Running with the Devil:
On the Use and the Misuse of Bibliometrics and
How Social Science Research can Profit from
Bibliometrics

Michael Ochsner, ETH Zurich & FORS
Research Methods Seminar University of Lausanne, 3. November 2015
Outlook

- Introduction to bibliometric indicators
- Bibliometrics and the SSH
- Quality of music: A non-(or un-)scientific example for what bibliometrics cannot do
- Light at the end of the tunnel: What bibliometrics can do
Introduction to Bibliometrics
What is Bibliometrics?

- Term coined by Pritchard (1969), when he defined bibliometrics as:
  - „… the application of mathematical and statistical methods to books and other media of communication.“ (Pritchard, 1969)

- Today, bibliometrics is:
  - The analysis of the number of publications (all kinds of document types) and/or citations using also other bibliographic (meta-)data

- Use of bibliometric studies by:
  - Bibliometricians (basic research)
  - Disciplines/scholars (strategic and informational uses)
  - Science administrators (research evaluation, research policy)
Bibliometrics vs Scientometrics vs. Altmetrics

- No exact distinction (Glänzel, 2003: synonyms)
- Bibliometrics (Pritchard, 1969):
  - Citation analysis, analysis of document types, author networks etc.
  - What can be done with meta data from publication data bases
- Scientometrics (Nalimov & Mulchenko, 1969):
  - Analysis of communication and research processes
  - Includes bibliometrics but makes use also of other data on scholarly work: prizes, presentations, curricula etc.
- Altmetrics:
  - Bibliometrics with Web 2.0 data: analysis of Twitter feeds, Mendeley, Research Gate, download statistics from journals etc.
Godfathers of bibliometrics

- Candolle describes the scientific strength of nations in 1873 according to memberships in associations
- Lotka (1926) described the frequency distributions of publications → Lotka’s Law
- Gross & Gross (1927) aimed to identify the most important journals in their field (chemistry) by counting the citations
- Bradford (1934) described the frequency distributions of papers in journals → Bradford’s Law
- Zipf (1935) studied the frequencies of words in papers and generalised Lotka’s and Bradford’s Laws → Zipf’s Law → not really a bibliometrician but a linguist
Roots of modern bibliometrics (1/2)

- Modern bibliometrics started with Eugene Garfield
  * 16.09.1925, New York
- Now Thomson Reuters
- 1976: Publication of the *Journal Citation Reports* (JCR) including the *Journal Impact Factor* (JIF)
  (note that the JIF was invented by Martyn and Gilchrist (1968)
Roots of modern bibliometrics (2/2)

- Derek J. de Solla Price
  22.01.1922, Leyton, UK – 3.9.1983
- “Little science, big science” (1963)
- Exponential growth of science (personnel and publications) Price’s Law
- From small scale (erudite scholar) to large scale (huge teams, government or private funding of infrastructure)
Data availability as a driver: growth
Lotka’s Law:
- The number of authors writing x papers is about $1/n^a$ of those with one paper ($a=2$ in most disciplines)
- About 75% write only one article, 25% two or more, 4% five or more

Bradford’s Law:
- The number of relevant articles diminishes exponentially
- $1:n:n^2 \rightarrow$ 10 core journals in subject will cover 12 articles. For the next 12 article, 20 journals have to be searched, then 40 etc.

Zipf’s Law:
- The frequency of any citation/item is inversely related to its rank in the frequency table
- If most cited document has 30 citations, the second most cited has 15, the third most has 10 etc.
Bibliometric indicators: JIF (1/2)

Journal Impact Factor (JIF) in the Web of Science (WoS):

- JIF = \( \frac{\text{Number of citations in year } t \text{ of the documents published in the journal from the years } t-1 \text{ und } t-2)}{\text{number of „Citable Items“ in the years } t-1 \text{ and } t-2} \)
- Documents = article, review, proceeding paper, notes, letter, editorial note, etc.
- Citable Item = article, review, proceeding paper
- Advantage: Librarians can select most used journals
- Critique:
  - Documents vs. citable items: artificial pushing of JIF possible
  - Short citation window of 2 years; SSH 5-year JIF, but still short
  - Ecological fallacy if used as an indicator for an article
  - Not comparable across disciplines: different citation practices
Bibliometric indicators: h-index (1/2)

h-Index (Hirsch, 2005)

