Some issues on Bibliometrics

the way I would like to be helped as a University Professor

Luis Borges Gouveia, lmbg@ufp.edu.pt
4th December, 2013
University Fernando Pessoa
Talk abstract:

**Some issues on Bibliometrics...**

The availability of digital based information services and the growing number of Web based tools provide new ways to deal with quantitative analysis and statistics to describe publication patterns, authors or journals impact and its influence and weight for R&D (or as placed now R&I) evaluation and even funding.

In a changing world of higher education where research and scientific production is getting more and more a global comparison bias, university staff support to harness their efforts and maximize results is more strategic. The presentation provides a personal report on the context, the challenges and current common restrictions felt when dealing with such issues.

Additionally, a number of ongoing activities and Internet tools are presented as the rise of Altmetrics and the use of Internet based tools for reputations and cross reference management, including individual research identifiers as ORCID.
One initial word...

• This work is a personal statement regarding the good and the not so good about bibliometrics

• It provides a view for sharing information and serve as a starting point to improve and provide support to direct and maximize ongoing efforts to produce more impact and assess achieved influence

• Still, it is working in progress, started 18 years ago...
Presentation content

- User digimetrics
- Current challenges
- From bibliometrics to altmetrics
A Word on Bibliometrics
Some critics

• Quantity different from quality
• Almost impossible to provide an all inclusive account

*(is there a place where we can collect/include all the work?)*
“Bibliometrix” (a dark secret art?)
“Bibliometrix”

• **Bibliometrics** – is the quantitative analysis of documents in scientific communication; the documents reflect the state of scientific knowledge.

• **Cybermetrics** – is the quantitative research of information sources, structures and technologies on the Internet; a study of discussion groups, email communication, etc.

• **Informetrics** – is focused on the information stream in networks and demonstrates on the basis of mathematical and statistical methods a variety of relations between them.

• **Scientometrics** – is focused on the evaluation of efficiency of scientific-research or individual researchers by citation counts.

• **Altmetrics** – refer to alternative metrics that combine both informetrics and bibliometrics and strive to become a substitute of Bibliometrics
Bibliometrics

• Bibliometrics is a research method used in library and information science
  – uses quantitative analysis and statistics to describe patterns of publication

• Bibliometrics as a means of evaluation:
  – to determine the influence of a researcher
  – the relationship between two or more researchers or works
  – a R&D group / lab / university / nation / area
  – a given journal / conference / scientific event

• A nice tutorial (MyRI, Measuring your Research Impact):
  http://www.ndlr.ie/myri/MyRI_Tutorial/player.html
Old times: from counting to digital

• The era of quantitative, computer-tabulated science metrics dates back to the 1950s, when linguist Eugene Garfield began indexing the scientific literature using punch cards
  – A company in Philadelphia, Pennsylvania, that Garfield founded in 1955
  – Renamed Institute for Scientific Information (ISI) in 1960, when it began to publish the Science Citation Index: a systematic effort to track citations — the footnotes by which journal authors acknowledge their intellectual debts. (ISI is now owned by Thomson Reuters)

• In 1965, Garfield and his colleagues used ISI's databases to show that Nobel laureates published five times the average number of papers, and that their work was cited 30 to 50 times the average
  – A finding that for decades established citations as the pre-eminent quantitative measure of a scientist's influence
  – Introduces the number of publications and the number of citations as quality work quantifiers
To much, to many...

Bibliometrics grows in importance and assessment use
When use tools to measure

• What to measure
• Metrics to be used
• Benchmarking context
• Comparing what or whom
• Measure for improving
• To assess a policy
• To meet a strategy
• For fun...
• As an excuse...
• Because the others do...
• As a requirement, mainly for funding
When use tools to measure

- What to measure
- Metrics to be used
- Benchmarking context
- Comparing what or whom
- Measure for improving
- To assess a policy
- To meet a strategy
- For fun...
- As an excuse...
- Because the others do...
- As a requirement, mainly for funding
Laws of Bibliometrics – Lotka's Law

• Lotka's Law describes the frequency of publication by authors in a given field. It states that "... the number (of authors) making n contributions is about 1/n² of those making one; and the proportion of all contributors, that make a single contribution, is about 60 percent".

• This means that out of all the authors in a given field:
  – 60% will have just one publication
  – 15% will have two publications (1/2² times .60)
  – 7% will have three publications (1/3² times .60), and so on.

• According to Lotka's Law of scientific productivity, only six percent of the authors in a field will produce more than 10 articles.

• Lotka's Law, when applied to large bodies of literature over a fairly long period of time, can be accurate in general, but not statistically exact.

