A Review of Theory and Practice in Scientometrics

John Mingers

Kent Business School, University of Kent, Canterbury CT7 2PE, UK

j.mingers@kent.ac.uk

01227 824008

Loet Leydesdorff

Amsterdam School of Communication Research (ASCoR), University of Amsterdam, PO Box 15793, 1001 NG Amsterdam, The Netherlands

loet@leydesdorff.net

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.

Keywords: altmetrics**,** bibliometrics, citations, h-index, impact factor, normalisation, scientometrics

1. HISTORY AND DEVELOPMENT OF SCIENTOMETRICS

Scientometrics is one of 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.

Scientometrics – "The quantitative methods of the research on the development of science as an informational process" (Nalimov & Mulcjenko, 1971, p. 2). This field concentrates specifically on science (and the social sciences and humanities).

1

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 (Egghe, L. & Rousseau, 1988).

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). 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, although we also discuss new developments in altmetrics. 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. This in turn stems largely from the work of one person – Eugene Garfield – who identified the importance of the citation and then created the *Science Citation Index* in the 1950's (and the company the *Institute for Scientific Information, ISI*, to maintain it) as a database for capturing citations (Garfield, E. , 1955; Garfield, E., 1979). 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 Science Citation Index* (SSCI) and the *Arts & Humanities Citation Index* (A&HCI; since 1980), and eventually taken over by Thompson Reuters who converted it into the *Web of Science* as part of their *Web of Knowledge* platform. In 2013, the *SCI* covered 8,539 journals, the *SSCI* 3,080 journals, and the *A&HCI* approximately 1,700 journals.

The SCI was soon recognized as having great value for the empirical study of the practice of science. The historian Derek de Solla 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 such as the idea of cumulative advantage (Price, 1976), a version of "success to the successful" (Senge, 1990) also

known as the Matthew[1] effect (Merton, 1968, 1988). Price identified some of the key problems to 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): 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 programmes 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 (1972) himself 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, Social Studies of Science* 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 conferences, 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 form the publishers Elsevier. However, the most interesting challenger is *Google Scholar* which works in an entirely different way – searching the web rather than collecting data directly. Whilst 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 have 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; Glänzel, W., 2006; Hirsch, 2005; Mingers, J., 2008b; Mingers, J., 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

---

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

3

areas of science such as biomedicine cite very highly and have many authors; other areas, particularly some of the social sciences, mathematics and the humanities do not.

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, E., 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, MDS 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) 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 (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, scientometrics methods are actually shaping that reality through their performative effects on academics and researchers (Wouters, P., 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

4

"altmetrics" (Priem, 2014), and whilst 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),* and *Mendeley (www.mendeley.com)* which in some cases have their own research metrics. *Google Scholar* automatically produces profiles of researchers, including their h-index, and *Publish or Perish* (www.harzing.com) enhances searches of *Scholar* as well as being a repository for multiple journals ranking lists in the field of business and management.

2. SOURCES OF CITATIONS

Clearly for the quantitative analysis of citations to be successful, there must be comprehensive and accurate sources of citation data. The major source of citations in the past was the Thompson Reuters ISI *Web of Science* (WoS) which is a specialised database covering all the papers in around 12,000 journals[2] . It also covers conferences[3] and is beginning to cover books[4]. Since 2004, a very similar rival database is available from Elsevier called *Scopus*[5] that covers 20,000 journals and also conferences and books. *Scopus* retrieves back until 1996, while WoS is available for all years since 1900. These two databases have been the traditional source for most major scientometrics exercises, for example by the *Centre for Science and Technology Studies* (CWTS) which has specialised access to them. More recently (2004), an alternative source has been provided by *Google Scholar* (GS). This works in an entirely different way, by searching the Web for references to papers and books rather than inputting data from journals. It is best accessed through a software program called *Publish or Perish*[6].

Many studies have shown that the coverage of WoS and Scopus differs significantly between different fields, particularly between the natural sciences, where coverage is very good, the social sciences where it is moderate and variable, and the arts and humanities where it is generally poor (HEFCE, 2008; Larivière, Archambault, Gingras, & Vignola-Gagné, 2006; Mahdi, D'Este, & Neely, 2008; Moed & Visser, 2008). In contrast, the coverage of GS is generally higher, and does not differ so much between subject areas, but the reliability and quality of its data can be poor (Amara & Landry, 2012).

Van Leeuwen (2006), in a study of Delft University between 1991 and 2001 found that in fields such as architecture and technology, policy and management the proportion of publication in WoS and the proportion of references to ISI material was under 30% while for applied science it was between 70% and 80%. Across the social sciences, the proportions varied between 20% for political science and

---

[2] http://wokinfo.com/essays/journal-selection-process/
[3] http://wokinfo.com/products_tools/multidisciplinary/webofscience/cpci/cpciessay/
[4] http://thomsonreuters.com/book-citation-index/
[5] http://www.elsevier.com/online-tools/scopus/content-overview
[6] http://www.harzing.com/pop.htm

50% for psychology. Mahdi et al. (2008) studied the results of the 2001 RAE in the UK and found that while 89% of the outputs in biomedicine were in WoS, the figures for social science and arts and humanities were 35% and 13% respectively. CWTS (Moed, Visser, & Buter, 2008) was commissioned to analyse the 2001 RAE and found that the proportions of outputs contained in WoS and Scopus respectively were: Economics (66%, 72%), Business and Management (38%, 46%), Library and Information Science (32%, 34%) and Accounting and Finance (22%, 35%).

There are several reasons for the differential coverage in these databases (Nederhof, 2006) and we should also note that the problem is not just the publications that are not included, but also that the publications that are included have lower citations recorded since many of the citing sources are not themselves included. The first reason is that in science almost all research publications appear in journal papers (which are largely included in the databases), but in the social sciences and even more so in humanities books are seen as the major form of research output. Secondly, there is a greater prevalence of the "lone scholar" as opposed to the team approach that is necessary in the experimental sciences and which results in a greater number of publications (and hence citations) overall. As an extreme example, a paper in *Physics Letters B* (Aad, et al., 2012) in 2012 announcing the discovery of the Higgs Boson has 2,932 authors and already has over 4000 citations. These outliers can distort bibliometrics analyses as we shall see (Cronin, B. , 2001). Thirdly, a significant number of social science and humanities journals have not chosen to become included in WoS, the accounting and finance field being a prime example. Finally, in social science and humanities a greater proportion of publications are directed at the general public or specialised constituencies such as practitioners and these "trade" publications or reports are not included in the databases.

There have also been many comparisons of WoS, Scopus and Google Scholar across a range of disciplines (Adriaanse & Rensleigh, 2013; Amara & Landry, 2012; Franceschet, 2010; García-Pérez, 2010; Harzing, A.-W. & van der Wal, 2008; Meho & Rogers, 2008; Meho & Yang, 2007). The general conclusions of these studies are:

- That the coverage of research outputs, including books and reports, is much higher in GS, usually around 90%, and that this is reasonably constant across the subjects. This means that GS has a comparatively greater advantage in the non-science subjects where Scopus and WoS are weak.
- Partly, but not wholly, because of the coverage, GS generates a significantly greater number of citations for any particular work. This can range from two times to five times as many. This is because the citations come from a wide range of sources, not being limited to the journals that are included in the other databases.
- However, the data quality in GS is very poor with many entries being duplicated because of small differences in spellings or dates and many of the citations coming from a variety of

non-research sources. With regard to the last point, it could be argued that the type of citation does not necessarily matter – it is still impact.

Typical of these comparisons is Mingers and Lipitakis (2010) who reviewed all the publications of three UK business schools from 1980 to 2008. Of the 4,600 publications in total, 3,023 were found in GS, but only 1,004 in WoS. None of the books, book chapters, conference papers or working papers were in WoS[7]. In terms of number of citations, the overall mean cites per paper (cpp) in GS was 14.7 but only 8.4 in WoS. It was also found that these rates varied considerably between fields in business and management, a topic to be taken up in the section on normalization. When taken down to the level of individual researchers the variation was even more noticeable both in terms of the proportion of outputs in WoS and the average number of citations. For example, the most prolific researcher had 109 publications. 92% were in GS, but only 40% were in WoS. The cpp in GS was 31.5, but in WoS it was 12.3.

With regard to data quality, Garcia-Perez (2010) studied papers of psychologists in WoS, GS, and PsycINFO. GS recorded more publications and citations than either of the other sources, but also had a large proportion of incorrect citations (16.5%) in comparison with 1% or less in the other sources. Adriaanse and Rensleigh (2013) studied environmental scientists in WoS, Scopus and GS and made a comprehensive record of the inconsistencies that occurred in all three across all bibliometric record fields. There were clear differences with GS having 14.0% inconsistencies, WoS 5.4%, and Scopus only 0.4%. Similar problems with GS were also found by Jacso (2008) and Harzing and van der Wal (2008).

To summarise this section, there is general agreement at this point in time that bibliometric data from WoS or Scopus is adequate to conduct research evaluations in the natural and formal sciences where the coverage of publications is high, but it is not adequate in the social sciences or humanities, although, of course, it can be used as an aid to peer review in these areas (Abramo & D'Angelo, 2011; Abramo, D'Angelo, & Di Costa, 2011; van Raan, A., 2005b). GS is more comprehensive across all areas but suffers from poor data, especially in terms of multiple versions of the same paper, and also has limitations on data access – no more than 1000 results per query. This particularly affects the calculation of cites per paper (because the number of papers is the divisor) but it does not affect the h-index which only includes the top h papers.

These varied sources do pose the problem that the number of papers and citations may vary significantly and one needs to be aware of this in interpreting any metrics. To illustrate this with a simple example, we have looked up data for one of the authors on WoS and GS. The results are shown in Table 1.