“A scientist has index h if h of his/her NP papers have at least h citations each, and the other (NP – h) papers have fewer than h citations each.”
Bibliometric indicators: h-index (2/2)

- **Advantages:**
  - Individual research performance (though possible on all levels)
  - Combining number of publications AND number of citations
  - Quite robust measure for “excellence” (high scores)

- **Disadvantages/Critique:**
  - Preference for researchers with long careers (cannot diminish)
  - Not comparable across disciplines: different citation practices
  - Poor discrimination (same h-index but different distribution)
Three commercial players
- Thomson Reuters, Elsevier and Google
- Plus: Web 2.0 and scholarly databases (PubMed or APA)

Different types
- ISI: Idea of the “core” journals: Only most important journals that cover about 80% of all scholarly output and most of the citations
- Web of Science and Scopus try to keep this ideal
- Google: anything but asymptotically everything
- Research Gate, Mendeley etc.: Only uploaded, no quality control, completely arbitrary
Different data base, different results

- My citations and h-index by Thomson Reuters
  - 3 Documents, 5 citations, h-index=2

This graph shows the number of times the articles on the publication list have been cited in each of the last 20 years. Note: Only articles from Web of Science Core Collection with citation data are included in the calculations. More information about these data.
Different data base, different results

- My citations and h-index by Scopus
  - 3 documents, 10 citations, h-index=2

Ochsner, Michael
Universität Lausanne Schweiz, Lausanne, Switzerland
Author ID: 55027245960

Documents: 3
- Citations: 10 total citations by 7 documents
- h-index: 2
- Co-authors: 2
- Subject area: Social Sciences

3 Documents | Cited by 7 documents | 2 co-authors

3 documents View in search results format
Export all | Add to my list | Set document alert | Set document feed

Voraussetzungen für die Beurteilung der Qualität wissenschaftlicher Forschung: Zusammenführung der Befunde aus vier empirischen Studien | Ochsner, M., Hug, S.E., Daniel, H.-D. | 2014 | Zeitschrift für Erziehungswissenschaft | 1


Four types of research in the humanities: Setting the stage for research quality criteria in the humanities | Ochsner, M., Hug, S.E., Daniel, H.-D. | 2013 | Research Evaluation | 6
Different data base, different results

- My citations and h-index by Google Scholar
  - 9 documents, 53 citations, h-index=5

<table>
<thead>
<tr>
<th>Title</th>
<th>Zitated from</th>
<th>Jahr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criteria for assessing research quality in the humanities: a Delphi study among scholars of English literature, German literature and art history</td>
<td>13</td>
<td>2013</td>
</tr>
<tr>
<td>Four types of research in the humanities: Setting the stage for research quality criteria in the humanities</td>
<td>13</td>
<td>2013</td>
</tr>
</tbody>
</table>
Different data base, different results

- My citations and RG-Score by ResearchGate
  - 10 documents, 14 citations
  - 4.77 RG-Score
  - 282 Reads, 2.68 impact points
Different data base, different results

Top researcher: Lutz Bornmann (1% SocSci)
- WoS: 199 documents, 2793 citations, h-index=27
- Scopus: n/a
- Google: 304 documents, 6487 citations, h-index=41
- ResearchGate: 246 documents, 3769 citations, RG=43.14, 10K reads, 659.2 impact points
Different data base, different results

- Impact Factors by data base:
  - WoS
    - Research Evaluation: 1.123
    - Zeitschrift für Erziehungswissenschaft: 0.299
  - Scopus
    - Research Evaluation: 1.103; SJR: 0.605
    - Zeitschrift für Erziehungswissenschaft: 0.436; SJR: 0.28
  - Research Gate
    - Research Evaluation: 0.85
    - Zeitschrift für Erziehungswissenschaft: 0.99
Different data base, different results

- **WoS**: Not all citations are detected:
  - Scopus: Misspelled or wrong citation = not captured
  - Google: Citation from anywhere, duplicates (same article = 3 items → citations divided)
  - See Sjögarde, 2014: Especially for non-hard science (citation styles?)
Differences also in interpretation

