European Journal of Operational Research 246 (2015) 1–19

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Invited Review

A review of theory and practice in scientometrics ☆

John Mingers [a,*], Loet Leydesdorff [b]

[a] Kent Business School, University of Kent, Parkwood Road, Canterbury CT7 2PE, UK
[b] Amsterdam School of Communication Research (ASCoR), University of Amsterdam, PO Box 15793, 1001 NG Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 14 January 2015
Accepted 1 April 2015
Available online 9 April 2015

Keywords:
Altmetrics
Citations
H-index
Impact factor
Normalisation

ABSTRACT

Scientometrics is the study of the quantitative aspects of the process of science as a communication system. It is centrally, but not only, concerned with the analysis of citations in the academic literature. In recent years it has come to play a major role in the measurement and evaluation of research performance. In this review we consider: the historical development of scientometrics, sources of citation data, citation metrics and the "laws" of scientometrics, normalisation, journal impact factors and other journal metrics, visualising and mapping science, evaluation and policy, and future developments.

© 2015 Elsevier B.V. and Association of European Operational Research Societies (EURO) within the International Federation of Operational Research Societies (IFORS). All rights reserved.

1. History and development of scientometrics

Scientometrics was first defined by Nalimov (1971, p. 2) as developing "the quantitative methods of the research on the development of science as an informational process". It can be considered as the study of the quantitative aspects of science and technology seen as a process of communication. Some of the main themes include ways of measuring research quality and impact, understanding the processes of citations, mapping scientific fields and the use of indicators in research policy and management. Scientometrics focuses on communication in the sciences, the social sciences, and the humanities among several related fields:

Bibliometrics – "The application of mathematics and statistical methods to books and other media of communication" (Pritchard, 1969, p. 349). This is the original area of study covering books and publications generally. The term "bibliometrics" was first proposed by Otlet (1934); see also Rousseau (2014).

Informetrics – "The study of the application of mathematical methods to the objects of information science" (Nacke, 1979, p. 220). Perhaps the most general field covering all types of information regardless of form or origin (Bar-Ilan, 2008; Egghe & Rousseau, 1990; Egghe & Rousseau, 1988; Wilson, 1999).

Webometrics – "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 & Ingwersen, 2004, p. 1217; Thelwall & Vaughan, 2004; Thelwall, Vaughan, & Björneborn, 2005). This field mainly concerns the analysis of web pages as if they were documents.

Altmetrics – "The study and use of scholarly impact measures based on activity in online tools and environments" (Priem, 2014, p. 266). Also called Scientometrics 2.0, this field replaces journal citations with impacts in social networking tools such as views, downloads, "likes", blogs, Twitter, Mendelay, CiteULike.

In this review we concentrate on scientometrics as that is the field most directly concerned with the exploration and evaluation of scientific research. In fact, traditionally these fields have concentrated on the observable or measurable aspects of communications – external borrowings of books rather than in-library usage; citations of papers rather than their reading – but currently online access and downloads provide new modes of usage and this leads to the developments in webometrics and altmetrics that will be discussed later. In this section we describe the history and development of scientometrics (de Bellis, 2014; Leydesdorff & Milojevic, 2015) and in the next sections explore the main research areas and issues.

Whilst scientometrics can, and to some extent does, study many other aspects of the dynamics of science and technology, in practice it has developed around one core notion – that of the citation. The act of citing another person's research provides the necessary linkages between people, ideas, journals and institutions to constitute an empirical field or network that can be analysed quantitatively. Furthermore, the citation also provides a linkage in time – between the previous publications of its references and the later appearance of its citations. This in turn stems largely from the work of one person – Eugene Garfield – who identified the importance of the citation

☆ We would like to acknowledge helpful comments from Anne-Wil Harzing, David Pendlebury, Ronald Rousseau and an anonymous referee.

* Corresponding author. Tel.: +44 1227 8240080.
E-mail addresses: j.mingers@kent.ac.uk (J. Mingers), loet@leydesdorff.net (L. Leydesdorff).

http://dx.doi.org/10.1016/j.ejor.2015.04.002
0377-2217/© 2015 Elsevier B.V. and Association of European Operational Research Societies (EURO) within the International Federation of Operational Research Societies (IFORS). All rights reserved.

and then promulgated the idea of the *Science Citation Index* (SCI) in the 1950s (and the company the *Institute for Scientific Information, ISI*, to maintain it) as a database for capturing citations (Garfield, 1955; Garfield, 1979).[1] Its initial purpose was not research evaluation, but rather help for researchers to search the literature more effectively – citations could work well as index or search terms, and also enabled unfamiliar authors to be discovered. The *SCI* was soon joined by the *Social Sciences Citation Index* (SSCI, in 1973) and the *Arts & Humanities Citation Index* (A&HCI; since 1978), and eventually taken over by Thomson Corporation who converted it into the *Web of Science* as part of their *Web of Knowledge* platform.[2] In 2013, the *SCI* covered 8539 journals, the *SSCI* 3080 journals, and the *A&HCI* approximately 1700 journals.

The SCI was soon recognised as having great value for the empirical study of the practice of science. The historian, Price (1963, 1965), was one of the first to see the importance of networks of papers and authors and also began to analyse scientometric processes, leading to the idea of cumulative advantage (Price, 1976), a version of "success to the successful" (Senge, 1990) or "success breeds success (SBS)" also known as the Matthew[3] effect (Merton, 1968; Merton, 1988). Price identified some of the key problems that would be addressed by scientometricians: mapping the "invisible colleges" (Crane, 1972) informally linking highly cited researchers at the research frontiers (cf co-authorship networks and co-citation analysis (Marshakova, 1973; Small, 1973)): studying the links between productivity and quality in that the most productive are often the most highly cited (cf the h-index); and investigating citation practices in different fields (cf normalization). In 1978, Robert K. Merton, a major sociologist, was one of the editors of a volume called *Towards a Metric of Science: The Advent of Science Indicators* (Elkana, Lederberg, Merton, Thackray, & Zuckerman, 1978) which explored many of these new approaches. Scientometrics was also developing as a discipline with the advent of the journal *Scientometrics* in 1978; a research unit in the Hungarian Academy of Sciences and scientific conferences and associations.