---

[7] Most studies do not include WoS for books, which is still developing (Leydesdorff & Felt, 2012).

	Cites from outputs in WoS using WoS	Cites from all sources using GS
	n, c, h, cpp	n, c, h, cpp
Cites to outputs in WoS	88, 1684, 21, 19.1	87, 4890, 31, 56.2
Cites to all outputs	349, 3796, 30, 10.8	316, 13,063, 48, 41.3

Table 1 Comparison of WoS and GS for one of the authors
N=no. of papers, c=no. of citations, h=h-index, cpp=cites per paper

The first thing to note is that there are two different ways of accessing citation data in WoS. a) One can do an author search and find all their papers, and then do a citation analysis of those papers. This generates the citations from WoS papers to WoS papers. b) One can do a cited reference search on an author. This generates all the citations from papers in WoS to the author's work whether the cited work is in WoS or not. This therefore generates a much larger number of *cited* publications and a larger number of citations for them. The results are shown in the first column of Table 1. Option a) finds 88 papers in WoS and 1684 citations for them from WoS papers. The corresponding h-index is 21. Option b) finds 349 (!) papers with 3796 citations and an h-index of 30. The 349 papers include many cases of illegitimate duplicates just as does GS. If we repeat the search in GS, we find a total of 316 cited items (cf 349) with 13,063 citations giving an h-index of 48. If we include only the papers that are in WoS we find 87 of the 88, but with 4890 citations and an h-index of 31. So, one could justifiably argue for an h-index ranging from 21 to 48, and a cpp from 10.8 to 56.2.

3. METRICS AND THE "LAWS" OF SCIENTOMETRICS

In this section we will consider the main areas of scientometrics analysis − indicators of productivity and indicators of citation impact.

3.1. Indicators of productivity

Some of the very early work, from the 1920s onwards, concerned productivity in terms of the number of papers produced by an author or research unit; the number of papers journals produce on a particular subject; and the number of key words that texts produce. They all point to a similar phenomenon – the Paretian one that a small proportion of producers are responsible for a high proportion of outputs. This also means that the statistical distributions associated with these phenomena are generally highly skewed.

Lotka (1926) studied the frequency distribution of numbers of publications, concluding that "the number of authors making *n* contributions is about *$1/n^2$* of those making one" from which can be derived de Solla Price's (1963) "square root law" that "half the scientific papers are contributed by the

top square root of the total number of scientific authors". Lotka's Law generates the following distribution:

$$P(X=k) = (6/\pi^2).k^{-2} \quad \text{where } k = 1, 2, \ldots$$

Glänzel and Schubert (1985) showed that a special case of the Waring distribution satisfies the square root law.

Bradford (1934) hypothesised that if one ranks journals in terms of number of articles they publish on a particular subject, then there will be a core that publish the most. If you then group the rest into zones such that each zone has about the same number of articles, then the number of journals in each zone follows this law:

$$N_n = k^n N_0$$

where k = Bradford coefficient, N_0 = number in core zone, N_n = journals in the nth zone;

Thus the number of journals needed to publish the same number of articles grows with a power law.

Zipf (1936) studied the frequency of words in a text and postulated that the rank of the frequency of a word and the actual frequency, when multiplied together, are a constant. That is, the number of occurrences is inversely related to the rank of the frequency. In a simple case, the most frequent word will occur twice as often as the second most frequent, and three times as often as the third.

$$rf(r) = C \qquad \text{r is the rank, f(r) is the frequency of that rank, C is a constant}$$

$$f(r) = C\ 1/r$$

More generally:

$$f(r) = \frac{1/r^s}{\sum_1^N \left(\frac{1}{n^s}\right)} \qquad \text{N is the number of items, s is a parameter}$$

The Zipf distribution has been found to apply in many other contexts such as the size of city by population. All three of these behaviours ultimately rest on the same cumulative advantage mechanisms mentioned above and, indeed, all three can be shown to be mathematically equivalent (Egghe, Leo, 2005).

However, empirical data on the distribution of publications by, for example, a particular author shows that the Lotka distribution by itself is too simplistic as it does not take into account productivity varying over time (including periods of inactivity) or subject. One approach is to model the process as a cumulation of distributions (Sichel, 1985). For example, we could assume that the number of papers

9

per year followed a Poisson distribution with parameter λ, but that the parameter itself varied with a particular distribution depending on age, activity, discipline. If we assume that the parameter follows a Gamma distribution, then this mixture results in a negative-binomial which has been found to have a good empirical fit (Mingers, J. & Burrell, 2006).

3.2. Indicators of Impact: Citations

We should begin by noting that the whole idea of the citation being a fundamental indicator of impact, let al.one quality, is itself the subject of considerable debate. This concerns: the reasons for citing others' work, Weinstock (1971) lists 15, or not citing it; the meaning or interpretation to be given to citations (Cozzens, 1989; Day, 2014; Leydesdorff, 1998); their place within scientific culture (Wouters, P., 2014); and the practical problems and biases of citation analysis (Chapman, 1989). This wider context will be discussed later, this section will concentrate on the technical aspects of citation metrics.

The basic unit of analysis is a collection of papers (or more generally research outputs including books reports etc. but as pointed out in Section 2 the main databases only cover journal papers) and the number of citations they have received over a certain period of time. In the case of an individual author, we are often interested in all their citations. In the case of evaluations of departments or journals, a particular window of three, five or ten years are usually considered. Usually the analysis occurs at a particular point in time but it can be done longitudinally or the dynamic behaviour of citations can be studied (Mingers, J., 2008a).

Citation patterns

If we look at the number of citations received by a paper over time it shows a typical birth-death process. Initially there are few citations; then the number increases to a maximum; finally they die away as the content becomes obsolete. There are many variants to this basic pattern, for example "shooting stars" that are highly cited but die quickly, and "sleeping beauties" that are ahead of their time (van Raan, A. J., 2004). There are also significantly different patterns of citation behaviour between disciplines that will be discussed in the normalization section. There are several statistical models of this process. Glänzel and Schoepflin (1995) use a linear birth process; Egghe (2000) assumed citations were exponential. Perhaps the most usual is to conceptualise the process as basically random from year to year but with some underlying mean (λ) and use the Poisson distribution. There can then be two extensions – the move from a single paper to a collection of papers with differing mean rates (Burrell, Q. , 2001), and the incorporation of obsolescence in the rate of citations (Burrell, Q., 2002, 2003).

If we assume a Gamma distribution for the variability of the parameter λ, then the result is a negative binomial of the form:

$$P(X_t = r) = \binom{r + \nu - 1}{\nu - 1} \left(\frac{\alpha}{\alpha + t} \right)^{\nu} \left(1 - \frac{\alpha}{\alpha + t} \right)^{r}, \quad r = 0, 1, 2, \ldots$$

With mean = $\nu t/\alpha$ variance = $\nu t(t + \alpha)/\alpha^2$ where ν and α are parameters to be determined empirically.

The negative binomial is a highly skewed distribution which, as we have seen, is generally the case with bibliometric data. Mingers and Burrell (2006) tested the fit on a sample of 600 papers published in 1990 in six MS/OR journals – *Management Science, Operations Research, Omega, EJOR, JORS* and *Decision Sciences* looking at fourteen years of citations. Histograms are shown in Figure 1 and summary statistics in Table 2. As can be seen, the distributions are highly skewed, and they also have modes (except *ManSci*) at zero, i.e, many papers have never been cited in all that time. The proportion of zero cites varies from 5% in *Management Science* to 22% in *Omega*.

	*JORS	Omega	EJOR	Dec Sci	Ops Res	Man Sci
Actual mean	7.3	7.2	11.3	11.1	14.6	38.6
Actual sd	17.9	15.5	19.0	14.0	28.6	42.4
% zero cites	18	22	14	12	10	5
Max cites	176	87	140	66	277	181

Table 2 Summary statistics for citations in six OR journals 1990-2004, from (Mingers, J. & Burrell, 2006)
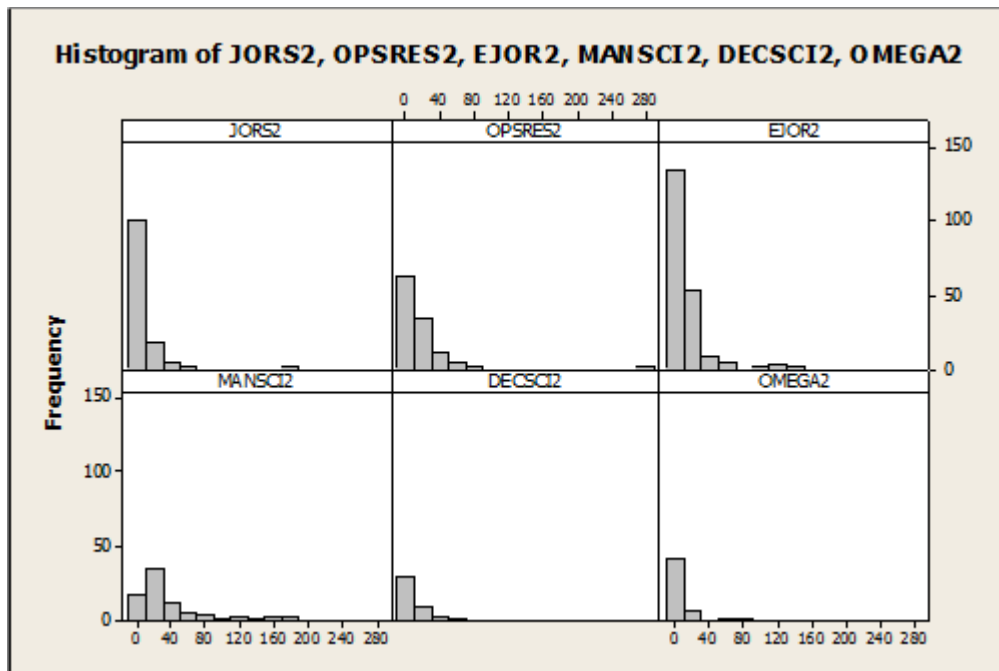
Figure 1 Histograms for papers published in 1990 in six management science journals, from (Mingers, J. & Burrell, 2006)

The issue of zero cites is of concern. On the one hand, that a paper has never been cited does not imply that it is of zero quality, especially when it has been through rigorous reviewing processes in a top journal, which is evidence that citations are not synonymous with quality. On the other, as Braun (1985). argues, a paper that has never been cited must at the least be disconnected from the field in question. The mean cites per paper (over 15 years) vary considerably between journals from 7.2 to 38.6 showing the major differences between journals (to be covered in a later section), although it is difficult to disentangle whether this is because of the intrinsically better quality of the papers or simply the reputation of the journal. Bornmann et al. (2013) found that the journal can be considered as a significant co-variate in the prediction of citation impact.

