Scientific Information Management: Bibliometrics

MIKAEL GRAFFNER
LUND UNIVERSITY LIBRARY
Outline

- Introduction to bibliometrics
- Research evaluation
- Common indicators
- Short break
- Exercises
- Discussion
- Altmetrics
- Bibliometrics and resource allocation
- Bibliometric services at LU
- Publishing strategies
What is bibliometrics?

• Biblion=book, metron=measure

• “The application of mathematics and statistical methods to books and other communication media.” (Pritchard, 1969)

• Related terms:
  – Informetrics (Nacke 1979)
  – Scientometrics (Nalimov, 1966)
  – Webometrics (Almind & Ingwersen, 1997)
Bibliometrics and the scientific communication system

Bibliometrics provide different ways of analyzing the traces left behind by scientific communication in the form of bibliographic information.
Bibliometric methods are used to study…

- Properties of texts
- Structures in research fields
- Production of research and aspects of quality

<table>
<thead>
<tr>
<th>Analysis of:</th>
<th>Examples:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequencies</td>
<td>Authors, references, key words</td>
</tr>
<tr>
<td>Distributions</td>
<td>Publications/journals, citations/author, words/document</td>
</tr>
<tr>
<td>Co-occurencies</td>
<td>Co-authorships, bibliographic coupling, key words</td>
</tr>
</tbody>
</table>
Maps of Science

A visualization of 7.2 million scholarly documents appearing in over 16,000 journals, proceedings, or symposia between Jan, 2001 and Dec, 2005.

Source: SciTech Strategies
http://mapofscience.com
LU collaborations
Visualization of terms in titles and abstracts

4,286 articles from Web of Science. TS=Melanoma, PY=2012,
Research evaluation – two main options

Published research

Peer review

Bibliometrics

Quality

Quantity
The peer review system is still the gold standard for evaluation of research but...

...there are also weaknesses:

• Subjectivity
• Bias
• Hard to find the right competencies for narrow research fields and interdisciplinary fields
• Expensive and time consuming
Research evaluation

- Bibliometric indicators for research evaluation were first developed during the 1970s.
- Response to a growing need for transparent measures which can be used for allocation of research funds, strategic planning and appointments.
- Bibliometric evaluations are based on number of publications and/or citations.
- Different indicators for different purposes:
  - Different indicators should be combined to illuminate different aspects of the research.
  - More reliable on higher aggregation levels.
Assumptions when using indicators in evaluations

• Basic assumptions:
  – Number of publications is an indicator of productivity
  – If many people cite a publication this is an indication of scientific impact
  – Science is a cumulative process, i.e. research is based on prior findings

• Works well for natural sciences, medicine and some areas of technology

• Less suitable for humanities and social science
Arguments for using bibliometrics in evaluations

• Measurement=counting standardized elements
• Publications in peer reviewed journals are the best and most easily accessible unit in most fields
• Publications can be seen as building blocks in the evolution of science
• Citations show influence and impact, which is an important aspect of scientific quality
But…

• In some disciplines journals aren’t the most important channel for communicating research results
• Citation patterns vary between different fields and it’s not always easy to find comparable units
• Bibliometric evaluations can create unforeseen effects: changed publications patterns, short term strategies etc.
Citation databases

• **Web of Science**
  - Database which indexes 12 000 journals
  - Varying coverage depending on research field:
    - very good coverage of science and medicine
    - medium good coverage of engineering
    - in general poor coverage of social science and humanities

• **Scopus**
  - 22 000 peer reviewed journals, citation data from 1996 1970
  - Better coverage of SSH and parts of engineering
Other resources

- **Google Scholar**
  - Broader in coverage but also includes publications which aren’t strictly scientific
  - Less reliable citation data
  - **Author profiles**
- **Dimensions**
  - New platform based on data from Open Citation Corpus
- **Microsoft Academic**
Cited & citing documents
Cited & citing documents, 2 generations

Source: ISI Web of Knowledge®, www.isinet.com
PROTEIN MEASUREMENT WITH THE FOLIN PHENOL REAGENT*

BY OLIVER H. LOWRY, NIKA J. ROSEDBROUGH, A. LEWIS FARR, AND ROSE J. RANDALL

(From the Department of Pharmacology, Washington University
School of Medicine, St. Louis, Missouri)

(Received for publication, May 28, 1951)

Since 1922 when Wu proposed the use of the Folin phenol reagent for the measurement of proteins (1), a number of modified analytical procedures utilizing this reagent have been reported for the determination of proteins in serum (2–6), in antigen-antibody precipitates (7–9), and in insulin (10).