- Selective vs. comprehensive
  - WoS and Scopus: all citations are from scholarly articles published in peer reviewed journals, in the natural sciences also the most used and most important journals (not so in the SSH)
  - Google: citation from anything like articles in peer reviewed journals to presentations or student’s qualifying works

- Quality
  - WoS and Scopus: clear rules for inclusion, quality control
  - Google: algorithms and crawlers, many duplicates, easy to manipulate
  - ResearchGate and others: Uploads by users, no quality controls, same article twice, missing authors etc.
Bibliometric analyses

- Poor Man’s Bibliometrics (Marx & Bornmann, 2013)
  - Direct use of indicators of data bases (worst case: Google)
  - Possible because of expansion of availability of data
  - Simple indicators: JIF, h-index, SNIP, i10
    - If data is not cleaned: they are very biased
    - If used on wrong aggregation level: they are very biased
    - If used across different fields: they are very biased

- Professional Bibliometrics
  - Acquired relevant data, enormous time to clean
  - Sophisticated field normalization
  - Inclusion of non-source items for some fields (SSH)
  - Large numbers of indicators
  - Professional interpretation of results
Altmetrics: A hype (but a sarcastic joke)

- Use of Twitter, Facebook, Google+, Mendeley, Research Gate
- For 1.3 million indexed papers in WoS:
  - Only Twitter is significantly correlated with citation, therefore, only Twitter usable (Costas, Haustein, Larivière, 2014)
  - However: N=1,300,000; r=0.195
- Also known: 15% of scholars use Twitter, about 30% of Tweets are automated → extreme selection bias
- Easy to manipulate → researchers as professional Tweeters?
- What do we want to measure? “We are ready to do professional altmetrics, we only have to find out, what the numbers are actually measuring” (One prominent author in a presentation at Nordic Workshop on Bibliometrics and Research Policy 2014)
- Absolutely no control over data, entries come and go
Bibliometrics and the SSH
Bibliometrics are not valid in the SSH

- Consensus among bibliometricians that bibliometrics are not (yet) applicable to the SSH
  - Different publication patterns (Hicks, 2004)
  - Different citation practices (van Leeuwen, 2006)
  - Lack of coverage in data bases (van Leeuwen, 2013)
  - Language issues (Nederhof, 2005)
  - US over-covered (60%) UK over-covered (20%); not only English-bias (70% in English) but especially nationality (Chi, 2014)

- Further problems:
  - No linear progress of research (cf. Price’s Law; Lack, 2008)
  - Interaction with public – non-scholarly publications are important
  - WoS and Scopus exhibit citation-matching problems for non-hard science publications → Loss of citations
Thomson Reuters reacted with inclusion of books
- However, not usable (Gorraiz, 2012)
- Mostly, natural sciences and engineering books
- Almost only edited volumes
- Monographs: entries chapter by chapter
- No transparency on which books are included and why (vs. clear selection criteria for journals)

Elsevier also includes books
- Same problems as above: chapter by chapter

Google
- Books included as long as they are open access or publisher has a contract with Google (bias!)
Coverage issues (internal)

Black: hard sciences
Blue: engineering/applied
Green: economics
Orange: social sciences
Red: humanities

Figure 2: Coverage of disciplinary output in WoS, 2010.
Coverage issues (internal)
Evaluative bibliometrics: citations and quality

- Bibliometrics widely used in STEM to evaluate research
- This comes with assumptions
  - Citations as “currency of science” (Merton, 1962, personal communication to Garfield)
  - Citation as a predictor for quality (but: citations measure many things, Moed, 2005; Bornmann et al., 2008)
  - Coverage: the data base must include most important research adequately (80%-rule)
  - Linear progress of research
  - Citation practices are similar in subjects that are evaluated (but: van Leeuwen, 2006)
  - There is nothing else that is not correlated with citations that is important for the quality of research (but: Ochsner et al., 2012; Hug et al., 2013)
Conclusion

- Bibliometrics not usable for evaluative purposes in SSH
- Nevertheless high pressure to do so
  - External: Science policy (e.g., REF)
  - Internal: Appointments are very often done using “poor man’s” bibliometrics
  - Internal: Scholars themselves start to use “poor man’s” bibliometrics
- However, even in natural sciences where data is quite reliable, scholars start to oppose (DORA, Leiden-Manifesto)
Music as an example
Spotify