Obsolescence can be incorporated by including a time-based function in the distribution. This would generally be an S-shaped curve that would alter the value of λ over time, but there are many possibilities (Meade & Islam, 1998) and the empirical results did not identify any particular one although the gamma and the Weibull were best. It is also possible to statistically predict how many additional citations will be generated if a particular number have been received so far. The main results are that, at time *t*, the future citations are a linear function of the citations received so far, and the slope of the increment line decreases over the lifetime of the papers. These results applied to collections of papers, but do not seem to apply to the dynamics of individual papers.

In a further study of the same data set, the citation patterns of the individual were modelled (Mingers, J., 2008a). The main conclusions were twofold: i) that individual papers were highly variable and it

was almost impossible to predict the final number of citations based on the number in the early years, in fact up to about year ten. This was partly because of sleeping beauty and shooting star effects. ii) The time period for papers to mature was quite long – the maximum citations were not reached until years eight and nine, and many papers were still being strongly cited at the end of 14 years. This is very different from the natural sciences where the pace of citations is very much quicker.

If we wish to use citations as a basis for comparative evaluation, whether of researchers, journals or departments, we must consider influences on citations other than pure impact or quality. The first, and most obvious, is simply the number of papers generating a particular total of citations. A journal or department publishing 100 papers per year would expect more citations than one publishing 20. For this reason the main comparative indicator that has been used traditionally is the mean cites per paper (CPP) or raw impact per paper (RIP). This was the basis of the Leiden (CWTS) "crown indicator" measure for evaluating research units suitably normalised against other factors. (Waltman, L., van Eck, van Leeuwen, Visser, & van Raan, 2010, 2011). We should note that this is the opposite of total citations – it pays no attention at all to the number of papers, so a researcher with a CPP of 20 could have one paper, or one hundred papers each with 20 citations.

These other factors include: the general disciplinary area – natural science, social science or humanities; particular fields such as biomedicine (high) or mathematics (low); the type of paper (reviews are high); the degree of generality of the paper (i.e., of interest to a large or small audience); reputational effects such as the journal, the author, or the institution; the language; the region or country (generally the US has the highest number of researchers and therefore citations) as well as the actual content and quality of the paper.

Another interesting issue is whether all citations should be worth the same? There are three distinct factors here – the number of authors of a paper, the number of source references, and the quality of the citing journals. In terms of numbers of authors, the sciences generally have many collaborators within an experimental or laboratory setting who all get credited. Comparing this with the situation of a single author who has done all the work themselves, should not the citations coming to that paper be spread among the authors? The extreme example mentioned above concerning the single paper announcing the Higgs Boson actually had a significant effect on the position of several universities in the 2014 Times Higher World University Ranking (Holmes, 2014). The paper, with 2896 "authors" affiliated to 228 institutions, had received 1631 citations within a year. All of the institutions received full credit for this and for some, who only had a relatively small number of papers, it made a huge difference.

The number of source references is a form of normalisation (fractional counting of citations) (Leydesdorff & Bornmann, 2011a) which will be discussed below. Taking into account the quality of the citing journal gives rise to new indicators that will be discussed in the section on journals.

The h-index

We have seen that the total number of citations, as a metric, is strongly affected by the number of papers, but does not provide any information on this. At the opposite extreme, the CPP is totally insensitive to productivity. In 2005, a new metric was proposed by a physicist called Hirsch (2005), that combined in a single, easy to understand, number both impact (citations) and productivity (citations). The h-index has been hugely influential since then, generating an entire literature of its own. Currently his paper has well over 4000 citations. In this section we will only be able to summarise the main advantages and disadvantages, for more detailed reviews see (Alonso, Cabrerizo, Herrera-Viedma, & Herrera, 2009; Bornman & Daniel, 2005; Costas & Bordons, 2007; Glänzel, W., 2006) and for mathematical properties see Glänzel (2006) and Franceschini & Maisano(2010).

The h index is defined as: "a scientist has index h if h of his or her N_p papers have at least h citations each and the other $(N_p – h)$ papers have $<= h$ citations each".

So h represents the top h papers, all of which have at least h citations. This one number thus combines both number of citations and number of papers. These h papers are generally called the h-core. The h-index ignores all the other papers below h, and it also ignores the actual number of citations received above h. The advantages are:

- It combines both productivity and impact in a single measure that is easily understood and very intuitive.
- It is easily calculated just knowing the number of citations either from WoS, Scopus or Google Scholar. Indeed, all three now routinely calculate it.
- It can be applied at different levels – researcher, journal or department.
- It is objective and a good comparator within a discipline where citation rates are similar.
- It is robust to poor data since it ignores the lower down papers where the problems usually occur. This is particularly important if using GS.

However, many limitations have been identified including some that affect all citation based measures (e.g., the problem of different scientific areas, and ensuring correctness of data), and a range of modifications have been suggested (Bornmann, Mutz, & Daniel, 2008).

- The first is that the metric is insensitive to the actual numbers of citations received by the papers in the h-core. Thus two researchers (or journals) with the same h-index could have dramatically

different actual numbers of citations. Egghe (2006) has suggested the g-index as a way of compensating for this. "A set of papers has a g-index of g if g is the highest rank such that the top g papers have, together, at least g^2 citations" (p. 132). The fundamental idea is that the h-core papers must have at least h^2 citations between them although in practice they may have many more. The more they have, the larger g will become and so it will to some extent reflect the total number of citations. The disadvantage of this metric is that it is less intuitively obvious than the h-index. Another alternative is the e-index proposed by Zheng (2009). There are several other proposals that measure statistics on the papers in the h-core, for example:

- o The a-index (Jin, 2006; Rousseau, 2006) which is the mean number of citations of the papers in the h-core.
- o The m-index (Bornmann, et al., 2008) which is the median of the papers in the h-core since the data is always highly skewed. Currently *Google Scholar Metrics*[8] implements a 5-year h-index and 5-year m-index.
- o The r-index (Jin, 2007) which is the square root of the sum of the citations of the h-core papers. This is because the a-index actually penalises better researchers as the number of citations are divided by h, which will be bigger for better scientists. A further development is the ar-index (Jun, Liang, Rousseau, & Egghe, 2007) which is a variant of the r-index also taking into account the age of the papers.

- The h-index is strictly increasing and strongly related to the time the publications have existed. This biases it against young researchers. It also continues increasing even after a researcher has retired. Data on this is available from Liang(2006)who investigated the actual sequence of h values over time for the top scientists included in Hirsch's sample. A proposed way round this is to consider the h-rate (Burrell, Q., 2007), that is the h-index at time t divided by the years since the researcher's first publication. This was also proposed by Hirsch as the m parameter in his original paper. Values of 2 or 3 indicate highly productive scientists.

- The h-index does not discriminate well since it only employs integer values. Given that most researchers may well have h-indexes between 10 and 30, many will share the same value. Guns and Rousseau (2009) have investigated real and rational variants of both g and h.

- As with all citation-based indicators, they need to be normalised in some way to citation rates of the field. Iglesias and Pecharroman (2007) collected, from WoS, the mean citations per paper in each year from 1995-2005 for 21 different scientific fields The totals ranged from under 2.5 for computer science and mathematics over 24 for molecular biology. From this data they constructed a table of normalisation factors to be applied to the h-index depending on the field and also the total number of papers published by the researcher. A similar issue concerns the number of authors. The sciences tend to have more authors per paper than the social sciences and humanities

and this generates more papers and more citations. Batista et al. (2006) developed the hI-index as the h-index divided by the mean number of authors of the h-core papers. They also claim that this accounts to some extent for the citation differences between disciplines. *Publish or Perish* also corrects for authors by dividing the citations for each paper by the number of authors before calculating the hI,norm-index. This matric has been further normalised to take into account the career length of the author (Harzing, Anne-Wil, Alakangas, & Adams, 2014).

- The h-index is dependent on or limited by the total number of publications and this is a disadvantage for researchers who are highly cited but for a small number of publications (Costas & Bordons, 2007). For example, Aguillo[9] has compiled a list of the most highly cited researchers in GS according to the h-index (382 with h's of 100 or more). A notable absentee is Thomas Kuhn, one of the most influential researchers of the last 50 years with his concept of a scientific paradigm. His book (Kuhn, 1970) alone has (14/11/14) 74,000 citations which, if the table were ranked in terms of total citations would put him in the top 100. His actual total citations are around 115,000 citations putting him in the top 20. However, his h-index is only 64. This example shows how different metrics can lead to quite extreme results – on the h-index he is nowhere; on total citations, in the top 20; and on cites per paper probably first!