“Although I really know it is not a great paper ... I secretly get a kick out of the response”
– Oliver Lowry 1977
Common bibliometric indicators...
Journal impact factor

• Shows how much the average publication within a journal is cited
• Published annually in the Journal Citation Report (JCR)
• In a given JCR year (e.g. 2016), the impact factor of a journal is the average number of citations to those papers that were published during the two preceding years (2014 & 2015)
Problems with JIF

• Criticism:
  – Too short citation window (2 years)
  – No field normalization
  – No normalization of document types
  – Incorrect calculation (total number of citations/”citeable documents”)
  – Skewed distributions of highly cited publications within journals

• Often used to evaluate the quality of individual articles but studies have shown that there is little correlation between JIF for a journal and citation rates for the individual publications in it
“At PLOS, we believe that research articles should primarily be judged on their individual merits, rather than on the basis of the journal in which they were published.”

“eLife will not promote the Impact Factor. We’ll promote your work, and provide quantitative and qualitative indicators about its reach and influence. eLife is working to expand and enrich the concept of research impact beyond the use of a single number and a journal name.”

(impact factor=7.73)
New journal metrics

- CiteScore
- SJR – SCImago Journal Rank
- SNIP – Source normalized Impact Per Paper
- Eigenfactor and Article Influence Score
H-index

• H-index for an author is the number of publications \((h)\) from this author which have been cited at least \(h\) times

• Example: An author has published 25 articles out of which 12 have been cited at least 12 times, H-index=12

• Advantage:
  – A measure of both production and citations
  – Easy to calculate

• Criticism:
  – Favours older authors with long publication lists/unfair to young researchers at the beginning of their careers
  – Unfair to researchers with few but very highly cited articles
  – Cannot be used for comparisons across different research fields
“The brilliance of the h-index is that it provides a single, easy to compute, quantitative measure of your cumulative impact. You want your impact to go up! So, it follows directly and easily that all decisions in your career should be considered in terms of their potential to boost your h-index.”

Citation potential depends on research field

Citation rates and document types

Figure 3 - Average citation rates for articles, letters and reviews in CI published in 1998-2004. Data from the citation indices produced by Thomson Scientific, self citations are included and whole counting is performed.

Lundberg, Jonas (2006). Bibliometrics as a research assessment tool – impact beyond the impact factor, s. 14
Field normalized citations

- CPP/FCSm – "crown indicator".
- The citation score for all publications are normalized according to field, document type and age
- Possible to compare research from different fields
- If the mean field normalized score is 1,00, this means that the analyzed articles have the same citation rate as the world average
- 1,12 = 12% more citations than the world average
Field normalized citations

Number of citations

Average number of citations in subject category

Publications in the same subject category

Same category, publishing year, and document type
Different indicators

- Top 5% - share of articles from a studied unit that belong to the top 5% most cited within the field
- Uncited – share of publications with 0 citations
- CPP/JCSm – normalized value based on journal set
- JCSm/FCSm – shows if a unit publishes in journals with a high or low impact compared to the journal category as a whole
Excercises

• Search for the H-index for your supervisor or a senior researcher within your department or group. Use Web of Science (apps.webofknowledge.com) or Scopus (www.scopus.com)

• Use a search based on key words to try and delineate your specific research area. Look at a couple of the most highly cited articles and compare the number of citations in Web of Science, Scopus and Google Scholar.

• If you look at the same result list, which are the most common journals in your field? Which authors have produced the most articles?

