Social Web and Bibliometrics

University of Koblenz-Landau,
SS 2013

Philipp Mayr
About me (Philipp)

I am a team leader at GESIS – Leibniz Institute for Social Sciences. Department „Knowledge technologies for the Social Sciences“.

My team „Portals and Value-Added Services“: sowiport / SSOAR – Information retrieval evaluation, information visualization, Digital Libraries

DFG-Project: Value-added services of information retrieval (IRM)

http://www.gesis.org/irm
Acknowledgements

Many slides are based on the lecture „Bibliometrics“ by Tefko Saracevic.

Please find out more about Tefko at:
http://comminfo.rutgers.edu/~tefko/
Aim of this topic:

- How do researchers publish and how can you measure joint research activities?
- Understanding of bibliometrics and its regularities
- Similarity of social web and bibliometric regularities
What is?

“… all studies which seek to quantify processes of written communication.”

Pritchard

“… the quantitative treatment of the properties of recorded discourse and behavior pertaining to it.”

Fairthorne

Recorded communication - ‘literature’->
quantitative methods
Alan Pritchard 1969

Coined the term "bibliometrics"
"the application of mathematics and statistical methods to books and other media of communication“

and other related metrics …

Also used to study broader than books, articles …

– Scientometrics
  • covering science in general, not just publications

– Infometrics
  • all information objects

– Web(o)metrics or cybermetrics
  • web connections, manifestations
  • using bibliometric techniques to study the relationship or properties of different sites on the web
Concepts

Basic (primitive) concepts:

1. **Subject** (research topics, ...)
2. **Recorded communication** -> document, information object
3. **Subject literature**

Bibliometrics related to:

- **science of science**
- sociology of science - numerical methods
Literature studies

Qualitative
  – often in humanities, librarianship

Quantitative
  – bibliometrics

Mixed approaches
Reasons for quantitative studies of literature

Analysis of structure and dynamics
  – search for **regularities** - **predictions** possible

Understanding of patterns
  – “order out of documentary chaos”
  – verification of models, assumptions

Rationale for policies & design
Why quantitative studies?

Qualitative methods often depend on assertions. - ‘authoritative’ statements, anecdotal evidence

Science searches for regularities

Success of statistical methods in social sciences

Need for justification & basis for decisions
- Something can be counted - irresistible
Application in ...

History of science
Sociology of science
Science policy; resource allocation
Library selection, weeding, policies
Information organization & management
Information retrieval (e.g. Bradfordizing)

- See IRM project  [www.gesis.org/irm](http://www.gesis.org/irm)
Historical note

Bibliometrics long precedes information science
But found **intellectual home in information science**
  – study of a basic phenomenon - literature
It is not ‘hot’ lately, but **still produces very interesting results**
**Branched out into web studies** (web is a “literature” as well)
  - e.g. -> this lecture
What studied?

Governed by **data available in documents or information resources in general** - that what can be counted

- author(s)
- origin
  - organization, country, language
- source
  - journal, publisher, patent …
what ... more

– contents
  • text, parts of text, subject, classes
– representation
– citations
  • to a document, in a document, co-citation
– utilization
  • circulation, various uses
– links
– any other quantifiable attribute
Tools

Science Citation Index (SSCI, AHI) and Scopus
Compilation of variables from journals in a subject
Use data
Publication counts from indexes, or other data bases
Web structures, links
New tool: Google Scholar
Variable: authors

- number in a subject, field, institution, country
- growth
- correlation with indicators like GNP, energy etc.
- productivity e.g. Lotka’s law
- collaboration - co-authorship, associated networks
- dynamics - productive life, transscience, epidemics
- papers/author in a subject
- mapping
Service: Reranking (Author centrality)

IRM-project: Ranking by Author Centrality: sorting central author papers on top (co-authorship analysis)

Koblenz, Summer 2013
Variable: origin

Rates of production, size, growth by
- country, institution, language, subject

Comparison between these
Correlation with economic & other indicators
Variable: sources

Concentration most often on journals

Growth, dynamics, numbers
  – information explosion - exponential laws
  – time movements, life cycles

Scatter - quantity/yield distribution
  – Bradford’s law

Various distributions
  – by subject, language, country
Service: Reranking (Bradfordizing)

IRM-project: Ranking core journals on top

Information Retrieval Value-added Services

Query: luhrmann AND collection:iz*

Only show metadata sets which include an abstract (slower)

Total hits: 687

   Zeitschrift für soziologische Theorie Jg. 2 H. 1 (0948-423X) toggle details
   Zeitschrift für soziologische Theorie Jg. 2 H. 1 (0948-423X) toggle details
3. Esser, Hartmut; Luhmann, Niklas (1996): *Individualismus und Systemdenken*

Check out the IRM project >>

- Soziale Systeme (0948-423X) [132]
- Zeitschrift für Soziologie (0340-1804) [68]
- Soziale Welt (0038-6073) [34]
- Kölner Zeitschrift für Soziologie und Sozialpsychologie (0023-2653) [29]
- Zeitschrift für Rechtsoziologie (0174-0202) [28]
- Sociologia internationalis (0038-0164) [21]
- Archiv für Rechts- und Sozialphilosophie
Variable: contents

Analysis of texts

– distribution of words – Zipf’s law
– words, phrases in various parts
– subject analysis, classification
– co-word analysis
Variable: representation

– frequency of use of index terms, classes
– distribution laws - key terms where?
– thesaurus structure
Service: Search Term Recommender

IRM-project
- Recommendation of strongly associated terms
- Automatic query expansion

Information Retrieval Value-added Services

Query: luhmann

Only show metadata sets which include an abstract (slower)

Suggest search terms

Controlled vocabulary
Thesaurus Sozialwissenschaften
Automatic query expansion

Rerank the result list

Rerank method
Default relevance ranking

Total hits: 3599

1. Luhmann, Niklas (2002): Einführung in die Systemtheorie (3-89670-292-0) toggle details
Variable: citations

Studied a lot; many pragmatic results
  – base for citation indexes, web of science, impact factors, co-citation studies etc

Derived:
  – number of references in articles
  – number of citations to articles
    • research front; citation classics
  – bibliographic coupling
Co-citation analysis

Articles that cite the same article are likely to both be of interest to the reader of the cited article.

These two articles are likely to be related.
citations … more

– co-citations
  • author connections, subject structure, networks, maps
– centrality
  • of authors, papers
– validation with qualitative methods
– impact
Variable: utilization

– frequency
– distribution of requests for sources, titles
  • e.g. 20/80 law
– relevance judgement distributions
– circulation patterns
– use patterns
Variable: links

Development of link-based metrics
  – in-links, out-links

Web structure
Web page depth; update
PageRank vs quality
Examples from classic studies

- Comparative publications over centuries
- Number of journals founded over time
- Number of papers published over time
- National share of papers in chemistry
- National scientific size vs. economy size
- Bibliographic coupling and co-citation
- Web structures, links
- Bibliometric indicators (e.g. h-index)
- ...

Koblenz, Summer 2013
Exercise

– Web of Science or Google Scholar
  • Download 1000 records
– Parscit Demo
  • Analyzing a pdf with references
Sources for bibliometric analyses

Google Scholar [http://scholar.google.com](http://scholar.google.com)
Publish or perish [http://www.harzing.com/pop.htm](http://www.harzing.com/pop.htm)
Citespace [http://cluster.cis.drexel.edu/~cchen/citespace/](http://cluster.cis.drexel.edu/~cchen/citespace/)
Any further questions?