There have been many comparisons of the h-index with other indicators. Hirsch himself performed an empirical test of the accuracy of indicators in predicting the future success of researchers and concluded, perhaps unsurprisingly, that the h-index was most accurate (Hirsch, 2007). This was in contrast to other studies such as (Bornmann & Daniel, 2007; Lehmann, Jackson, & Lautrup, 2006; van Raan, A., 2005a). Generally, such comparisons show that the h-index is highly correlated with other bibliometric indicators, but more so with measures of productivity such as number of papers and total number of citations, rather than with citations per paper which is more a measure of pure impact (Alonso, et al., 2009; Costas & Bordons, 2007; Todeschini, 2011).

There have been several studies of the use of the h-index in business and management fields such as information systems (Oppenheim, 2007; Truex III, Cuellar, & Takeda, 2009), management science (Mingers, J., 2008b; Mingers, J., et al., 2012), consumer research (Saad, 2006), Marketing (Moussa & Touzani, 2010) and business (Harzing, A.-W. & Van der Wal, 2009).

Overall, the h-index may be somewhat crude in compressing information about a researcher into a single number, and it should always be used for evaluation purposes in combination with other measures or peer judgement but it has clearly become well-established in practice being available in all the citation databases.

Another approach is the use of percentile measures which we will cover in the next section.

---

[9] http://www.webometrics.info/en/node/58

4. Normalisation Methods

In considering the factors that affect the number of citations that papers receive, there are many to do with the individual paper – content, type of paper, quality, author, or institution (Mingers, John & Xu, 2010) – but underlying those there are clear disciplinary differences that are hugely significant. As mentioned above, Iglesias and Pecharroman (2007) found that mean citation rates in molecular biology were ten times those in computer science. The problem is not just *between* disciplines but also *within* disciplines such as business and management which encompass different types of research fields. Mingers and Leydesdorff (2014) found that management and strategy papers averaged nearly four times as many citations as public administration. This means that comparisons between researchers, journals or institutions across fields will not be meaningful without some form of normalisation. It is also important to normalise for time period because the number of citations always increases over time (Leydesdorff, Bornmann, Opthof, & Mutz, 2011; Waltman, L. & van Eck, 2013b).

4.1. Field Classification Normalisation

The most well established methodology for evaluating research centres was developed by the *Centre for Science and Technology Studies* (CWTS) at Leiden University and is known as the crown indicator or Leiden Ranking Methodology (LRM) (van Raan, A., 2005c). Essentially, this method compares the number of citations received by the publications of a research unit over a particular time period with that which would be expected, on a world-wide basis across the appropriate field and for the appropriate publication date. In this way, it normalises the citation rates for the department to rates for its whole field. Typically, top departments may have citation rates that are three or four times the field average. Leiden also produces a ranking of world universities based on bibliometric methods that will be discussed elsewhere (Waltman, Ludo, et al., 2012).

This is the traditional "crown indicator", but this approach to normalisation has been criticised (Leydesdorff & Opthof, 2011; Lundberg, 2007; Opthof & Leydesdorff, 2010) and an alternative has been used in several cases (Cambell, Archambaulte, & Cote, 2008; Rehn & Kronman, 2008; Van Veller, Gerritsma, Van der Togt, Leon, & Van Zeist, 2009). This has generated considerable debate in the literature (Bornmann, 2010; Bornmann & Mutz, 2011; Moed, 2010a; van Raan, A., van Leeuwen, Visser, van Eck, & Waltman, 2011; Waltman, L., et al., 2010, 2011). The criticism concerns the order of calculation in the indicator and the use of a mean when citation distributions are highly skewed. It is argued that, mathematically, it is wrong to sum the actual and expected numbers of citations and then divide them. Rather, the division should be performed first, for each paper and then these ratios should be averaged. It might be thought that this is purely a technical issue, but it has been argued that it can affect the results significantly. In particular, the older CWTS method tends to weight more

highly publications from fields with high citation numbers whereas the new one weights them equally. Also, the older method is not consistent in its ranking of institutions when both improve equally in terms of publications and citations. Eventually this was accepted by CWTS, and Waltman et al. (2010, 2011) (from CTWS) have produced both theoretical and empirical comparisons of the two methods and concluded that the newer one is theoretically preferably but does not make much difference in practice. The new method is called the "mean normalised citation score" (MNCS). Gingras et al. (2011) commented that the "alternative" method was not alternative but in fact the correct way to normalise, and had been in use for fifteen years.

4.2. Source Normalisation

The normalisation method just discussed normalised citations against other citations, but an alternative approach was suggested, initially by Zitt and Small (2008) in their "audience factor", which considers the *sources* of citations, that is the reference lists of citing papers, rather than citations themselves. This general approach is gaining popularity and is also known as the "citing-side approach" (Zitt, M., 2011), source normalisation (Moed, 2010c) (SNIP), fractional counting of citations (Leydesdorff & Opthof, 2010) and a priori normalisation (Glänzel, Wolfgang, Schubert, Thijs, & Debackere, 2011).

The essential difference in this approach is that the reference set of journals is not defined in advance, according to WoS or Scopus categories, but rather is defined at the time specifically for the collection of papers being evaluated (whether that is papers from a journal, department, or individual). It consists of all the papers, in the given time window, that cite papers in the target set. Each collection of papers will, therefore, have its own unique reference set and it will be the lists of references from those papers that will be used for normalisation. This approach has obvious advantages – it avoids the use of WoS categories which are ad hoc and outdated (Leydesdorff & Bornmann, 2014; Mingers, J. & Leydesdorff, 2014) and it allows for journals that are interdisciplinary and that would therefore be referenced by journals from a range of fields.

Having determined the reference set of papers, the methods then differ in how they employ the number of references in calculating a metric. The audience factor (Zitt, M., 2011; Zitt, Michel & Small, 2008) works at the level of a citing journal. It calculates a weight for citations from that journal based on the ratio between the average number of active references[10] in all journals to the average number of references in the citing journal. This ratio will be larger for journals that have few references compared to the average because they are in less dense citation fields. Citations to the

---

[10] An "active" reference is one that is to a paper included in the database (e.g., WoS) within the time window. Other references are then ignored as "non-source references".

target (cited) papers are then weighted using the calculated weights which should equalise for the citation density of the citing journals.

Fractional counting of citations (Leydesdorff & Bornmann, 2011a; Leydesdorff & Opthof, 2010; Leydesdorff, Radicchi, Bornmann, Castellano, & Nooy, 2013) begins at the level of an individual citation and the paper which produced it. Instead of counting each citation as one, it counts it as a fraction of the number of references in the citing paper. This if a citation comes from a paper with m references, the citation will have a value of $1/m$. It is then legitimate to add all these fractionated citations to give the total citation value for the cited paper. An advantage of this approach is that statistical significance tests can be performed on the results. One issue is whether all references are included (which Leydesdorff et al. do) or whether only the active references should be counted. The third method is essentially that which underlies the SNIP indicator for journals (Moed, 2010b) which will be discussed in Section 5. In contrast to fractional counting, it forms a ratio of the mean number of citations to the journal to the mean number of references in the citing journals. A later version of SNIP (Waltman, L., van Eck, van Leeuwen, & Visser, 2013) used the harmonic mean to calculate the average number of references and in this form it is essentially the same as fractional counting except for an additional factor to take account of papers with no active citations.

Some empirical reviews of these approaches have been carried out. Waltman and van Eck (2013a, 2013b) compared the three source-normalising methods with the new CWTS crown indicator (MNCS) and concluded that the source normalisation methods were preferable to the field classification approach, and that of them, the audience factor and revised SNIP were best. This was especially noticeable for interdisciplinary journals. The fractional counting method did not fully eliminate disciplinary differences (Radicchi & Castellano, 2012) and also did not account for citation age.

4.3. Percentile-Based Approaches

We have already mentioned that there is a general statistical problem with metrics that are based on the mean number of citations, which is that citations distributions are always highly skewed (Seglen, P. O., 1992) and this invalidates the mean as a measure of central tendency; the median is better. There is also the problem of ratios of means discussed above. A non-parametric alternative based on percentiles (an extension of the median) has been suggested for research groups (Bornmann & Mutz, 2011), individual scientists (Leydesdorff, Bornmann, Mutz, & Opthof, 2011) and journals (Leydesdorff & Bornmann, 2011b). This is also used by the US National Science board in their *Science and Engineering Indicators*.

The method works as follows:

19

1. For each paper to be evaluated, a reference set of papers published in the same year, of the same type and belonging to the same WoS category is determined.

2. These are rank ordered and split into percentile rank (PR) classes, for example the top 1% (99th percentile), 5%, 10%, 25%, 50% and below 50%. For each PR, the minimum number of citations necessary to get into the class is noted[11]. Based on its citations, the paper is then assigned to one of the classes. This particular classification is known as 6PR.

3. The procedure is repeated for all the target papers and the results are then summated, giving the overall percentage of papers in each of the PR classes. The resulting distributions can be statistically tested against both the field reference values and against other competitor journals or departments[12].

The particular categories used above are only one possible set (Bornmann, Lutz, Leydesdorff, Loet, & Mutz, Rüdiger, 2013) – others in use are [10%, 90%] and [0.01%, 0.1%, 1%, 10%, 20%, 50%] (used in ISI Essential Science indicators) and the full 100 percentiles (100PR) (Bornmann, Lutz, Leydesdorff, Loet, & Wang, Jian, 2013; Leydesdorff, Bornmann, Mutz, et al., 2011). This approach provides a lot of information about the proportions of papers at different levels, but it is still useful to be able to summarise performance in a single value. The suggested method is to calculate a mean of the ranks weighted by the proportion of papers in each. The minimum is 1, if all papers are in the lowest rank; the maximum is 6 if they are all in the top percentile. The field average will be 1.91 - (.01, .04, 05, .15, .25, .50) x (6,5,4,3,2,1) - so a value above that is better than the field average. A variation of this metric has been developed as an alternative to the journal impact factor (JIF) called *I3* (Leydesdorff, 2012; Leydesdorff & Bornmann, 2011b). Instead of multiplying the percentile ranks by the proportion of papers in each class, they are multiplied by the actual numbers of papers in each class thus giving a measure that combines productivity with citation impact. In the original, the 100PR classification was used but other ones are equally possible.