- Streaming service
- Provides statistics for each group, album, song etc.
- Provides playlists according to these statistics “Most listened” but also “hot”, “Best of” etc.
# Van Halen

## Popular Songs

<table>
<thead>
<tr>
<th>Rank</th>
<th>Song Title</th>
<th>Views</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Jump - 2015 Remastered Version</td>
<td>48,128,824</td>
</tr>
<tr>
<td>2</td>
<td>Runnin' With The Devil - 2015 Remastered Version</td>
<td>12,349,044</td>
</tr>
<tr>
<td>3</td>
<td>Why Can't This Be Love</td>
<td>7,727,813</td>
</tr>
<tr>
<td>4</td>
<td>Panama - 2015 Remastered Version</td>
<td>15,237,918</td>
</tr>
<tr>
<td>5</td>
<td>Dreams</td>
<td>3,932,693</td>
</tr>
</tbody>
</table>
And the Beatles?

<table>
<thead>
<tr>
<th>#</th>
<th>Song Title</th>
<th>Plays</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ain't She Sweet</td>
<td>5,176,798</td>
</tr>
<tr>
<td>2</td>
<td>Cry For A Shadow</td>
<td>2,294,081</td>
</tr>
<tr>
<td>3</td>
<td>If You Love Me, Baby (Take Out Some Insurance On Me, Baby)</td>
<td>1,714,448</td>
</tr>
<tr>
<td>4</td>
<td>Why (Can't You Love Me Again)</td>
<td>1,290,240</td>
</tr>
<tr>
<td>5</td>
<td>Plastic Beetle</td>
<td>881,283</td>
</tr>
</tbody>
</table>
ARTIST

Ludwig van Beethoven

OVERVIEW RELATED ARTISTS ABOUT

LATEST RELEASE

Violin Sonatas Nos. 6, 7 And 10

30 OCTOBER 2015

POPULAR

1  
Sonata No. 14 "Moonlight" in C-Sharp Minor*, Op. 27 No. 2: I. Adagio sostenuto 1,006,825

2  

3  
Piano Sonata in D Major, Op. 28 - 'Pastoral': I. Allegro 3,300,433

4  
Piano Sonata No. 4 in E-Flat, Op. 7: III. Allegro 1,391,440

5  
Concerto for Piano and Orchestra No. 5 in E-Flat Major, Op. 73: II. Adagio un poco moto... 937,500

Research Methods Seminar, Unil  |  03.09.2018  |  36
OK, maybe followers?
Albums sold?

- Beatles: 600 Millions
- Van Halen: 96.5 Millions
- Beethoven: approx. 10 Millions
- Macklemore & Ryan Lewis: <2 Millions

→ Obviously, Spotify’s data base is biased, so are album sales: coverage issues, language issues, music styles

→ And, just like WoS, Scopus, and even Google:
Where do the listeners (or listens) come from?
Light at the end of the tunnel
Spotify: What are “listens” really good for?
What can bibliometrics be used for?

- Analogous to Spotify: discover related authors, papers, emerging fields
- What it has been developed for:
  - Help for librarians (most used journals, core journals of subject etc.)
  - For retrieval problems
- But also for science policy:
  - Detect emerging fields
  - Build new organisational units (university level)
- Yet, in SSH always be cautious and look for results of different data bases (languages, regions, fields etc.)
Conclusions
Take-home messages

- Never use evaluative bibliometrics in SSH
- Don’t trust the data – any data
- Never use the IF to judge an article
  - Do never sum up your articles according to JIF (ResearchGate!)
- Do not select your journal ONLY according to JIF
- If you talk bibliometrics: mention the data base you are referring to (“I have an h-index of 12” – Google or WoS?)
- Never compare bibliometric indicators across disciplines
- Use citations to find new references and new authors
- Use citation networks to discover new topics
Thank you for your attention

- More information:
  - European Association for Research Evaluation in the SSH: www.evalhum.eu
  - Swiss project on research evaluation in the SSH: www.performances-recherche.ch/ (Website under construction but partly available)
  - Our own project on quality criteria for research in the humanities: http://www.psh.ethz.ch/forschung/anwendung-von-bottom-up-kriterien-zur-beurteilung-von-geisteswis.html