Contents | Author index | Subject index | Search | Home

Proceedings of the Seventh International Conference on Conceptions of Library and Information Science—"Unity in diversity" — Part 2

Webometrics: emergent or doomed?

Mike Thelwall
Statistical Cybermetrics Research Group, School of Computing and Information Technology, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1SB, UK.

Abstract

Introduction. For some years bibliometric studies have identified webometrics as one of the largest information science fields. Nevertheless, most published webometric research is relatively theoretical and, as a new research field, seems unlikely to survive unless it is useful in some way. In general it is very important for academics to identify hot research topics and avoid unpromising ones and so methods are needed to help this identification process.
Method. This article uses citation analysis and a survey of webometricians to assess the extent to which webometrics has found applications outside of its parent discipline.
Results. The results suggest that there has been a turn towards applied webometrics with several externally-financed studies being contracted. Moreover, there is a significant amount of citation of webometrics research by disciplines outside information science, including computing, communication science and health. Nevertheless, it seems that the potential user base for current webometric techniques is wider still, creating a need for awareness-raising.
Conclusions. Whilst webometrics already has several claims to usefulness, there is still progress to be made.

CHANGE FONT

Introduction

webometrics, the quantitative study of Web-based phenomena (defined more precisely later) was born in 1997 (Almind and Ingwersen 1997) and was driven rapidly forward by a number of pioneering researchers and investigations (Aguillo 1998; Bar-Ilan 1999; Ingwersen 1998; Rodríguez and Gairín 1997; Rousseau 1997; Smith 1999) to establish itself as apparently the largest coherent field within library and information science (Åström 2007; Zhao and Strotmann 2008). There were three main directions for early research: link impact evaluation, link relationship mapping and search engine results analysis.

Link impact evaluation studies typically identified methods to count the number of hyperlinks to sets of academic-related Web sites or pages (e.g., universities, departments, open access journal Web sites) and evaluated whether the number of links found could reasonably be used to indicate the research impact of the target site or page (Ingwersen 1998; Smith 1999). The idea that counts of links to a Web site or page could indicate its research impact derived from hyperlinks being structurally similar to academic citations and the reasonable belief that a Web page or site targeted by many hyperlinks may well contain valuable information. This was principally encapsulated in Ingwersen's (1998) Web Impact Factor, the main variant of which defined the Web impact of a page/site/space as being the average number of links received per page from outside of the space measured. Despite the attractive simplicity of this impact factor and numerous studies using it and developing alternatives, the discovery that UK university Web sites attracted links in proportion to their size rather than the quality of information in their individual pages seriously undermined the main variant (Thelwall and Harries 2004) and it does not seem to have been used in webometric applications. In addition, studies of links to journal Web sites seem to have disappeared, probably because most journals now seem to have publisher Web sites, many of which have complex URLs that are difficult to use in webometrics. Studies of links to sets of whole university Web sites continue to be popular, but mainly outside of the developed nations (e.g., Kousha and Horri 2004; Qiu *et al.* 2004), perhaps because emerging academic Web spaces are more interesting.

Link relationship mapping investigations typically chart a coherent set of academic Web sites in two dimensions, either as a network diagram or as coordinates in a two-dimensional space using multi-dimensional scaling. The similarity measures used to construct the diagrams are either the number of links between pairs of Web sites or co-inlink counts: the number of Web pages linking to both target pages or sites (Björneborn and Ingwersen 2004). Early studies all mapped sets of university Web sites within a country or within Europe (Boudourides *et al.* 1999; Thelwall 2002) but later projects obtained more interesting results by mapping smaller-scale academic units, such as departments (Heimeriks *et al.* 2003; Li 2005; Tang and Thelwall 2003) or by mapping themed collections of commercial Web sites (Vaughan and Wu 2004).

The third early type of research, search engine results analysis, had the objective to assess the comprehensiveness and consistency of commercial search engine results, either because of their use for the raw data in much webometrics research or because of their wider importance as an information retrieval tool by Web users (Bar-Ilan 2004). Perhaps surprisingly, given that a significant minority of the planet's population uses Web search engines, this type of research has not expanded by has continued at a low, steady rate (e.g., Bar-Ilan 1999, 2004, 2004; Mettrop and Nieuwenhuysen 2001; Rousseau 1999; Uyar 2009a, 2009b; Vaughan and Thelwall 2004). Note that this area contains many individual papers by people who many not consider themselves webometricians but the papers can be regarded as webometrics in the bibliometric sense that their citations are predominantly from other webometric papers.

Later webometrics research included many studies of visualisations (Ortega *et al.* 2008; Ortega and Aguillo 2009) and covered an increasing variety of topics, such as longitudinal quantitative studies of Web pages (Koehler 2002, 2004), Web citation analysis (Kousha and Thelwall 2006, 2007; Vaughan and Shaw 2003, 2005), blog analysis (Smith 2007; Thelwall *et al.* 2007), business Web sites (Vaughan 2005; Vaughan *et al.* 2009), digital library analysis (Zuccala *et al.* 2007), information retrieval factors (Jepsen *et al.* 2004), social network site investigations (Ackland 2009; Thelwall 2008) (although these may not be seen as part of webometrics) and theoretical contributions (Bar-Ilan 2001; Björneborn 2006; Björneborn and Ingwersen 2004; Fry 2006; Thelwall 2006). Despite the bibliometric studies suggesting the importance of webometrics to information science (Åström 2007; Zhao and Strotmann 2008), it may soon be surpassed by other emerging topics, such as the h-index or Web 2.0, and so it is a suitable time to conduct a review of the field in order to assess whether it is delivering enough useful knowledge to safeguard its future within library and information science. More specifically, since arguably only search engine evaluation seems to offer value to general information science goals - because it is relevant to Web information seeking - this review focuses on applied webometrics. The objective is to survey current applications of webometrics in order to assess whether webometrics has long-term viability as a research field. The belief underpinning this is that new fields within library and information science, as a professionally-oriented discipline (Whitley 2000), are likely to whither and die without clear applications. From a wider perspective, this may shed light on the normal process of the birth and death of research fields and suggest reasons why particular fields can be successful or not.

