. Centralities

- Highest degree centrality
- Highest betweenness centrality
- Highest closeness centrality
- Highest eigenvector centrality
Discussion

- What metadata fields would enhance the researcher’s experience with bibliometrics?
- What are some ways you can imagine improving the literature review process?
- What games might you play with the metadata to gain access to the information you need?
- Semantic shifts
- Emerging Research
- Topics
Thank you