• It is often used to estimate the frequency with which authors will appear in a catalog

Alfred J. Lotka (1880, 1949)
Laws of Bibliometrics – Bradford's Law

• Bradford's Law serves as a general guideline to librarians in determining the number of core journals in any given field.

• States that journals in a single field can be divided into three parts, each containing the same number of articles:
  1. The core of journals on the subject, relatively few in number, that produces approximately one-third of all the articles;
  2. A second zone, containing the same number of articles as the first, but a greater number of journals;
  3. A third zone, containing the same number of articles as the second, but a still greater number of journals.

• The mathematical relationship of the number of journals in the core to the first zone is a constant $n$ and to the second zone the relationship is $n^2$.

• Bradford expressed this relationship as $1:n:n^2$.

• Bradford's Law is not statistically accurate, but it is used as a general rule of thumb.

Samuel Clement Bradford (1878 — 1948)
Laws of Bibliometrics – Zipf's Law

• Zipf's Law is often used to predict the frequency of words within a text. It states that in a relatively lengthy text, “if you list the words occurring within that text in order of decreasing frequency, the rank of a word on that list multiplied by its frequency will equal a constant”. The equation for this relationship is: \( R \times F = C \) where \( R \) is the rank of the word, \( F \) is the frequency, and \( K \) is the constant.

• Use not only in language context, but in many other collections of data as the case of Internet or even the distribution of income within nations (for example, the rank versus distribution of individual incomes in a country approximates this law. Breaks in this normal distribution create pressure for change, even revolution).

• Zipf's Law, again, is not statistically perfect, but it is very useful for indexers.

George Kingsley Zipf (1902-1950)
**First: count work (not all equal)**

- Journals
  - International journals
  - Local journals
- Published books
  - International editors
  - National or even a lower level of local editors (excluding or adding self publication)
- Book chapters
  - International editors
  - National or even a lower level of local editors (excluding or adding self publication)
- Conferences
  - International
  - National
- Other work
  - Prizes and honors
  - Patents
  - Projects
- Academic work
  - Thesis and dissertations and other public or documental formal referred documents
  - Internal reports as of a clear selection overflow
- Observations:
  - All can be distinguished by (being indexed on major databases and ranked by impact factor and/or A, B, C rankings)
  - All works must be refereed by peers and recognized as genuine academic work
  - Other media and new media work can be listed but not count as the case for talks, keynotes, media, interviews, Web materials and contents, videos and audio content, multimedia and even transmedia. Creative work, art and collective work, versioning, trademarks, brands, software
Second: Measuring influence (not the same of work...)

- **Number of citations**
  - The sum of all the citations from the authors papers
  - We can differentiate self citations from others, first authors and the order position in authorship

- **H-index**
  - The h-index measures the broad impact of a researcher’s work
  - A scholar with an index of $h$ has published $h$ papers each of which has been cited in other papers at least $h$ times
  - Created in 2005 by Jorge E. Hirsch

- **G-index**
  - The g-index is similar to h-index, but takes into account highly cited papers
  - A g-index of $n$ show that an author produces $n$ articles who have, on average, $n$ citations for each of them
  - Created in 2006 by Leo Egghe
Third: Why can this be important

• A professor asks
  – Who is citing my articles?
  – How many times have I been cited? Where and from whom?
  – What are my h-index and g-index?
• A student asks
  – How do I know if this article is important?
• Another professor asks
  – Which journal should I publish in for more impact?
  – What are the best journals to publish my research?
• A different student
  – What are the best journals in my studying field?
• Many other questions possible from other parts than professors and students (university staff, services and boards, management, external assessment boards, etc.)
Sources and Tools for bibliometric data

• Sources
  – **Scopus**: [http://www.elsevier.com/online-tools/scopus](http://www.elsevier.com/online-tools/scopus)
  – Google Scholar: [http://scholar.google.com](http://scholar.google.com)

• Tools
  – Harzing’s **Publish or Perish**: [http://www.harzing.com](http://www.harzing.com)
  – Loet Leydesdorff collection: [http://www.leydesdorff.net/software.htm](http://www.leydesdorff.net/software.htm)
  – BibExcel: [http://www8.umu.se/inforsk/Bibexcel/](http://www8.umu.se/inforsk/Bibexcel/)
  – Pajek: [http://pajek.imfm.si/doku.php](http://pajek.imfm.si/doku.php)
  – CiteSpace: [http://cluster.cis.drexel.edu/~cchen/citespace/](http://cluster.cis.drexel.edu/~cchen/citespace/)
As so...