Definitions and scope

Research fields or specialisms within a discipline are difficult to precisely define due to changes over time in topics and personnel and a lack of organised groupings and journals (Becher and Trowler 2001: 67). In addition, the perspective taken is important: is a field defined by the topic, the people who identify themselves as part of the field or by citation structures? A precise definition is essential here because of the need to identify applications of webometrics in other fields. The original and dominant definition of webometrics is, "The study of the quantitative aspects of the construction and use of information resources, structures and technologies on the Web drawing on bibliometric and informetric approaches." (Björneborn and Ingwersen 2004), and a later proposed definition is "the study of Web-based content with primarily quantitative methods for social science research goals using techniques that are not specific to one field of study" (Thelwall 2009).

If webometrics was defined to be the relevant works of people identifying themselves as to some extent webometricians then this might be restricted to information scientists like Aguillo, Almind, An, Barjak, Bar-Ilan, Björneborn, Fairclough, Fry, Harries, Heimeriks, Holmberg, Ingwersen, Kousha, Kretschmer, Lamirel, Li, Ortega, Payne, Prime, Rousseau, Smith, Stuart, Thelwall, Uyar, Vaughan, and Wilkinson, all of whom have first-authored a webometrics article. This would exclude the works of computer scientists and physicists that have conducted large-scale quantitative Web analyses, including the hyperlink power law research of Barabasi *et al.* (Barabási 2002) and the longitudinal search engine evaluations of Microsoft researchers (Fetterly *et al.* 2003). Similarly, it would exclude computer science "Web metrics" research that includes Web page similarity measurements and usage information (Dhyani *et al.* 2002). These exclusions could potentially also be argued for on a topic basis using the second webometrics definition since the more abstract research seems to lack a real social sciences goal. One area that would be excluded is sociology research using Richard Rogers' IssueCrawler link analysis software, which is a type of link relationship mapping but seems to have very little interaction with webometrics, at least in terms of shared citations. A number of studies have also investigated politics-related hyperlinks, such as relating to Korean politicians' blogs (Park *et al.* 2001), US election-related blogs (Foot and Schneider 2006), and extremist Web sites (Ackland and Gibson 2004). These would be excluded, even though some of the authors may regard themselves as webometricians, as they are not primarily information scientists. This is a grey area, however, due to the overlap between LIS and communication science (Paisley 1984). The current study is concerned with webometrics as an information science field and hence with research that is concerned with quantitative Web analysis and is conducted by information scientists identifying themselves as at least partly webometricians. This excludes the computer science, physics and sociology research discussed above but includes work by information scientists that might not consider themselves to be webometricians.

The next task is to identify applications of webometrics. Applied science has been defined to be research that is "directed towards a specific practical aim or objective" (Stokes 1997: 65) but this definition is inadequate here. There seem to be two clear-cut criteria that could be used to identify applied information science research, however. First, paid research conducted for an external organisation desiring scientific outputs is necessarily to some degree applied. Much information science applied research seems to fall into this category, with the "scientific outputs" probably tending to be commissioned reports. Second, information science methods used in other disciplines, even if for theoretical purposes, have applications in the sense that they are aiding an external endeavour. This criterion is relatively clear cut in the sense that it could be operationalised as published research cited outside of the recognised library and information science literature. A third type of value for a field would be in becoming core knowledge or

competence within this discipline. This is applied in the sense of educating the profession to do its job. It is less clear-cut, however, because even if extensive teaching of a topic was a practical metric, academic degrees necessarily contain a quantity of theory. As a result, in the current study, only the first two criteria are used to operationalise applications of webometrics.

Methods

There are many different ways in which research can be said to have impact. For example, a survey of health related research found four different types: future research; service changes; policy changes; general societal value (Kuruvilla *et al.* 2007; 2006). Of these, the final one is arguably the most important but the hardest to measure: in contrast future research impact can easily be estimated (Moed 2005) or argued for (e.g., Charness 1992) through citations, service changes could be measured by surveys and policy impacts could perhaps be estimated through surveys or citation analysis of policy documents. Others have emphasised the importance of research at times of key decision making (Lavis *et al.* 2003). In general, however, it seems clear that it is reasonable to assess a range of indicators of success other than the (arguably) final societal impact goal of modern research.

Studies of the impact of an individual scientific field often take a narrow perspective, such as evaluating key journal impact factors (e.g., Wikström 2009), which is not appropriate for webometrics since it is published primarily in general library and information science journals. Although there have been many investigations of research applications (Cohen *et al.* 2002), these seem to take the narrow perspective of seeking specific non-academic uses (Estublier *et al.* 2002; Sherman and Cohn 1989) and no previous study seems to have evaluated an entire field for applications and hence new techniques were needed. Two methods were used to identify applications of webometrics: citation analysis and a survey of active webometrics researchers. A third method was also used to identify cases where webometrics could have been used but was not.

The citation analysis method sought to identify published articles that were not in LIS journals but which cited webometrics research. The LIS discipline has been identified as not widely cited by other disciplines (Tang 2004) and so citations from outside are particularly valuable. Papers citing webometrics research were sought using Scopus rather than the Web of Science, due to its superior search facilities for this purpose and adequate coverage of the topic area. The following process was used.

1. The articles of each of the webometricians in the list above (the second paragraph of the Definitions and scope section) were searched for and saved as a list.
2. Each of the above lists was manually filtered for articles by other authors, articles unrelated to webometrics, and articles previously found for another author.
3. A cited reference search was used to identify articles citing any article in the above lists.
4. Articles from recognised information science journals and conferences were manually excluded from the above list.
5. The above lists were combined and duplicate citations were manually excluded.

The above steps gave 461 citations that were apparently non-information science articles that cited webometrics research. Note that self-citations are included in the sense of studies published outside library and information science in which at least one author cites their own work. These were then manually classified by topic and discipline and qualitatively reviewed.

The survey of webometricians was conducted by e-mailing all people in the list above, except for those already at Wolverhampton (Fairclough, Harries, Payne, Price, Stuart, Thelwall, Wilkinson) or deceased (Almind). Each e-mail requested information about any funded research they had conducted using webometrics, other than research council-funded research (see Appendix 1).