• Look up some key journals in your research area in Journal Citation Reports. https://jcr.incites.thomsonreuters.com
Discussion

• What have you found out regarding the publication patterns in your field?
• Were there any differences between the results in WoS and Google scholar?
• What are the h-indices for the researchers you’ve focused on? Why are they different?
• Which journals are the most important within your field?
Altmetrics

• Umbrella term for different kinds of metrics which offer an alternative to traditional citation analysis
• Is usually based on the individual publication
• A number of commercial actors provide altmetrics:
  – ImpactStory
  – Plum Analytics
  – Altmetric
• Response to demands from funders and researchers who wish to measure the impact of published research with a broader set of tools
Different types of use and impact

http://www.niso.org/publications/isq/2013/v25no2/lin/
Advantages with altmetrics

• Gives a quick indication of the impact of recently published research
• Gives a broader view of the visibility and uptake, also outside academia
• Can be used to measure output types that aren’t normally covered by the citation databases
Drawbacks

• Measures popularity and says nothing about the quality
• Possible to manipulate
• Many different service providers, black box solutions and lack of standardization
• Gamification of research
Bibliometrics and resource allocation
Bibliometrics affect economic decisions on many different levels in academia

- Individual researcher ↔ research funders
- Research groups ↔ research funders
- Department/faculties ↔ faculty/university
- University ↔ state government
The bibliometric indicator for allocation of research funds in Sweden

• Confusion since the previous model (2009-2016) was scrapped

• In the budget for 2018, three quality indicators are used:
  – Bibliometrics
  – External funding
  – Collaboration

• The bibliometric indicator is based on each university’s share of articles belonging to the 10% most cited

• A new model is currently under investigation...
The resistance is growing…

- San Francisco Declaration on Research Assessment (DORA)
- The Leiden Manifesto for research metrics https://vimeo.com/133683418
The UK Research Councils have today (Wednesday 7th February) published a statement on the responsible use of metrics and signed the San Francisco Declaration on Research Assessment (DORA).

The Research Councils want to encourage active discussion by the research community on this policy area and, where appropriate, set out good practice in the way that outputs of scientific research such as journal articles are evaluated by funding agencies and academic institutions. This is important as the research councils rely on high quality independent expert peer review to select the most promising research to support.

The Research Councils’ statement includes a commitment to update the principles that councils will use to guide researchers applying for research council funding, to guide reviewers in assessing these applications, and to guide evaluations of the work that results. Central to this guidance is a steer to not place undue emphasis on the journal in which papers are published, but assess the content of specific papers, when considering the impact of an individual researcher’s contribution. The statement sets out a brief action plan that will be developed further over time.
The Leiden Manifesto for research metrics

1. Quantitative evaluation should support qualitative, expert assessment.
2. Measure performance against the research missions of the institution, group or researcher.
3. Protect excellence in locally relevant research.
4. Keep data collection and analytical processes open, transparent and simple.
5. Allow those evaluated to verify data and analysis.
6. Account for variation by field in publication and citation practices.
7. Base assessment of individual researchers on a qualitative judgement of their portfolio.
8. Avoid misplaced concreteness and false precision.
9. Recognize the systemic effects of assessment and indicators.
10. Scrutinize indicators regularly and update them.

Nature 520, 429–431 (23 April 2015) doi:10.1038/520429a
Bibliometrics at LU

- Lund University Library offers bibliometric services for researchers and administrators within LU
- All scientific publications from LU researchers are to be registered in LUCRIS
- Records from all LU-affiliated authors are imported from Scopus every month
  - About 400+ records/month
  - The records are reviewed and supplemented manually before they’re made public
  - Authors are verified using Lucat-id’s
  - Each record is assigned to one or more LU departments or divisions
Questions and discussion

• How can bibliometrics be used for evaluating funding proposals?
• Can they be used for enhancement of funding applications?
• How trustworthy are bibliometrics?
• How long has bibliometrics been used?
• Why use bibliometric analyses? And in which context are they appropriate to gain meaningful insights?
• How does bibliometrics help me to get updated with the new findings in my research area?
• What metrics are most frequently used and why?
• How do you weigh different measures against each other?
• Are there any guidelines regarding bibliometrics at Lund University when employing researchers?
• Are there any general guidelines worldwide?
• What are the main limitations of bibliometrics?
Publishing strategies?

- Which is the best audience for your research?
- Choose the right (WoS/Scopus-indexed?) journal
- Open Access?
- Consistent name syntax, correct author address
- Descriptive title, informative and clear abstract + key words
- Claim your publications: ORCID, Google Scholar, ResearcherID etc
- Co-publish?
- Use different platforms to promote your research: ResearchGate, academia.edu, Mendeley, blogs, twitter etc.
- Make sure the publication is registered in LUCRIS
Thank you for listening!

Contact:

mikael.graffner@ub.lu.se

046-222 15 30