The main drawback of this method is that it relies on the field definitions in WoS or another database which are unreliable, especially for interdisciplinary journals. It might be possible to combine it with some form of source normalisation (Colliander, 2014).

5. Indicators of Journal Quality: The Impact Factor and Other Metrics

So far, we have considered the impact of individual papers or researchers, but of equal importance is the impact of journals in terms of library's decisions about which journals to take (less important in the age of e-journals), authors' decisions about where to submit their papers, and in subsequent

---

[11] There are several technical problems to be dealt with in operationalising these classes (Bornmann, Leydesdorff, & Mutz, 2013; Bornmann, Lutz, Leydesdorff, Loet, & Wang, Jian, 2013).

[12] Using Dunn's test or the Mann-Whitney U test (Leydesdorff, Bornmann, Mutz, et al., 2011).

20

judgements of the quality of the paper. Indeed journal ranking lists such as the UK *Association of Business Schools'* (ABS) has a huge effect on research behaviour (Mingers, J. & Willmott, 2013). Until recently, the journal impact factor (JIF) has been the pre-eminent measure. This was originally created by Garfield and Sher (1963) as a simple way of choosing journals for their SCI but, once it was routinely produced in WoS (who have copyright to producing it), it became a standard. Garfield recognised its limitations and also recommended a metric called the "cited half-life" which is a measure of how long citations last for. Specifically, it is the median age of papers cited in a particular year, so a journal that has a cited half-life of five years means that 50% of the citations are to papers published in the last five years.

JIF is simply the mean citations per paper for a journal over a two year period. For example, the 2014 JIF is the number of citations in 2014 to papers published in a journal in 2012 and 2013, divided by the number of such papers. WoS also published a 5-year JIF because in many disciplines two years is too short a time period. It is generally agreed that the JIF has few benefits for evaluating research, but many deficiencies (Brumback, 2008; Cameron, 2005; Seglen, P., 1997). Even Garfield (1998) has warned about its over-use.

- JIF depends heavily on the research field. As we have already seen, there are large differences in the publishing and citing habits of different disciplines and this is reflected in huge differences in JIF values. Looking at the WoS journal citation reports 2013, in the area of cell biology the top journal has a JIF of 36.5 and the 20th one of 9.8. *Nature* has a JIF of 42.4. In contrast, in the field of management, the top journal (*Academy of Management Review*) is 7.8 and the 20th is only 2.9. Many journals have JIFs of less than 1. Thus, it is not appropriate to compare JIFs across fields (even within business and management) without some form of normalisation

- The two-year window. This is a very short time period for many disciplines, especially given the lead time between submitting a paper and having it published which may itself be two years. In management, many journals have a cited half-life of over 10 years while in cell biology it is typically less than 6. The 5-year JIF is better in this respect (Campanario, 2011).

- There is a lack of transparency in the way the JIF is calculated and this casts doubt on the results. Brumback (2008) studied medical journals and could not reproduce the appropriate figures. It is highly dependent on which types of papers are included in the denominator. In 2007, the editors of three prestigious medical journals published a paper questioning the data (Rossner, Van Epps, & Hill, 2007). Pislyakov (2009) has also found differences between JIFs calculated in WoS and Scopus for economics resulting from different journal coverage.

- It is possible for journals to deliberately distort the results by, for example, publishing many review articles which are more highly cited; publishing short reports or book reviews that get

cited but are not included in the count of papers; or pressuring authors to gratuitously reference excessive papers from the journal (Wilhite & Fong, 2012). The *Journal of the American College of Cardiology*, for example, publishes each year an overview of highlights in its previous year so that the IF of this journal is boosted (DeMaria, et al., 2008).

- If used for assessing individual researchers or papers the JIF is unrepresentative (Oswald, 2007). As Figure 1 shows, the distribution of citations within a journal is highly skewed and so the mean value will be distorted by a few highly cited papers, and not represent the significant number that may never be cited at all.

In response to criticisms of the JIF, several more sophisticated metrics have been developed, although the price for sophistication is complexity of calculation and a lack of intuitiveness in what it means .

The first metrics we will consider take into account not just the quantity of citations but also their quality in terms of the prestige of the citing journal. They are based on iterative algorithms over a network, like Googles's PageRank, that initially assign all journals an equal amount of prestige and then iterate the solution based on the number of citations (the links) between the journals (nodes) until a steady state is reached. The first such was developed by Pinsky and Narin (1976) but that had calculation problems. Since then, Page et al. (Ma, Guan, & Zhao, 2008; Page, Brin, Motwani, & Winograd, 1999) have an algorithm based directly on PageRank but adapted to citations, Bergstrom (Bergstrom, 2007) has developed the Eigenfactor which is implemented in WoS and Gonzalez-Pereira et al. (2010) have developed SCImago Journal Rank (SJR) which is implemented in Scopus.

The Eigenfactor is based on the notion of a researcher taking a random walk following citations from one paper to the next, measuring the relative frequency of occurrence of each journal as a measure of prestige. It explicitly excludes journal self-citations unlike most other metrics. Its values tend to be very small, for example the largest in the management field is Management Science with a value of 0.03 while the 20th is 0.008 which is not very meaningful. The Eigenfactor measures the total number of citations and so is affected by the total number of papers published by a journal. A related metric is the article influence score (AI) which is the Eigenfactor divided by the proportion of papers in the database belonging to a particular journal over five years. It can therefore be equated to a 5-yr JIF. A value of 1.0 shows that the journal has average influence; values greater than 1.0 show greater influence. We can see that in cell biology the largest AI is 22.1 compared with 6.6 in management. Fersht (2009) and Davies (2008) argue empirically, that the Eigenfactor gives essentially the same information as total citations as it is size-dependent, but West et al. (2010) dispute this. It is certainly the case that the rankings of journals with the Eigenfactor are significantly different to those based on total citations, the JIF or AI, which are all quite similar (Leydesdorff, 2009).

Metric	Description	Advantages	Disadvantages	Maximum values for: a) cell biology b) management	No of papers	field	prestige
Impact factor (JIF) and cited half-life (WoS)	Mean citations per paper over a 2 or 5 year window. Normalised to number of papers. Counts citations equally	Well-known, easy to calculate and understand.	Not normalised to discipline; short time span; concerns about data and manipulation	From WoS a) 36.5 b) 7.8	Y	N	N
Eigenfactor and article influence score (AI) (WoS)	Based on PageRank, measures citations in terms of the prestige of citing journal. Not normalised to discipline or number of papers. Correlated with total citations. Ignores self-citations. AI is normalised to number of papers, so is like a JIF 5-yr window	Very small values, difficult to interpret, not normalised	The AI is normalised to number of papers. A value of 1.0 shows average influence across all journals	From WoS Eigenfactor: a)0.599 b)0.03 AI: a) 22.2 b) 6.56	N Y	N N	Y Y
SJR and SJR2 (Scopus)	Based on citation prestige but also includes a size normalisation factor. SJR2 also allows for the closeness of the citing journal. 3-year window	Complex calculations and not easy to interpret. Not field normalised	Normalised number of papers but not to field so comparable to JIF. Most sophisticated indicator		Y	N	Y
h-index (Scimago website and Google Metrics)	The h papers of a journal that have at least h citations. Can have any window – Google metrics uses 5-year	Easy to calculate and understand. Robust to poor data	Not normalised to number of papers or field. Not pure impact but includes volums	From Google Metrics h5: a) 223 b) 72 h5 median: a) 343 b)122	N	N	N
SNIP Revised SNIP (Scopus)	Citations per paper normalised to the relative database citation potential, that is the mean number of references in the papers that cite the journal	Normalises both to number of papers and field.	Does not consider citation prestige. Complex and difficult to check. Revised version is sensitive to variability of number of references		Y	Y	N

| I3 | Combines the distribution of citation percentiles with respect to a reference set with the number of papers in each percentile class | Normalises across fields. Does not use the mean but is based on percentiles which is better for skewed data | Needs reference sets based on pre-defined categories such as WoS | Not known | N | Y | N |

Table 2 Characteristics of Metrics for Measuring Journal Impact

The SJR works in a similar way to the Eigenfactor but includes within it a size normalisation factor and so is more akin to the article influence score. Each journal is a node and each directed connection is a normalised value of the number of citations from one journal to another over a three year window. It is normalised by the total number of citations in the citing journal for the year in question. It works in two phases:

1. An un-normalised value of journal prestige is calculated iteratively until a steady state is reached. The value of prestige actually includes three components: A fixed amount for being included in the database (Scopus); an amount dependent on the number of papers the journal produces; a citation amount dependent on the number of citations received, and the prestige of the sources. However, there are a number of arbitrary weighting constants in the calculation.
2. The value from 1., which is size-dependent, is then normalised by the number of published articles and adjusted to give an "easy-to-use" value.

Gonzales-Pereira et al. (2010) carried out extensive empirical comparisons with a 3-yr JIF (on Scopus data). The main conclusions were that the two were highly correlated, but the SJR showed that some journals with high JIFs and lower SJRs were gaining citations from less prestigious sources. This was seen most clearly in the computer science field where the top ten journals, based on the two metrics, were entirely different except for the number one, which was clearly a massive outlier (*Briefings in Bioinformatics*). Values for the JIF are significantly higher than for SJR. Falagas et al. (Falagas, Kouranos, Arencibia-Jorge, & Karageorgopoulos, 2008) also compared the SJR favourably with the JIF.