The third approach is to identify articles that could potentially cite webometrics but do not. These studies therefore bypass webometrics either through not knowing its contribution or by regarding it as unimportant. This section focuses on link analysis through keyword searches for "hyperlink analysis" and "link analysis" in Google Scholar and discarding articles citing webometrics research. Google Scholar was used, despite its limitations in accuracy (Jascó 2005; Mayr and Walter 2007), to get wide coverage of different disciplines

Results 1: Commissioned webometrics research

The answers from the e-mail survey about externally-funded webometrics research were combined by apparent type to facilitate discussion. Each section below discusses one type.

Standard academic webometric analysis

It seems that only one main standard webometric report has been commissioned. It was standard in the sense of being primarily designed to calculate webometric indicators for an academic Web space rather than using the webometric indicators as an intermediate step to an additional goal. This contract was Iberoamerican Spaces 2008: Innovations, Science and Technology funded by the Comisión Económica para América Latina (CEPAL), United Nations Economic Commission for Latin America and the Caribbean, UNECLAC, Business and Productive Division (Ortega and Aguillo 2009). The purpose was to analyse the Iberoamerican Web space primarily through network visualisations and social network analysis statistics, and it identified factors such as linguistic divides and national subnetworks.

A set of indicators for European research interlinking has also been calculated for the European Community Science, Technology and Competitiveness

Key Figures Report 2008 (Directorate-General for Research 2008) (Aguillo, Ortega), giving webometrics a showcase within the European science policy community. The outputs included a geographic map of Europe showing the main directions of international academic links.

Taken together, these two studies give recognition that academic hyperlinks are viewed as important in their own right in some international science policy circles.

International research communication and mobility

The Directorate-General for Research of the European Commission has commissioned a number of studies containing a webometric component that were designed to shed light on aspects of internationalism in research. Project NetReAct (Barjak and Robinson 2007; Barjak and Thelwall 2008) used hyperlinks to life science research group Web sites in nine European countries in order to get an indication of the degree of internationalism in each country. The assumption for this was that the proportion of hyperlinks originating in foreign countries would be a reasonable approximate indicator of the degree of internationalism of the target country's academic life science research. The same link data was also used to identify the most prominent life science research groups in each country for interviews and questionnaires, with the assumption that better research groups would tend to attract more links. The latter shows that webometrics can play a helper role in more traditional research methods. Project RESCAR (Barjak, Robinson, Thelwall) had similar goals to NetReAct but for the wider remit of social sciences and engineering.

The above two projects derived from the same initiative and the funding of the second to the same research team was an indicator of the success of the first. At least two further projects have given results for the Commission with related goals, however. A separate project, Rindicate, (Thelwall, Klitkou, Arnold, Stuart, Vincent) was a more exploratory initiative, charged with demonstrating that webometrics could provide useful indicators about international technology networks in emerging transdisciplinary research fields, such as second generation boifuels.

In addition to the above projects, a webometrician has been employed by the Directorate-General for Research: Gaston Heimeriks at the Institute for Prospective Technological Studies in Seville. Overall, this points to a significant uptake of webometrics by the Directorate-General but this uptake does not seem to have been replicated in any other similar national or international body. This is perhaps because webometrics for research seems particularly relevant for its international coverage and relative ease with which it can be applied to many different countries compared to bibliometrics. As a result it is most relevant to powerful international research monitoring and funding agencies, of which the Directorate-General is the prime example.

Online impact of digital resources

The UK Joint Information Systems Council has funded two projects that use webometrics to help evaluate the impact of collections of digital resources. The Council is in charge of the higher education infrastructure in the UK and is not a research funding agency but does fund projects including an element of research. The first project, User Needs and Potential Users of Public Repositories: An Integrated Analysis, (Thelwall, Zuccala, Oppenheim, Dhiensa) developed a standard set of webometric methods to evaluate digital repositories, including inlink evaluations and colink diagrams. Although the Council had not specified webometrics in its funding call for repository evaluation methods, it selected the webometric proposal as a solution to its needs.

The second Joint Information Systems Council project, TIDSR Toolkit (microsites.oii.ox.ac.uk/tidsr/), (Meyer, Thelwall, Eccles, Stuart) had the objective to develop an integrated method to evaluate the digital archives funded by the Council. Webometrics was proposed as part of the toolkit, along with interviews, surveys and Web log file analysis.

Although digital repositories are an international phenomenon, it seems that only the Council has funded evaluation research including webometrics. The clear reason for this difference seems to be the availability of UK webometrics researchers willing to bid for the project. Since webometrics is a logical, but not necessary, solution to the online evaluation problem, this seems to be the most likely explanation.

Note that this research relates somewhat to an earlier webometric study of the characteristics of online scholarly documents (Jepsen *et al.* 2004).

Webometrics ranking of world universities

The *Ranking Web of World Universities* (Aguillo, *et al.* 2006) is not an externally-funded initiative but is included here because it is a high profile webometrics initiative aimed primarily at universities rather than information scientists. It ranks world universities' Web presences based upon a weighted sum of webometric indicators, such as the number of rich files (e.g., pdf) in their Web sites and the number of hyperlinks pointing to their Web sites (Aguillo *et al.* 2006). Evidence of its impact includes the award of an honorary doctorate to its initiator Aguillo, as well as over 16,000 Web pages linking to it, according to Yahoo, in 15 December 2009.

The lack of further similar rankings probably reflects the scarcity of expertise and knowledge to create it and its almost universal coverage, making competitors redundant. The value of the ranking to the field of webometrics is hard to judge but it seems to have given webometrics a much higher visibility in academia than it would otherwise have.

Results 2: Citations to webometrics research from outside library and information science

This section surveys research fields or types of research outside library and information science that have cited webometrics research. To qualify, the research area must make significant use of the webometrics research rather than just giving it a perfunctory mention. This section is organised by discipline but discusses the types of webometrics research cited.

