Introduction to Webometrics: Web-based Methods for Research Evaluation

Mike Thelwall
Statistical Cybermetrics Research Group
University of Wolverhampton, UK
What is webometrics?
Webometrics in *traditional* research evaluation
Webometrics for *new* research evaluations
Webometrics for *social science*
1. What is webometrics?
1. What is webometrics?

- Webometrics is gathering data on the Web, and measuring aspects of the Web:
  - web sites
  - web pages
  - hyperlinks
  - web search engine results
  - YouTube video commenter networks
  - MySpace Friend networks

- ...for varied social science purposes
New problems: Web-based phenomena

- Webometrics can analyse online academic communication
  - Why do academic web sites interlink?
  - Which academic web sites interlink?
  - What academic interlinking patterns exist?
  - Which web sites/groups/universities have the most online impact, and why?
Old problems: Offline phenomena reflected online

- Some offline phenomena have measurable online reflections
  - International communication
  - Inter-university collaboration
  - University-business collaboration
  - The impact or spread of ideas
  - Public opinion about science
Example: The online impact of research groups (NetReAct)

Link source site for pages linking to Hungarian LSRGs
Example: Links between EU universities

Normalised linking, smallest countries removed
International biofuels research network
Data Gathering/Processing Tools

- LexiURL Searcher – links, web text, YouTube, Flickr, Technorati
  - Submits thousands of queries to Live Search/Yahoo and summarises the results in standard ways
- SocSciBot – links, web text
- Web Crawler & analyser
2. Webometrics in traditional research evaluation
2. Webometrics in traditional research evaluation

- Webometrics can supplement traditional citation impact non-traditional online impact
  - E.g., educational, discussion-based
- It is typically weaker than citation data but is useful for research groups that have non-standard types of impacts
The Integrated Online Impact Indicator (IOI)

- Combines a range of online sources into one indicator
  - Google Scholar +
  - Google Books +
  - Course reading lists +
  - Google Blogs +
  - PowerPoint presentations = IOI

- OR select individual separate components
New source 1: Google Scholar

- Wider evidence of academic impact
- Wider types of academic publications, some non-academic publications
- Not reliable
- Coverage variable
- Can’t be automatically queried
- Free
New source 2: Google Books

- Books typically not indexed in WoS or Scopus
- Relevant in book-based disciplines (arts, humanities, some social sciences)
- Reliability unknown but probably not good
- Coverage variable
- Can be automatically queried
- Free  
  (Clifford Lynch)
New source 3: Course reading lists

- Evidence of educational impact
- Can automatically construct queries to detect individual articles in online syllabuses
- Get results via advanced Google/Yahoo/Live Search queries
- Works for most articles
  - Fails for short common article titles
New source 4: Blogs

- Evidence of impact on discussions
- Educational impact, public dissemination evidence, academic impact in discursive subjects?
- Not possible to automate in the largest database (Google Blogs)?
- Not a well researched area
New source 5: PowerPoint Presentations

- Evidence of educational/scholarly impact
- Especially relevant for discursive subjects?
- Automated Live Search/Yahoo advanced queries

\[ IOI = a \times \text{Scholar} + b \times \text{PowerPoint} + c \times \text{Blogs} + d \times \text{Syllabus} + e \times \text{Books} \]
3. Webometrics for new research evaluations
3. Webometrics for new research evaluations

- Some organisations produce non-academic research & need impact evaluation
  - Think tanks, government departments, charities
- Target audience not academic
  - WoS/Scopus citations irrelevant (and almost non-existent)
  - May use LexisNexis media mentions
Web citation analysis

- Count mentions of report on the web
- Evidence of wider public and media interest
- Could focus on just blogs
- Can be automated (LexiURL Searcher)
- Lots of Spam
  - Need manual checking and content analysis for the best results (expensive)
Web citation analysis of online PDF and word documents

- Evidence of “professional impact”
  - Academic documents, newsletters, white papers & the grey literature
- Can be automated via Yahoo!/Live Search document type-specific queries (filetype:pdf)
- Variable quality documents
- High quality documents can be very interesting
NESTA Web Reports

- National Endowment for Science, Technology and the Arts
- Conducted twice-yearly for several years
- Evaluate the online impact of the most recent 20 reports
- Identify successful and unsuccessful reports
- Identify successful and unsuccessful types of report
<table>
<thead>
<tr>
<th>Citing document title and information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
</tr>
<tr>
<td>Conference paper</td>
</tr>
<tr>
<td>Cited NESTA document</td>
</tr>
<tr>
<td>Demanding Innovation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Citing document title and information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
</tr>
<tr>
<td>Journal article</td>
</tr>
<tr>
<td>Cited NESTA document</td>
</tr>
<tr>
<td>Demanding Innovation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Citing document title and information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Innovation Nation, DIUS, March 2009</td>
</tr>
<tr>
<td>Type</td>
</tr>
<tr>
<td>Research Report</td>
</tr>
<tr>
<td>Cited NESTA document</td>
</tr>
<tr>
<td>Demanding Innovation</td>
</tr>
</tbody>
</table>

*Full index contains 100s of records. It is provided for browsing by NESTA*
4. Webometrics for social science

Large-scale quantitative web measurements for social science research goals
Example 1: Blog searching

The graph is evidence that the issue started in the West at this date.

No other source of information will tell you that people were *not* thinking about the issue earlier!

= large scale quantitative Web measurements for social science research

Free at blogpulse.com
YouTube debates about Justin Bieber
The web contains a wide variety of web and web 2.0 content posted by many different people in many different formats.

Webometric methods can exploit this for evidence of a wide variety of types of impact.

New clients and social scientists that we can help!

Happier old clients!?


http://lexiurl.wlv.ac.uk  http://webometrics.wlv.ac.uk