There are several limitations of these 2nd generation measures: the values for "prestige" are difficult to interpret as they are not a mean citation value but only make sense in comparison with others; they are still not normalised for subject areas (Lancho-Barrantes, Guerrero-Bote, & Moya-Anegón, 2010); and the subject areas themselves are open to disagreement (Mingers, J. & Leydesdorff, 2014).

A further development of the SJR indicator has been produced (Guerrero-Bote & Moya-Anegón, 2012) with the refinement that, in weighting the citations according to the prestige of the citing

journal, the relatedness of the two journals is also taken into account. An extra term is added based on the cosine of the angle between the co-citation vectors of the journals so that the citations from a journal in a highly related area count for more. It is claimed that this also goes some way towards reducing the disparity of scores between subjects. However, it also makes the indicator even more complex, hard to compute, and less understandable.

The h-index can also be used to measure the impact of journals as it can be applied to any collection of cited papers (Braun, Glänzel, & Schubert, 2006; Schubert & Glänzel, 2007; Xu, Liu, & Mingers, 2015).Studies have been carried out in several disciplines: marketing (Moussa & Touzani, 2010), economics (Harzing, A.-W. & Van der Wal, 2009), information systems (Cuellar, Takeda, & Truex, 2008) and business (Mingers, J., et al., 2012). The advantages and disadvantages of the h-index for journals are the same as the h-index generally, but it is particularly the case that it is not normalised for different disciplines, and it is also strongly affected by the number of papers published. So a journal that publishes a small number of highly cited papers will be disadvantaged in comparison with one publishing many papers, even if not so highly cited. Google Metrics (part of Google Scholar) uses a 5-year h-index and also shows the median number of citations for those papers in the h core to allow for differences between journals with the same h-index. It has been critiqued by Delgado-López-Cózar and Cabezas-Clavijo (2012).

Another recently developed metric that is implemented in Scopus but not WoS is SNIP – source normalised impact per paper (Moed, 2010b). This normalises for different fields based on the citing-side form of normalisation discussed above, that is, rather than normalising with respected to the *citations* that a journal receives, it normalises with respect to the number of *references* in the citing journals. The method proceeds in three stages:

1. First the raw impact per paper (RIP) is calculated for the journal. This is essentially a three year JIF – the total citations from year *n* to papers in the preceding three years is divided by the number of citable papers.
2. Then the database citation potential for the journal (DCP) is calculated. This is done by finding all the papers in year *n* that cite papers the journal over the preceding ten years, and then calculating the arithmetic mean of the number of references (to papers in the database – Scopus) in these papers.
3. The DCP is then relativized (RDCP). The DCP is calculated for *all* journals in the database and the median value is found. Then $RDCP_j = DCP_j/Median\ DCP$. Thus a field that has many references will have an RDCP above 1.
4. Finally, $SNIP_j = RIP_j / RDCP_j$

The result is that journals in fields that have a high citation potential will have their RIP reduced, and vice versa for fields with low citation potential. This is an innovative measure both because it normalises for both number of publications and field, and because the set of reference journals are specific to each journal rather than being defined beforehand somewhat arbitrarily. Moed presents empirical evidence from the sciences that the subject normalisation does work even at the level of pairs of journals in the same field. Also, because it only uses references to papers within the database, it corrects for coverage differences – a journal with low database coverage will have a lower DCP and thus a higher value of SNIP.

A modified version of SNIP has recently been introduced (Waltman, L., et al., 2013) to overcome certain technical limitations, and also in response to criticism from Leydesdorff and Opthof (2010; Moed, 2010c) who favour a fractional citation approach. The modified version involves two main changes: i) the mean number of references (DCP), but not the RIP, is now calculated using the harmonic mean rather than the arithmetic mean. ii) The relativisation of the DCP to the overall median DCP is now omitted entirely, now SNIP = RIP/DCP.

Mingers (2014) has pointed out two problems with the revised SNIP. First, because the value is no longer relativized it does not bear any particular relation to either the RIP for a journal, or the average number of citations/references in the database which makes it harder to interpret. Second, the harmonic mean, unlike the arithmetic, is sensitive to the variability of values. The less even the numbers of references, the lower will be the harmonic mean and this can make a significant difference to the value of SNIP which seems inappropriate. There is also a more general problem with these sophisticated metrics that work across a whole database, and that is that the results cannot be easily replicated as most researchers do not have sufficient access to the databases (Leydesdorff, 2013).

Two other alternatives to the JIF have been suggested (Leydesdorff, 2012) – fractional counting of citations, which is similar in principle to SNIP, and the use of non-parametric statistics such as percentiles which avoids using means which are inappropriate with highly skewed data. A specific metric, based on percentiles, called I3 has been proposed by Leydesdorff (2011b) which combines relative citation impact with productivity in terms of the numbers of papers but is normalised through the use of percentiles (see Section 4.3 for more explanation).

6. Visualizing and mapping science

In addition to its use as an instrument for the evaluation of impact, citations can also be considered as an operationalization of a core process in scholarly communication, namely, referencing. Citations refer to texts other than the one that contains the cited references, and thus induce a dynamic vision of the sciences developing as networks of relations (Price, 1965). The development of co-citation

analysis (Marshakova, 1973; Small, 1973) and co-word analysis (Callon, et al., 1983) were achievements of the 1970s and 1980s. Aggregated journal-journal citations are available on a yearly basis in the Journal Citation Reports of the Science Citation Index since 1975. During the mid-80s several research teams began to use this data for visualization purposes using multidimensional scaling and other such techniques (Doreian & Fararo, 1985; Leydesdorff, 1986; Tijssen, de Leeuw, & van Raan, 1987). The advent of graphical user-interfaces in Windows during the second half of the 1990s stimulated the further development of network analysis and visualization programs such as Pajek (de Nooy, et al., 2011) that enable users to visualize large networks. Using large computer facilities, Boyack *et al..* (2005) first mapped "the backbone" of all the sciences (De Moya-Anegón, et al., 2007).

Bollen *et al..* (2009) developed maps based on clickstream data; Rosvall & Bergstrom (2010) proposed to use alluvial maps for showing the dynamics of science. Rafols *et al..* (2010) first proposed to use these "global" maps as backgrounds for overlays that inform the user about the position of specific sets of documents, analogously to overlaying institutional address information on geographical maps like Google Maps. More recently, these techniques have further be refined, using both journal (Leydesdorff, Rafols, & Chen, 2013) and patent data (Kay, Newman, Youtie, Porter, & Rafols, 2014)).

Nowadays, scientometric tools for the visualization are increasingly available on the internet. Some of them enable the user directly to map input downloaded from Web of Science or Scopus. VOSviewer[13] (Van Eck & Waltman, 2010) can generate, for example, co-word and co-citation maps from this data.

6.1. *Visualisation techniques*

The systems view of multidimensional scaling (MDS) is deterministic, whereas the graph-analytic approach can also begin with a random or arbitrary choice of a starting point. Using MDS, the network is first conceptualized as a multi-dimensional space that is then reduced stepwise to lower dimensionality. At each step, the stress increases. Kruskall's stress function is formulated as follows:

$$S = \sqrt{\frac{\sum_{i \neq j} (\|x_i - x_j\| - d_{ij})^2}{\sum_{i \neq j} d_{ij}^2}} \qquad (1)$$

In this formula $\|x_i - x_j\|$ is equal to the distance on the map, while the distance measure d_{ij} can be, for example, the Euclidean distance in the data under study. One can use MDS to illustrate factor-analytic results in tables, but in this case the Pearson correlation is used as the similarity criterion.

---

[13] http://www.vosviewer.com

Spring-embedded or force-based algorithms can be considered as a generalization of MDS, but were inspired by developments in graph theory during the 1980s. Kamada and Kawai (1989) were the first to reformulate the problem of achieving target distances in a network in terms of energy optimization. They formulated the ensuing stress in the graphical representation as follows:

$$S = \sum_{i \neq j} s_{ij} \quad \text{with} \quad s_{ij} = \frac{1}{d_{ij}^{2}} (\|x_i - x_j\| - d_{ij})^2 \tag{2}$$

Equation 2 differs from Equation 1 by taking the square root in Equation 1, and because of the weighing of *each* term with $1/d_{ij}^2$ in the numerator of Equation 2. This weight is crucial for the quality of the layout, but defies normalization with $\sum d_{ij}^2$ in the denominator of Equation 1; hence the difference between the two stress values.

The ensuing difference at the conceptual level is that spring-embedding is a graph-theoretical concept developed for the topology of a network. The weighting is achieved for each individual link. MDS operates on the multivariate space as a system, and hence refers to a different topology. In the multivariate space, two points can be close to each other without entertaining a relationship. For example, they can be close or distanced in terms of the correlation between their *patterns* of relationships.

In the network topology, Euclidean distances and geodesics (shortest distances) are conceptually more meaningful than correlation-based measures. In the vector space, correlation analysis (factor analysis, etc.) is appropriate for analysing the main dimensions of a system. The cosines of the angles among the vectors, for example, build on the notion of a multi-dimensional space. In bibliometrics, Ahlgren et al. (2003) have argued convincingly in favour of the cosine as a non-parametric similarity measure because of the skewedness of the citation distributions and the abundant zeros in citation matrices. Technically, one can also input a cosine-normalized matrix into a spring-embedded algorithm. The value of (1 – *cosine*) can then be considered as a distance in the vector space (Leydesdorff & Rafols, 2011).

Newman & Girvan (2004) developed an algorithm in graph theory that searches for (latent) community structures in networks of observable relations. An objective function for the decomposition is recursively minimized and thus a relative "modularity" Q can be measured (and normalized between zero and one). Blondel et al. (2008) improved community-finding for large networks; this routine is implemented in Pajek and Gephi, whereas Newman & Girvan's original routine can be found in the Sci2 Toolset for "the science of science".[14] VOSviewer provides its own algorithms for the mapping and the decomposition .