Overall results

Figure 1 summarises the results of the classification of the non-library and information science sources of articles citing webometrics. The main sources are all related to information science. The discipline of computing has many overlaps with LIS, including the topics of information retrieval, information systems, human factors in computing, digital libraries and Web technologies. These overlaps also allow some articles that could be classified as webometrics to be published in computing conferences, including: *'A Webomatrics Analysis between Taiwan and China Academic Web by Sitation Analysis'* in Proceedings of the 2008 International Conference on Semantic Web and Web Services, and *'Agreeing to disagree: Search engines and their public interfaces'* in Proceedings of the ACM International Conference on Digital Libraries. In the latter case, the conference could arguably also be classified as a library and information science conference. The overlap is also reflected in the inclusion of some papers authored by webometricians (Bar-Ilan, Björneborn, Thelwall) but published in computer science venues. Interestingly, although there is significant computer science (information retrieval) research into search engine design from the main commercial search engines, this seems never to cite webometrics although academic search engine research and other articles discussing the role of search engines do sometimes cite it (e.g., *'Web evolution and incremental crawling'*).
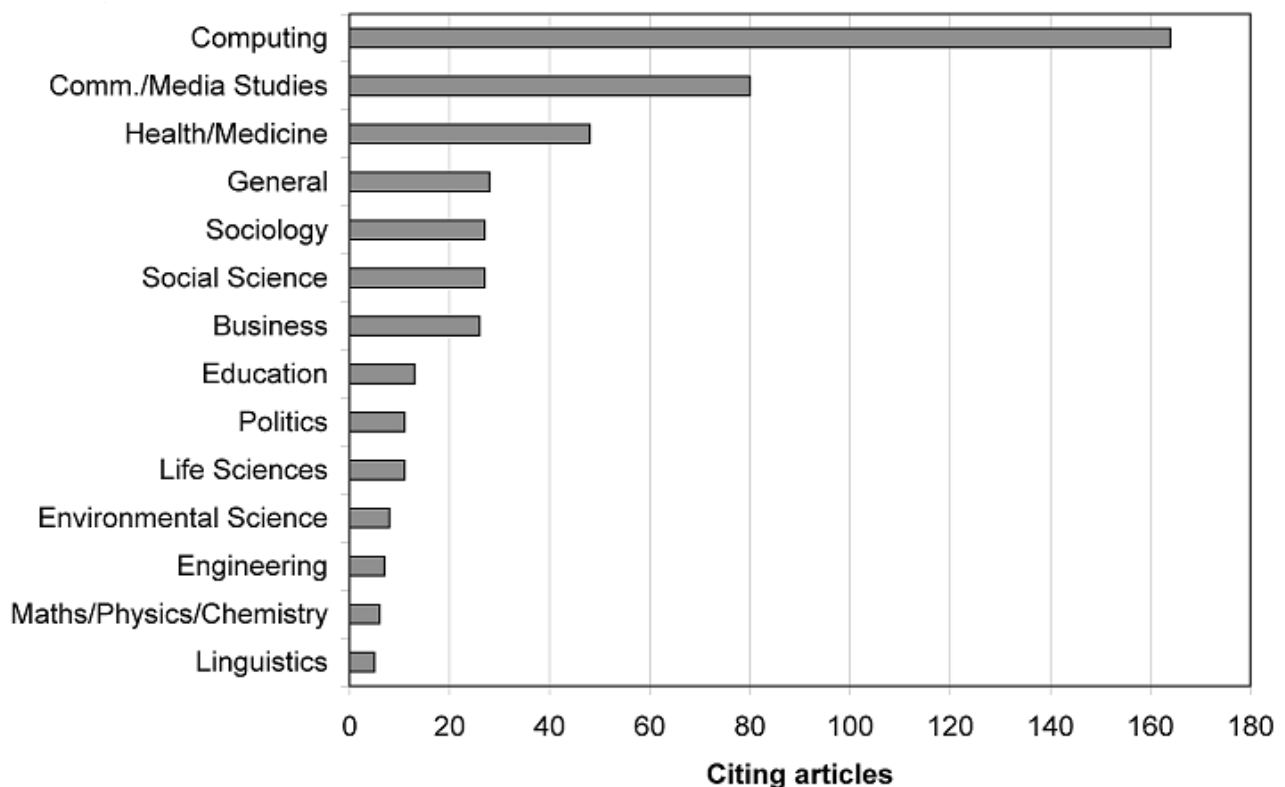

Figure 1. Disciplines of non-library and information science articles citing webometrics, classified by the source journals and conferences.

Communication and media studies also have significant overlaps with library and information science (Paisley 1984) and webometrics authors that publish in its journals (Bar-Ilan, Shaw, Thelwall, Vaughan). It also has some authors that could be described as webometricians because of their focus on quantitative Web methods (Ackland, Elmer, Park and perhaps Foot). The main sources in this category are the journals New Media and Society, Journal of Computer-Mediated Communication and First Monday (a general Internet journal but with a communication/media focus). The use of hyperlinks to investigate politics or other sets of Web sites is a common theme in the articles citing webometrics (e.g., *'The north/south divide in NGO hyperlink networks'*, *'The Italian extreme right on-line network: An exploratory study using an integrated social network analysis and content analysis approach'*), and webometrics articles investigating blogs or search engines also receive some citations (e.g., *'Election bloggers: methods for determining political influence'*, *'Blogging practices: an analytical framework'*, *'The social, political, economic, and cultural dimensions of search engines: an introduction'*).

Within communication and media studies, social network site research is also represented (e.g., *'Examining social media usage: yechnology clusters and social network site membership'*, *'An anonymous social network site to share pictures'*). The rise of social network sites saw a rapid growth in research into how they are used, who uses them and their impact on society. The leading light in this trend has been danah boyd (e.g., boyd and Ellison 2007) and the research seems to be concentrated in media studies, communication studies and sociology and is represented at many conferences in these fields. There has also been related research in health and psychology (Hinduja and Patchin 2008; Sheldon 2008). Although there have been only a few webometrics studies of social network sites, using a data mining approach rather than link analysis, these have been cited for the range of basic data provided about social network sites, rather than for methods or theoretical insights.