At the same time as scientometrics research programs were beginning, the first links to research evaluation and the use of citation analysis in policy making also occurred. For example, the ISI data was included in the (US) National Science Board's *Science Indicators Reports* in 1972 and was used by the OECD. Garfield and Sher (1963) developed a measure for evaluating journals – the *impact factor (IF)* – that has been for many years a standard despite its many flaws. Journals with this specific policy focus appeared such as *Research Policy, R&D Management* and *Research Evaluation.*

During the 1990s and 2000s several developments have occurred. The availability and coverage of the citation databases has increased immensely. The WoS itself includes many more journals and also conference proceedings, although its coverage in the social sciences and humanities is still limited. It also does not yet cover books adequately although there are moves in that direction. A rival, *Scopus*, has also appeared from the publisher Elsevier. However, the most interesting challenger is *Google Scholar* which works in an entirely different way – searching the web rather than collecting data directly. While this extension of coverage is valuable, it also leads to problems of comparison with quite different results appearing depending on the databases used.

Secondly, a whole new range of metrics has appeared superseding, in some ways, the original ones such as total number of citations and citations per paper (cpp). The h-index (Costas & Bordons, 2007; Egghe, 2010; Glänzel, 2006; Hirsch, 2005; Mingers, 2008b; Mingers, Macri, & Petrovici, 2012) is one that has become particularly prominent, now available automatically in the databases. It is transparent and robust but there are many criticisms of its biases. In terms of journal evaluation, several new metrics have been developed such as SNIP (Moed, 2010b) and SCImago Journal Rank (SJR) (González-Pereira, Guerrero-Bote, & Moya-Anegón, 2010; Guerrero-Bote & Moya-Anegón, 2012) which aim to take into account the differential citation behaviours of different disciplines, e.g., some areas of science such as biomedicine cite very highly and have many authors per paper; other areas, particularly some of the social sciences, mathematics and the humanities do not cite so highly.

A third, technical, development has been in the mapping and visualization of bibliometric networks. This idea was also initiated by Garfield who developed the concept of "historiographs" (Garfield, Sher, & Thorpie, 1964), maps of connections between key papers, to reconstruct the intellectual forebears of an important discovery. This was followed by co-citation analysis which used multivariate techniques such as factor analysis, multi-dimensional scaling and cluster analysis to analyse and map the networks of highly related papers which pointed the way to identifying research domains and frontiers (Marshakova, 1973; Small, 1973). And also co-word analysis that looked at word pairs from titles, abstracts or keywords and drew on the actor network theory of Callon and Latour (Callon, Courtial, Turner, & Bauin, 1983). New algorithms and mapping techniques such as the Blondel algorithm (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008) and the Pajek mapping software have greatly enhanced the visualization of high-dimensional datasets (de Nooy, Mrvar, & Batgelj, 2011).

But perhaps the most significant change, which has taken scientometrics from relative obscurity as a statistical branch of information science to playing a major, and often much criticised, role within the social and political processes of the academic community, is the drive of governments and official bodies to monitor, record and evaluate research performance. This itself is an effect of the neo-liberal agenda of "new public management" (NPM) (Ferlie, Ashburner, Fitzgerald, & Pettigrew, 1996) and its requirements of transparency and accountability. This occurs at multiple levels – individuals, departments and research groups, institutions and, of course, journals – and has significant consequences in terms of jobs and promotion, research grants, and league tables. In the past, to the extent that this occurred it did so through a process of peer review with the obvious drawbacks of subjectivity, favouritism and conservatism (Bornmann, 2011; Irvine, Martin, Peacock, & Turner, 1985). But now, partly on cost grounds, scientometrics are being called into play and the rather ironic result is that instead of merely reflecting or mapping a pre-given reality, scientometric methods are actually shaping that reality through their performative effects on academics and researchers (Wouters, 2014).

At the same time, the discipline of science studies itself has bi- (or tri-) furcated into at least three elements – the quantitative study of science indicators and their behaviour, analysis and metrication from a positivist perspective. A more qualitative, sociology-of-science, approach that studies the social and political processes lying behind the generation and effects of citations, generally from a constructivist perspective. And a third stream of research that is interested in policy implications and draws on both the other two.

Finally, in this brief overview, we must mention the advent of the Web and social networking. This has brought in the possibility of alternatives to citations as ways of measuring impact (if not quality) such as downloads, views, "tweets", "likes", and mentions in blogs. Together, these are known as "altmetrics" (Priem, 2014), and while they are currently underdeveloped, they may well come to rival citations in the future. There are also academic social networking sites such as *ResearchGate (www.researchgate.net), CiteULike (citeulike.org), academia.edu (www.academia.edu), RePEc (repec.org)* and *Mendeley (www.mendeley.com)* which in some cases have their own research metrics. *Google Scholar* can produce profiles of researchers, including their h-index, and *Publish or Perish* (Harzing, 2007) enhances searches of *Scholar* with the Harzing

[1] It was first realised in 1964.

[2] Web of Knowledge has now reverted to Web of Science.

[3] Named after St Matthew (25:29): "For unto everyone that hath shall be given … from him that hath not shall be taken away".

Download English Version:

https://daneshyari.com/en/article/6896502

Download Persian Version:

https://daneshyari.com/article/6896502

Daneshyari.com