---

[14] https://sci2.cns.iu.edu/user/index.php

6.2. Local and global maps

To illustrate some of these possibilities, we analysed the 505 documents published in the *European Journal of Operational Research* in 2013[15]. Among the 1,555 non-trivial words in the titles of these documents, 58 words occur more than ten times and form a large component. A semantic map of these terms is shown in Figure 2.
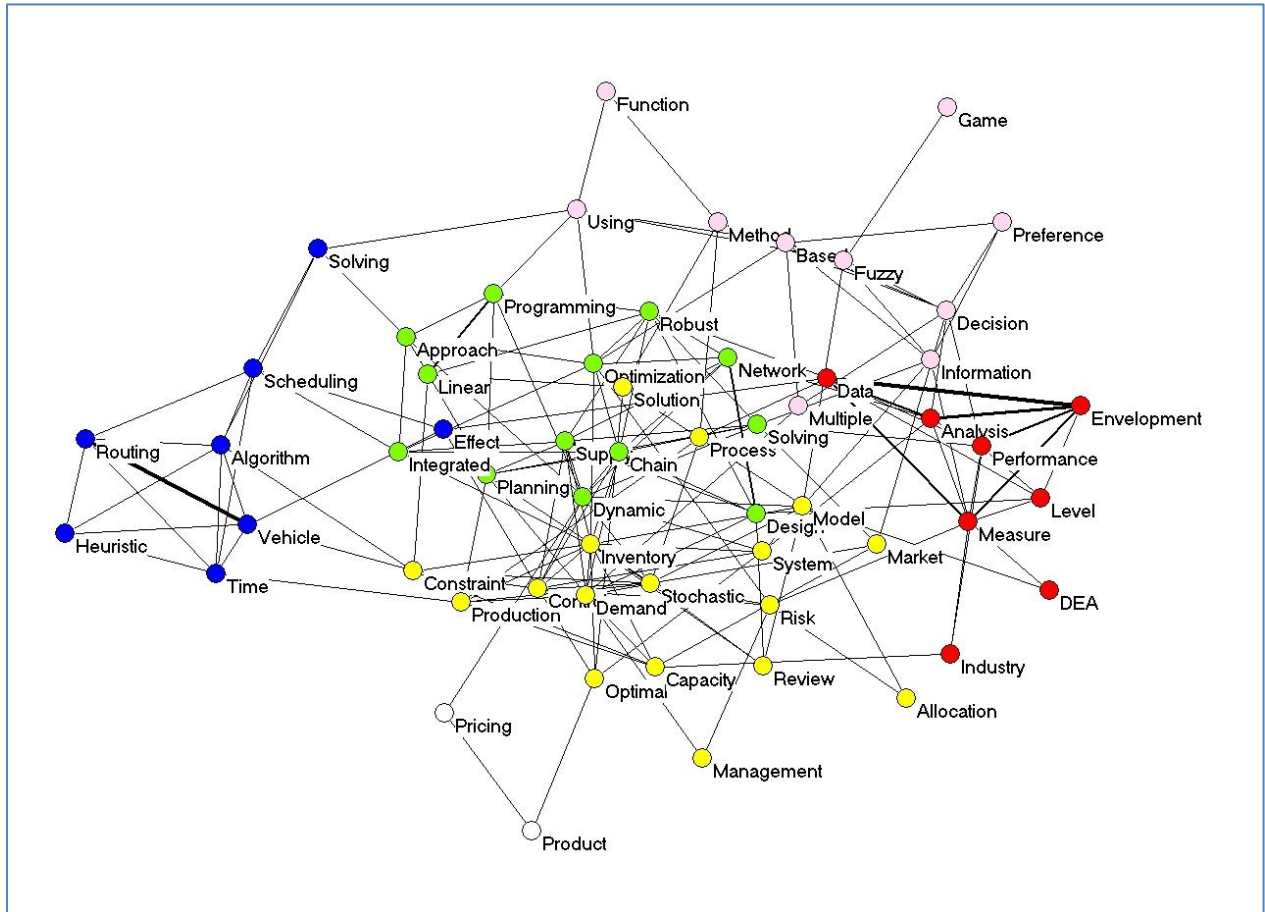

Figure 2: Cosine-normalized map of the 58 title words which occur ten or more times in the 505 documents published in EJOR during 2013. (cosine > 0.1; modularity Q = 0.548 using Blondel et al.., (2008); Kamada & Kawai (1989) used for the layout; see http://www.leydesdorff.net/software/ti.)

Figure 3 shows the 613 journals that are most highly cited in the same 505 *EJOR* papers (12,172 citations between them) overlaid on to a global map of science (Leydesdorff, L., et al., 2013). The cited references can, for example, be considered as an operationalization of the knowledge base on which these articles draw. It can be seen that, apart from the main area around OR and management, there is significant citation to the environmental sciences, chemical engineering, and biomedicine. Rao-Stirling diversity — a measure for the interdisciplinarity of this knowledge base (Rao, 1982) — however, is low (0.1187). In other words, citation within the specialty prevails.

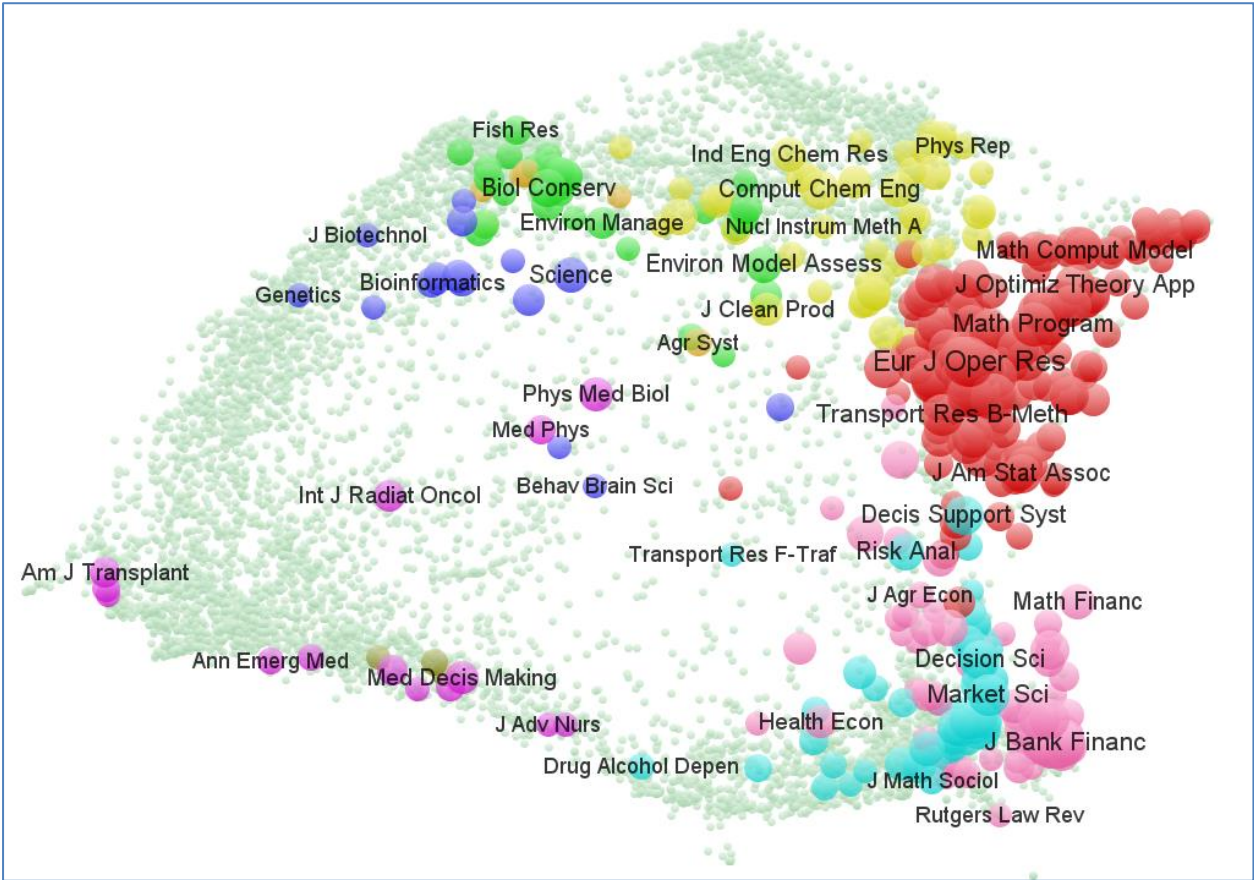[15] Using a routine available at http://www.leydesdorff.net/software/ti

Figure 3: 613 journals cited in 505 documents published in EJOR during 2013, overlaid on the global map of science in terms of journal-journal citation relations. (Rao-Stirling diversity is 0.1187; Leydesdorff et al.. (in press); see at http://www.leydesdorff.net/journals12).