There is also a logical connection between library and information science and the areas of health and medicine, through the field of medical informatics, the existence of which signals the particular importance of the field. Although information science is relevant to all academic fields because all require information seeking, there do not seem to be equivalent hybrid fields for all disciplines. There is only one webometrics author represented in the health and medicine set (Aguillo), however, suggesting that the overlap occurs mainly at the topic level. A wide variety of journals are represented however, from those with no obvious connection like the *International Journal of Colorectal Disease* to a few with an informatics connection, such as *Journal of Medical Internet Research*. One common topic forming an obvious crossover is a concern with online medical information searching and quality (e.g., *'Searching of information in evidence-based pediatrics (I): 'Infoxicacion' and Internet'*, *'Can examination of WWW usage statistics and other indirect quality indicators help to distinguish the relative quality of medical Websites?'*, *'Analysis of governmental Web sites on food safety issues: A global perspective'*, *'Gastrointestinal cancer Web sites: How do they address patients' concerns?'*). The most common topic seems to be bibliometrics, however (e.g., *'Citations to trials of nicotine replacement therapy were biased toward positive results and high-impact-factor journals'*), with webometric literature cited for its evaluations of Google Scholar (e.g., *'Acimed in Scholar Google: a citation analysis of the Cuban Journal of Health Information and Communication Professionals'*, *'Comparisons of citations in Web of science, Scopus, and Google Scholar for articles published in general medical journals'* or for its relevance to Web references (e.g., *'Web Citations in the Nursing Literature: How Accurate Are They?'*). Overall, health and medicine seem be particularly fertile areas for a range of webometrics research. Over a third of the articles are in Spanish (17/48), suggesting that webometrics is most visible amongst Spanish-language medical and health researchers.

Like communication/media studies, the sociology citations are often related to political or issue hyperlink networks (e.g., *'A hyperlink network analysis of citizen blogs in South Korean politics'*, *'Cyberplace and cyberspace: Two approaches to analyzing digital intercity linkages'* and search engines (e.g., *'Search engines: Terrorism's killer app'* but there are also more general theoretical articles (e.g., *'Opening the black box of link formation: Social factors underlying the structure of the Web'*, *'New techniques in online research: Challenges for research ethics'*). Social science citations display a similar pattern, with webometrics authors contributing many of them (11/27).

The last area with at least 5% of the citations is business. Most of the citing articles covered online business and particularly marketing and e-commerce (e.g., *'Internet-based scanning of the competitive environment'*, *'Competition in internet retail markets: The impact of links on Web site traffic'*, *'Website adoption and sales performance in Valais' hospitality industry'*). The webometrics articles seemed to supply background information for these. A few bibliometrics articles were also present, discussing Google Scholar or online sources (e.g., *'Google scholar visibility and tourism journals'*, *'A citation-based ranking of the business ethics scholarly journals'*, *'Ranking, rating and scoring of tourism journals: Interdisciplinary challenges and innovations'*).

Research areas bypassing webometrics

The Google Scholar searches for webometrics and link analysis studies not citing webometrics produced articles mainly from computer science and sociology. Some were isolated studies but others formed coherent bodies of work. The latter are discussed below, broadly by discipline. In addition, physics research known to the author but not found by the searches is discussed.

Computer science

Many different Web-related computer science topics use link analysis but do not cite webometrics research. The topics include search engine ranking algorithms, Web information retrieval, question answering, link analysis, link mining, crawling, Web page classification, and spam filtering. This shows that there is a significant body of link analysis research that does not need webometrics. This could be because the computer science research is typically concerned with algorithm construction (e.g., *'Discovering authorities in question answer communities by using link analysis'*, *'Link analysis ranking: algorithms, theory, and experiments'*) whereas webometrics normally analyses the outputs of algorithms, giving a different perspective. In addition, there is some computer science research into traditional LIS topics like scholarly communication (e.g., *'Semantic Web link analysis to discover social relationships in academic communities'*) and collaborative knowledge building (e.g., *'Link analysis for collaborative knowledge building'*).

Sociology

A type of link analysis has been successfully promoted by Richard Rogers, a sSociologist at the University of Amsterdam. His variant focuses on *'issue networks'*, collections of networks around a topic of interest. His IssueCrawler software supports the generation of networks using hyperlink data from any (themed) set of Web sites entered by the user. This approach is described with examples in two books (Rogers 2002, 2005) and has been adopted by others to research topics such as the Web usage patterns of diaspora communities (Kissau and Hunger 2008). Relatively little of this research seems to be published in recognised academic journals, however, perhaps due to its humanities-oriented nature.

It is not clear why issue network research does not cite webometrics from a methodological perspective because much webometrics research relates to the accuracy of data from commercial search engines or the validity of using links in social science research. The reason may be because of the humanities focus on exploration and idea generation rather than a social science focus on data validity and methodological robustness.

Physics and complexity science

Much physics research is concerned with fitting simple mathematical models to large-scale data. In around 1999 a number of physicists noticed that links to Web pages did not have a standard Gaussian distribution, showing that they approximately fitted a power law (Adamic, Barabasi). webometrics research was probably not cited because it was not known or did not provide relevant methods due to its reliance upon commercial search engines (e.g., Ingwersen). In fact an early webometrics article (Rousseau 1997) seems to have been the first to observe power law phenomena in the Web, albeit on a

small scale, but this went unnoticed or unrecognised.

This physics research was closely related to some computer science research into power laws (e.g., Pennock, Flake, Lawrence, Glover, and Giles 2002). In fact a small computer science/physics research field of Web dynamics arguable began to subsequently form (e.g., book) with some participation by webometricians Judit Bar-Ilan who injected a search engine evaluation perspective. Most of the research was quite theoretical and abstract, however, often simulating Web link growth and attempting to build a simple mathematical model for it. This abstract approach perhaps rendered the more concrete webometrics research less relevant, and since commercial search engines were rarely used for raw data, the search engine evaluation webometrics was not found useful. Nevertheless, some webometrics contributions, such as the alternative document model concept for link counting could arguably have made a contribution to generating cleaner link data but was not used. Perhaps the computer scientists and physicists did not know of its existence and looked for relevant prior work mainly from within their own fields or did not regard webometrics as meeting other criteria for citing.

Limitations and Conclusions

The methods used in this article have a number of limitations. First, there may be researchers outside the sample e-mailed who have conducted commissioned research. This may have occurred in non-English speaking nations with developed national literatures outside the Web of Science (e.g., China), or by research teams not emphasising academic publications for example. Similarly, for the citation analysis, the coverage of Scopus is incomplete and therefore there will be citations that have been missed and possibly even whole areas of science. In particular, there may be national literatures or even whole disciplines that have been missed. In addition, the definition of webometrics selected, based on researchers more than content, would not necessarily be universally accepted as the best choice.

From a wider perspective, the methods used (e-mails and citation analysis) seem to have worked to give insights into the extent to which webometrics has found applications but perhaps more detailed methods would have helped. In particular, some kind of benchmarking of citations would have helped to decide whether webometrics had attracted an unusually high or low number of citations from outside library and information science.