• First: count work (not all equal)
• Second: Measuring influence
• Third: Why can this be important
• Sources and Tools for bibliometric data

• Explore, experiment, seek help...
As so...

- First: count work (not all equal)
- Second: Measuring influence
- Third: Why can this be important
- Sources and Tools for bibliometric data

- Explore, experiment, seek help...
A Word on being Professor
Be a University Professor

• Think of it as the need to use different hats
  – One for each main role
  – One that requires alternative skills
  – One that uses proper balance to deal with time and other people
  – They are somewhat overlapping but essential they compete for effort, resources, will and time of each one of us
  – Need to respond to all of them to be complete
Not the same of 6 thinking hats…
(Edward De Bono)

but somewhat inspired in 😊
No magic... just a working perspective

Helps organizing activity and for information management
The process

• It take some time (possible to jump some)
  – Starts with a PhD
  – Can goes to a Post doc
  – A Research Assistant (RA) position
  – Lecturer...
  – Then tenure...
  – Become Assistant Professor, Associate Professor
  – Full Professor, and stay there

• Causion: is this still a valid/viable path?
A path that relates degrees and pubs
*TRS Technology, Networks & Society

Networks

knowledge
information
tecnology

E-planning

E-government
E-health
E-learning

November, 2013 I&D+i framework
Five hats for university Professor

- Teaching
- Research
- Supervising
- Consulting
- Managing
Five hats for university Professor

- teaching
- subjects
- levels
- impact
- hours
- students

Consulting
Research
Five hats for university Professor

- teaching
- supervising
- managing
- consulting

- supervising
- degree
  - PhDs
  - MScs
  - other
- type
- other
Five hats for university Professor

managing

context

level

roles

time

scope
Five hats for university Professor

- Teaching
- Research
- Consulting
- Industry work
- Managing
- Support
- Brands
- Projects
- Value
Five hats for university Professor

- teaching
- research
- consulting
- publications
- patents
- projects
- citations
- outcomes
Five hats for university Professor

Enterprise side

Institucional side

- teaching
- research
- supervising
- consulting
- managing
• Which hat fits you?
• Which side is yours?

...hope these are the wrong questions!

[my position]
To be a Professor it is needed to be the full circle, although that some balance is not to ask ideal perfection (or, for that matter, the impossible)
A Word on HEI Challenges
Current context (I)

• Higher education are on change
  – its global and a non return issue
• In the Portuguese case, it was introduced a minimal member rule for R&D units:
  – Phd until 2009: 4 journal or book publications
  – Phd after 2009: 2 publications (being one, the thesis)
• Why five years?
  – Absolescence knowledge rate: 20%
    • Within a five years timeframe, knowledge need to be renewed...
## Current context (II)

<table>
<thead>
<tr>
<th>year</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>2009</td>
<td>2010</td>
<td>2011</td>
<td>2012</td>
<td>2013</td>
</tr>
<tr>
<td></td>
<td>Phd A, 2 pubs</td>
<td>Phd B, 2 pubs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>2010</td>
<td>2011</td>
<td>2012</td>
<td>2013</td>
<td>2014</td>
</tr>
<tr>
<td></td>
<td>A, 4 pubs!</td>
<td>Phd B, 2 pubs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Phd B, 2 pubs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B, 4 pubs!</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Current context (II)

Publishing
- Internal reports
- Conferences
- Book chapters
- Journals
- Books

Citations
- Listed
  - Scopus, ISI
  - DOAJ, ACM, IEEE, other...
- Cited
  - ISI, Scopus
  - Google Scholar

Funding
- Grants
- Publication support
- Consulting
- Projects
Challenges (within HEI context)

• More global
• More competitive
• A market for students not for university staff
• Real time & easy to compare and monitor
• Digital offerings
  – No cost
  – Open access
  – High quality
  – Sofisticated
  – Fast pace, growing alternative to choose from
Things are changed a lot in HEI
HEI – Higher Education Institutions
new highlights to consider...

OPEN  EXPERIENCE  POPULATED
GLOBAL  DIVERSE  SHARE
INTERACTIVE  VALUE  EMOTIONAL
FLEXIBLE  DIGITAL
IMPACTANT  INNOVATIVE
CONNECTED  FASHION
NETWORKED
A Word on Web time
Old and New (bibliometrics) gamers

- Research ID (Thomson Reuters)
- Scopus Author ID (Elsevier)
- Google citations (Google)
- Research Gate

- Information services
  - Information collectors (e.g. DEGOIS)
  - Information aggregators (e.g. ORCID)
ISI Research ID

http://www.researcherid.com/rid/E-9682-2010
Scopus author ID

http://www.scopus.com/authid/detail.url?authorId=7004135841
Google citations

http://scholar.google.com/citations?user=xF1eGXcAAAAJ&hl=en
Research Gate

https://www.researchgate.net/profile/Luis_Borges_Gouveia/
ResearchGate as an open score

5.96
RG Score

A new way to measure scientific reputation.
The RG Score takes all your research and turns it into a source of reputation.