From the limitations, the findings should be interpreted as the minimum impact of webometrics rather than an estimate of its actual impact. Nevertheless, it is clear that webometrics has gained a small foothold in the European Union and Iberoamerica in terms of researchers being commissioned for applied studies. In addition, webometrics research is having some impact in other disciplines in the sense of attracting citations. These include the cognate disciplines of computer science and communication (and media) studies as well as the broad area of health and medicine. In contrast, related sociology and physics/complexity research seems to be bypassing webometric studies, although this is also true for some computer science research too. In summary, webometrics is having some impact outside of its field but has not influenced some areas that it perhaps should have. Most promising for the future may be health-related research because its lack of cognate connection with information science, except through medical informatics, means that there may be little awareness of it within the webometrics community. More health-related webometrics research is therefore recommended in the future to help strengthen this connection.

The results are probably insufficient to give a definitive answer about whether webometrics is likely to have a future, however, since this is a qualitative question and the results are not clear-cut enough to demonstrate clear and lasting applications. Instead, webometrics may still be at a vulnerable growing stage with too few clear applications to make a strong case for its future value and vitality.

About the author

Mike Thelwall is a Professor of Information Science in the School of Computing & IT, University of Wolverhampton, UK. He received his Bachelor's degree in Mathematics and his PhD from Lancaster University, UK. He can be contacted at m.thelwall@wlv.ac.uk

References

- Ackland, R. (2009). Social network services as data sources and platforms for e-researching social networks. *Social Science Computer Review*, **27**(4), 481-492.
- Ackland, R. & Gibson, R. (2004). *Mapping political party networks on the WWW.* Paper presented at the Australian Electronic Governance Conference, 14-15 April, 2004. Retrieved 12 December, 2009 from http://voson.anu.edu.au/papers/political_networks.pdf. (Archived by WebCite® at http://www.Webcitation.org/5st786zmS)
- Aguillo, I.F. (1998). STM information on the Web and the development of new Internet R & D databases and indicators. In D. Raitt, (Ed.). *Proceedings, Online Information 98.* (pp. 239-243). London: Learned Information.
- Aguillo, I.F., Granadino, B., Ortega, J.L. & Prieto, J.A. (2006). Scientific research activity and communication measured with cybermetrics indicators. *Journal of the American Society for Information Science and Technology*, **57**(10), 1296-1302.
- Almind, T.C. & Ingwersen, P. (1997). Informetric analyses on the World Wide Web. Methodological approaches to 'webometrics'. *Journal of Documentation*, **53**(4), 404-426.
- Åström, F. (2007). Changes in the LIS research front: time-sliced cocitation analyses of LIS journal articles, 1990-2004. *Journal of the American Society for Information Science and Technology*, **58**(7), 947-957.
- Bar-Ilan, J. (1999). *Search engine results over time - a case study on search engine stability*. Retrieved 26 January, 2006, from

http://www.cindoc.csic.es/cybermetrics/articles/v2i1p1.html (Archived by WebCite$^®$ at http://www.Webcitation.org/5st7fEK0F)

- Bar-Ilan, J. (2001). Data collection methods on the Web for informetric purposes - a review and analysis. *Scientometrics*, **50**(1), 7-32.
- Bar-Ilan, J. (2004). Search engine ability to cope with the changing Web. In M. Levene & A. Poulovasilis (Eds.), *Web dynamics* (pp. 195-215). Berlin: Springer-Verlag.
- Bar-Ilan, J. (2004). The use of Web search engines in information science research. *Annual Review of Information Science and Technology*, **38**, 231-288.
- Barabási, A. L. (2002). *Linked: The new science of networks*. Cambridge, MA: Perseus Publishing.
- Barjak, F. & Robinson, S. (2007). International collaboration, mobility and team diversity in the life sciences. Impact on research performance. *Social Geography Discussions*, **3**(1), 121-157.
- Barjak, F. & Thelwall, M. (2008). A statistical analysis of the Web presences of European life sciences research teams. *Journal of the American Society for Information Science and Technology*, **59**(4), 628-643.
- Becher, T. & Trowler, P. (2001). *Academic tribes and territories (2ed)*. Milton Keynes, UK: Open University Press.
- Björneborn, L. (2006). 'Mini small worlds' of shortest link paths crossing domain boundaries in an academic Web space. *Scientometrics*, **68**(3), 395-414.
- Björneborn, L. & Ingwersen, P. (2004). Toward a basic framework for webometrics. *Journal of the American Society for Information Science and Technology*, **55**(14), 1216-1227.
- Boudourides, M. A., Sigrist, B. & Alevizos, P. D. (1999). *webometrics and the self-organization of the European information society*. [Draft report on Task 2.I of the SOEIS Project, Rome meeting, June 17-19.] Retrieved 29 May 2002 from http://hyperion.math.upatras.gr/webometrics (Archived by WebCite$^®$ at http://www.Webcitation.org/5st86DA8h)
- Boyd, D. & Ellison, N. (2007). Social network sites: definition, history, and scholarship. *Journal of Computer-Mediated Communication*, **13**(1), Retrieved 7 May, 2009 from http://jcmc.indiana.edu/vol13/issue2001/boyd.ellison.html. (Archived by WebCite$^®$ at http://www.Webcitation.org/5st8mJDcP)
- Charness, N. (1992). The impact of chess research on cognitive science. *Psychological Research*, **54**(1), 4-9.
- Cohen, W.M., Nelson, R.R. & Walsh, J. P. (2002). Links and impacts: the influence of public research on industrial R&D. *Management Science*, **48**(1), 1-23.
- Dhyani, D., Ng, W. & Bhowmick, S. (2002). A survey of Web metrics. *ACM Computing Surveys*, **34**(4), 469-503.
- European Commission. *Directorate-General for Research*. (2008). *A more research-intensive and integrated European Research Area: science, technology and competitiveness key figures report 2008/2009*. Brussels: European Commission. Retrieved 28 October0, 2009 from http://ec.europa.eu/research/era/pdf/key-figures-report2008-2009_en.pdf.
- Estublier, J., Leblang, D., Clemm, G., Conradi, R., Tichy, W., van der Hoek, A., *et al.* (2002). Impact of the research community on the field of software configuration management: summary of an impact project report. *ACM SIGSOFT Software Engineering Notes*, **27**(5), 31-39.
- Fetterly, D., Manasse, M., Najork, M. & Wiener, J. (2003). A large-scale study of the evolution of Web pages. In *Proceedings of the 12th International World Wide Web Conference, Budapest, Hungary, May 2003* Retrieved 20 September, 2010 from http://research.microsoft.com/en-us/projects/pageturner/p97-fetterly.pdf (Archived by WebCite$^®$ at http://www.Webcitation.org/5st9uHuCB).
- Foot, K. & Schneider, S. (2006). *Web campaigning*. Cambridge, MA: The MIT Press.
- Fry, J. (2006). Studying the scholarly Web: how disciplinary culture shapes online representations. *Cybermetrics*, **10**(1), paper 2. Retrieved 20 September, 2010 from http://www.cindoc.csic.es/cybermetrics/articles/v10i1p2.html (Archived by WebCite$^®$ at http://www.Webcitation.org/5stA9mMLT).
- Heimeriks, G., Hörlesberger, M. & van den Besselaar, P. (2003). Mapping communication and collaboration in heterogeneous research networks. *Scientometrics*, **58**(2), 391-413.
- Hinduja, S. & Patchin, J. W. (2008). Personal information of adolescents on the Internet: a quantitative content analysis of MySpace. *Journal of Adolescence*, **31**(1), 125-146.
- Ingwersen, P. (1998). The calculation of Web Impact Factors. *Journal of Documentation*, **54**(2), 236-243.
- Jascó, P. (2005). Google Scholar: the pros and the cons. *Online Information Review*, **29**(2), 208-214.
- Jepsen, E. T., Seiden, P., Ingwersen, P. & Björneborn, L. (2004). Characteristics of scientific Web publications: preliminary data gathering and analysis. *Journal of the American Society for Information Science and Technology*, **55**(14), 1239-1249.
- Kissau, K. & Hunger, U. (2008). Political online-participation of migrants in Germany. *German Policy Studies*, **4**(4), 5-31. Retrieved 20 September, 2010 from http://www.spaef.com/file.php?id=992. (Archived by WebCite$^®$ at http://www.Webcitation.org/5stAUC5M7)
- Koehler, W. (2002). Web page change and persistence - a four-year longitudinal study. *Journal of American Society for Information Science*, **53**(2), 162-171.
- Koehler, W. (2004). A longitudinal study of Web pages continued: a report after six years. *Information Research*, **9**(2), paper 174. Retrieved 20 September, 2007 from http://informationr.net/ir/9-2/paper174.html (Archived by WebCite$^®$ at http://www.Webcitation.org/5stAh7wF6)
- Kousha, K. & Horri, A. (2004). The relationship between scholarly publishing and the counts of academic inlinks to Iranian university Web sites: Exploring academic link creation motivations. *Journal of Information Management and Scientometrics*, **1**(2), 13-22.
- Kousha, K. & Thelwall, M. (2006). Motivations for URL citations to open access library and information science articles. *Scientometrics*, **68**(3), 501-517.
- Kousha, K. & Thelwall, M. (2007). Google Scholar citations and Google Web/URL citations: A multi-discipline exploratory analysis. *Journal of the American Society for Information Science and Technology*, **58**(7), 1055 -1065.

- Kuruvilla, S., Mays, K. & Walt, G. (2007). Describing the impact of health services and policy research. *Journal of Health Services & Research Policy*, **12**(1), 23-37.
- Kuruvilla, S., Mays, N., Pleasant, A. & Walt, G. (2006). Describing the impact of health research: a Research Impact Framework. *BMC Health Services Research*, **6**, 134. Retrieved 20 September, 2010 from http://www.biomedcentral.com/1472-6963/6/134 (Archived by WebCite® at http://www.Webcitation.org/5stAyZkRA)
- Lavis, J., Ross, S., McLeod, C. & Gildiner, A. (2003). Measuring the impact of health research. *Journal of Health Services Research Policy*, **8**, 165-170.
- Li, X. (2005). *National and international university departmental Web site interlinking: a webometric analysis*. Unpublished doctoral dissertation, University of Wolverhampton, Wolverhampton, UK .
- Mayr, P. & Walter, A.K. (2007). An exploratory study of Google Scholar. *Online Information Review*, **31**(6), 814-830.
- Mettrop, W. & Nieuwenhuysen, P. (2001). Internet search engines - fluctuations in document accessibility. *Journal of Documentation*, **57**(5), 623-651.
- Moed, H.F. (2005). *Citation analysis in research evaluation*. New York, NY: Springer.
- Ortega, J.L. & Aguillo, I. (2009). Structural analysis of the Iberoamerican academic Web. *Revista Espanola de Documentacion Cientifica*, **32**(3), 51-65.
- Ortega, J. L., Aguillo, I., Cothey, V. & Scharnhorst, A. (2008). Maps of the academic Web in the European Higher Education Area: an exploration of visual Web indicators. *Scientometrics*, **74**(2), 295-308.
- Ortega, J. L. & Aguillo, I. F. (2009). Mapping world-class universities on the Web. *Information Processing & Management*, **45**(2), 272-279.
- Paisley, W. (1984). Communication in the communication sciences. *Progress in Communication Sciences*, **V**, 1-43.
- Park, H. W., Barnett, G. A. & Kim, C. S. (2001). Internet communication structure in Korean National Assembly: a network analysis. *Korean Journal of Journalism & Communication Studies, Special English Edition*, 185-203.
- Pennock, D., Flake, G. W., Lawrence, S., Glover, E. J. & Giles, C. L. (2002). Winners don't take ng the competition for links on the Web. *Proceedings of the National Academy of Sciences*, **99**(8), 5207-5211.
- Qiu, J. P., Chen, J. Q. & Wang, Z. (2004). An analysis of backlink counts and Web Impact Factors for Chinese university Websites. *Scientometrics*, **60**(3), 463-473.
- Rodríguez i Gairín, J. M. (1997). Valorando el impacto de la información en Internet: AltaVista, el "Citation Index" de la Red. [Assessing the impact of information on the Internet: AltaVista, the "Citation Index" of the Network] *Revista Española de Documentación Científica*, **20**(2), 175-181.
- Rogers, R. (2002). Operating issue networks on the Web. *Science as Culture*, **11**(2), 191-214.
- Rogers, R. (2005). *Information politics on the Web*. Cambridge, MA: MIT Press.
- Rousseau, R. (1997). Sitations: an exploratory study. *Cybermetrics*, **1**(1), Retrieved July 25, 2006 from: http://www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html (Archived by WebCite® at http://www.Webcitation.org/5stBoPIrC).
- Rousseau, R. (1999). Daily time series of common single word searches in AltaVista and NorthernLight. *Cybermetrics*, **2**/3, Retrieved July 25, 2006 from http://www.cindoc.csic.es/cybermetrics/articles/v2i1p2.html. (Archived by WebCite® at http://www.Webcitation.org/5stC0kSKl)
- Sheldon, P. (2008). The relationship Between unwillingness-to-communicate and students Facebook use. *Journal of Media Psychology: Theories, Methods, and Applications*, **20**(2), 67-75.
- Sherman, L. W. & Cohn, E. G. (1989). The impact of research on legal policy: the Minneapolis domestic violence experiment. *Law & Society Review*, **23**(1), 117-144.
- Smith, A. G. (1999). A tale of two Web spaces; comparing sites using Web Impact Factors. *Journal of Documentation*, **55**(5), 577-592.
- Smith, A. (2007). Issues in 'blogmetrics' - case studies using BlogPulse to observe trends in Weblogs. In Daniel Torres-Salinas and Henk F. Moed (Eds.), *Proceedings of 11th International Conference of the International Society for Scientometrics and Informetrics* (pp. 726-730). Madrid: Centre for Scientific Information and Documentation.
- Stokes, D. E. (1997). *Pasteur's quadrant: Basic science and technological innovation*. Washington, DC: Brookings Institution.
- Tang, R. (2004). Visualizing interdisciplinary citations to and from information and library science publications. In *Eighth International Conference on Information Visualisation (IV'04)*. (pp. 972-977). Washington, DC: IEEE Computer Society.
- Tang, R. & Thelwall, M. (2003). U.S. academic departmental Web-site interlinking: disciplinary differences. *Library & Information Science Research*, **25**(4), 437-458.
- Thelwall, M. (2002). An initial exploration of the link relationship between UK university Web sites. *ASLIB Proceedings*, **54**(2), 118-126.
- Thelwall, M. (2006). Interpreting social science link analysis research. A theoretical framework. *Journal of the American Society for Information Science and Technology*, **57**(1), 60-68.
- Thelwall, M. (2008). Social networks, gender and friending: an analysis of MySpace member profiles. *Journal of the American Society for Information Science and Technology*, **59**(8), 1321-1330.
- Thelwall, M. (2009). *Introduction to webometrics: Quantitative Web research for the social sciences*. New York, NY: Morgan & Claypool.
- Thelwall, M., Byrne, A. & Goody, M. (2007). Which types of news story attract bloggers? *Information Research*, **12**(4), Retrieved 16 October, 2007 from http://informationr.net/ir/12-4/paper327.html (Archived by WebCite® at http://www.Webcitation.org/5stGwALA4).
- Thelwall, M. & Harries, G. (2004). Do better scholars' Web publications have significantly higher online impact? *Journal of American Society for Information Science and Technology*, **55**(2), 149-159.
- Uyar, A. (2009a). Google stemming mechanisms. *Journal of Information Science*, **35**(5), 499-514.