RG SCORE OVER TIME

How does the RG Score work?
Your RG Score is calculated based on how other researchers interact with your content, how often, and who they are. The higher their score, the more yours will increase.

Learn more
Open score...

- Transform the process
  - Networked
  - Cross information
  - Gamification
  - Still use ISI impact factors from published journals
Information services (DEGOIS & ORCID)
DEGOIS

http://www.degois.pt/visualizador/curriculum.jsp?key=4336687962859213
ORCID

http://orcid.org/0000-0002-2079-3234

Luis Borges Gouveia

http://orcid.org/0000-0002-2079-3234

Keywords: e-government, e-learning

Websites:
About me page
Homepage

Other IDs:
ResearcherID: E-9682-2010
Scopus Author ID: 7004135841

Personal Information

Biography
Aggregate in Engineering and Industrial Management (University of Aveiro, 2010)
PhD in Computer Science (University of Lancaster, 2002)
MSc in Electronic and Computer Engineering (University of Porto, 1995)
Dip in Informatics (University Portucalense, 1989)

Publications

Using Games for Primary School: Concepts, Methodologies, Tools, and Applications 2013


A worked proposal on eparticipation for state wide elections: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 2012

The power of freely content publishing in Web

• Self publication still matters (and has its own role)
  – Lots of people, worldwide seek free and non structured and/or curated information
  – Easy to access, find and use
  – No need to get a context of training and support
  – Can be a problem, but it is gained more and more adepts
  – Fast, no cost, availability, but depends on user knowledge
  – ...also starts to be explored for open access (see http://www.slideshare.net/lmbg/the-library-the-digital-and-the-quest)
Luis Borges Gouveia

Associate Professor
Professor Associado com Agregação
Science and Technology Faculty
Faculdade de Ciência e Tecnologia
University Fernando Pessoa (UFP)
Universidade Fernando Pessoa

publications (date & type)
publicações e comunicações

published books
livros publicados

online materials
textos disponibilizados

work on thesis, dissertations and monographs
trabalhos de graduação e pós-graduação

credits | my portal | blogue (pt) | blog (en) | Grupo *TRS
Social Web includes a growing number of networked based platforms to promote, connect and share information.
Still some use for old fashion Web pages
An alternative data source
institucional open access repositories

http://bdigital.ufp.pt/handle/10284/3787
• Provide alternative metrics to use
  – this example is from the link of the open access presentation
• File deposit made in end June
  – 145 days availability
  – get 138 downloads
  – get 273 hits
  – from 25 different locations
URLs to Explore
About scientific journals

- Journals search; [http://journalseek.net/](http://journalseek.net/)
- Journals for free: [http://www.journals4free.com](http://www.journals4free.com)
Assorted resources

• Portuguese access portal to scientific data, b-on: http://www.b-on.pt/
• Portuguese open access search engine, RCAAP: http://www.rcaap.pt/
• I&D search engine: http://worldwidescience.org/
• Proofreader: http://www.grammarly.com
• Ethical and journal procedures: http://publicationethics.org/resources/flowcharts
• Author copyrights, licencing and auto-archive policies: http://www.sherpa.ac.uk/romeo/
• CrossRef: http://www.crossref.org
• Get the URL for a DOI: http://dx.doi.org/
About citations

• ISI Web of Knowledge: http://wokinfo.com/
  Journal Citation Reports (JCR): http://thomsonreuters.com/journal-citation-reports/

• SCOPUS: http://www.scopus.com

• Google Citations: http://scholar.google.pt/citations
  Academic Google: http://scholar.google.pt

• Microsoft Academic Research:
  http://academic.research.microsoft.com/

• Computer Science h-index authors listing:
  http://www.cs.ucla.edu/~palsberg/h-number.html
Explore & create your IDs
The word cloud of this presentation
Presentation conducted by

• Luis Borges Gouveia is an Associate Professor of Knowledge Management, Science and Technology Faculty at University Fernando Pessoa

• Among his research interests are the impact of digital information and information society issues in our day-to-day life.

• He publishes 12 books and is involved in several projects concerning the use of information in e-learning and e-government settings
  – More information is available at http://about.me/lbgouveia