- Uyar, A. (2009b). Investigation of the accuracy of search engine hit counts. *Journal of Information Science*, **35**(4), 469-480.
- Vaughan, L. (2005). Exploring Website features for business information. *Scientometrics*, **61**(3), 467-477.
- Vaughan, L., Du, J. & Tang, J. (2009). *Constructing business profiles based on keyword patterns on Websites.* PowerPoint presentation for the 12th International Conference of the International Society for Scientometrics and Informetrics.
- Vaughan, L. & Shaw, D. (2003). Bibliographic and Web citations: what is the difference? *Journal of the American Society for Information Science and Technology*, **54**(14), 1313-1322.
- Vaughan, L. & Shaw, D. (2005). Web citation data for impact assessment: a comparison of four science disciplines. *Journal of the American Society for Information Science & Technology*, **56**(10), 1075-1087.
- Vaughan, L. & Thelwall, M. (2004). Search engine coverage bias: evidence and possible causes. *Information Processing & Management*, **40**(4), 693-707.
- Vaughan, L. & Wu, G. (2004). Links to commercial Websites as a source of business information. *Scientometrics*, **60**(3), 487-496.
- Whitley, R. (2000). *The intellectual and social organization of the sciences.* 2nd ed. Oxford: Oxford University Press.
- Wikström, M. A. (2009). *Two of the "new" sciences - nanomedicine and systems biology in the United States*. Stockholm: VINNOVA, Swedish Governmental Agency for Innovation Systems. Retrieved 5 January, 2010 from http://www.vinnova.se/upload/EPiStorePDF/va-09-16.pdf. (Archived by WebCite® at http://www.Webcitation.org/5stHX0fJV)
- Zhao, D. & Strotmann, A. (2008). Information science during the first decade of the Web: An enriched author co-citation analysis. *Journal of the American Society for Information Science & Technology*, **59**(6), 916-937.
- Zuccala, A., Thelwall, M., Oppenheim, C. & Dhiensa, R. (2007). Web intelligence analyses of digital libraries: A case study of the National electronic Library for Health (NeLH). *Journal of Documentation*, **64**(3), 558-589.

How to cite this paper

Thelwall, M. (2010). "Webometrics: emergent or doomed?" *Information Research*, **15**(4) colis713. [Available at http://InformationR.net/ir/15-4/colis713.html]

Find other papers on this subject

Scholar Search   Google Search   Bing

Bookmark This Page

Appendix: e-mail survey

Dear [name],

[personal message]

I am carrying out a study of applications of webometrics research and would be very grateful if you could assist this by giving an answer to the following question. A draft of the article is available at [URL] so that you can see the context in which the information would be used and also my own answers to the question built into the text.

Question: Please could you name, give an URL for (if possible) or describe projects that you have been involved in that have (a) used webometrics AND (b) been applied in the sense of being funded either as contract research or as other contract work. This excludes research funded by research funding agencies, except for specifically commissioned studies.

Best wishes and many thanks,

Mike

Contents | Author index | Subject index | Search | Home