Figure 4 shows a map of the field of OR based on the 29 journals most highly cited in papers published in *Operations Research* in 2013. In this map three groupings have emerged – the central area of OR, including transportation; the lower left of particularly mathematical journals; and the upper region of economics and finance journals which includes *Management Science*.
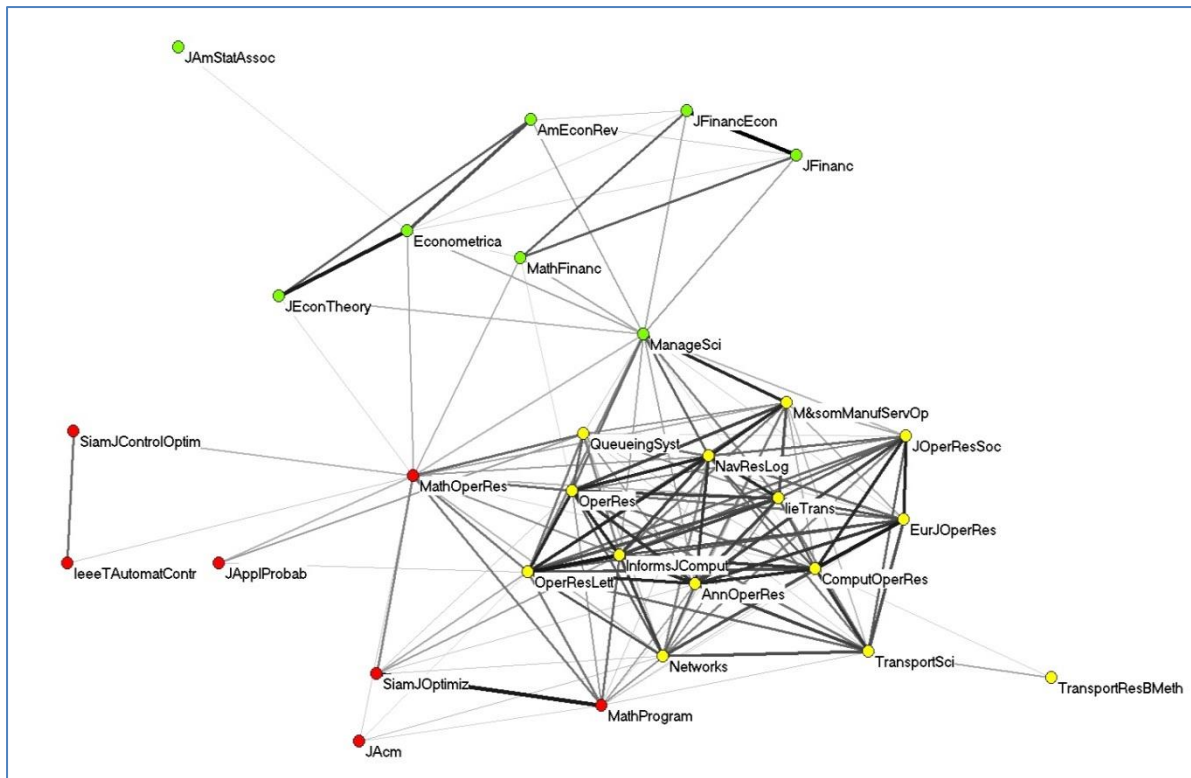
Figure 4: local map of the 29 journals cited in articles of Operations Research in 2013 (1% level; cosine > 0.2; Kamada & Kawai, 1989; Blondel et al.., 2008; Q = 0.213).

7. Evaluation and Policy

As we said in the introduction, scientometrics has come to prominence because of its use in the evaluation and management of research performance, whether at the level of the researcher, research group, institution or journal (Bornmann & Leydesdorff, 2014). Many countries, especially the UK, Australia, New Zealand and Italy, carry out regular reviews of university performance affecting both the distribution of research funds and the generation of league tables. On a macro scale, world university league tables have proliferated (e.g., QS[16], Times Higher[17] and Shanghai[18]) including one from Leiden[19] based purely on bibliometrics (Waltman, Ludo, et al., 2012) while on a micro scale personal employment and promotion is shaped by journal and citation data. Much of this is based on the citation metrics that we have discussed above.

The traditional method of research evaluation was peer review (Abramo, et al., 2011; Moed, 2007). However, this has many drawbacks – it is very time consuming and costly (Abramo & D'Angelo, 2011), subject to many biases and distortions (Horrobin, 1990; Moxham & Anderson, 1992),

---

generally quite opaque (panel members in the 2008 UK RAE were ordered to destroy all notes for fear of litigation) (Reale, Barbara, & Costantini, 2007) and limited in the extent to which it actually provides wide-ranging and detailed information (Butler & McAllister, 2009). The UK did investigate using bibliometrics in 2008 (HEFCE, 2009), used them to a limited extent in 2014, and are expected to employ them more fully in 2020. In contrast, bibliometrics has the potential to provide a cheaper, more objective and more informative mode of analysis, although there is general agreement that bibliometrics should only be used in combination with some form of transparent peer review Moed (2007; van Raan, A., 2005b). Abramo and D'Angelo (2011) compared informed peer review (including the UK RAE) with bibliometrics on the natural and formal sciences in Italy and concluded that bibliometrics were clearly superior across a range of criteria – accuracy, robustness, validity, functionality, time and cost. They recognized that there were problems in the social sciences and humanities where citation data is often not available.

The effective use of bibliometrics has a number of requirements, not all of which are currently in place.

First, we need robust and comprehensive data. As we have already seen, the main databases are reliable but their coverage is limited especially in the humanities and social sciences and they need to enlarge their scope to cover all forms of research outputs (Leydesdorff, 2008). Google Scholar is more comprehensive, but unreliable and non-transparent. At this time, full bibliometric evaluation is feasible in science and some areas of social science, but not in the humanities or some areas of technology (Archambault, Vignola-Gagné, Côté, Larivière, & Gingras, 2006; Nederhof, 2006; van Leeuwen, 2006). Abramo and D'Angelo (2011)suggest that nations should routinely collect data on all the publications published within its institutions so that it is scrutinised and available on demand rather than having to be collected anew each time a research evaluation occurs.

Second, we need suitable metrics that measure what is important in as unbiased way as possible. These should not be crude ones such as simple counts of citations or papers, the h-index (although this has its advantages) or journal impact factors but more sophisticated ones that take into account the differences in citation practices across different disciplines as has been discussed in Section 4. This is currently an area of much debate with a range of possibilities (Gingras, Y., 2014). The traditional crown indicator (now MNCS) is subject to criticisms concerning the use of the mean on highly cited data and also on the use of WoS field categories (Ruiz-Castillo & Waltman, 2014). There are source normalised alternatives such as SNIP or fractional counting (Aksnes, Schneider, & Gunnarsson, 2012) and metrics that include the prestige of the citing journals such as SJR. There are also moves towards non-parametric methods based on percentiles. One dilemma is that the more sophisticated the metrics become, the less transparent and harder to replicate they are.

A third area for consideration is inter- or trans- disciplinary work, and work that is more practical and practitioner oriented. How would this be affected by a move towards bibliometrics? There is currently little research in this area (Larivière & Gingras, 2010) and we are not aware of any empirical studies to test whether interdisciplinary work gains more or less citations. Indeed, bibliometrics is still at the stage of establishing reliable and feasible methods for defining and measuring interdisciplinarity (Wagner, et al., 2011). Huutoniemi et al. (2010) developed a typology and indicators to be applied to research proposals, and potentially research papers as well; Leydesdorff and Rafols (2011) have developed citation-based metrics to measure the interdisciplinarity of journals; and Silva et al. (Silva, Rodrigues, Oliveira Jr, & da F. Costa, 2013) evaluated the relative interdisciplinarity of science fields using entropy measures.

Fourth, we must recognise, and try to minimise, the fact that the act of measuring inevitably changes the behaviour of the people being measured. So, citation-based metrics will lead to practices, legitimate and illegitimate, to increase citations; an emphasis on 4* journals leads to a lack of innovation and a reinforcement of the status quo. For example, Moed (2008) detected significant patterns of response to UK research assessment metrics, with an increase in total publications after 1992 when numbers of papers were required; a shift to journals with higher citations after 1996 when quality was emphasised; and then in increase in the apparent number of research active staff through greater collaboration during 1997-2000. Michels and Schmoch (2014) found that German researchers changed their behaviour to aim for more US-based high impact journals in order to increase their citations.

Fifth, we must be aware that often problems are caused not by the data or metrics themselves, but by their inappropriate use either by academics or by administrators (Bornmann & Leydesdorff, 2014; van Raan, A., 2005b). There is often a desire for "quick and dirty" results and so simple measures such as the h-index or the JIF are used indiscriminately without due attention being paid to their limitations and biases. This also reminds us that there are ethical issues in the use of bibliometrics for research evaluation and Furner (2014) has developed a framework for evaluation that includes ethical dimensions.

8. Future Developments

8.1. Alternative metrics

Although citations still form the core of scientometrics, the dramatic rise of social media has opened up many more channels for recording the impact of academic research (Bornmann, 2014; Konkiel & Scherer, 2013; Priem, 2014; Roemer & Borchardt, 2012). These go under the name of "altmetrics"

both as a field, and as particular alternative metrics[20]. One of the interesting characteristics of altmetrics is that it throws light on the impacts of scholarly work on the general public rather than just the academic community. The Public Library of Science (PLoS) (Lin & Fenner, 2013) has produced a classification of types of impacts which goes from least significant to most significant:

- **Viewed**: institutional repositories, publishers, PLoS, Academia.com, ResearchGate. Perneger (2004) found a weak correlation with citations.
- **Downloaded**/**Saved**: as viewed plus CiteUlike, Mendelay.
- **Discussed**: Wikipedia, Facebook, Twitter, Natureblogs[21], ScienceSeeker[22], general research blogs. Eysenbach (2011) suggested a "twimpact factor" based on the number of tweets
- **Recommended**: F1000Prime[23]
- **Cited**: Wikipedia CrossRef, WoS, Scopus, Google Scholar,

Altmetrics is still in its infancy and the majority of papers would have little social networking presence. There are also a number of problems: i) altmetrics can be gamed by "buying" likes or tweets; ii) there is little by way of theory about how and why altmetrics are generated (this is also true of traditional citations); iii) a high score may not mean that the paper is especially good, just on a controversial or fashionable topic; and iv) because social media is relatively new it will under-represent older papers.

8.2. The shape of the discipline

Citations refer to texts other than the one that contains the cited references, and thus induce a dynamic vision of the sciences developing as networks of relations (Price, 1965). In the scientometric literature, this has led to the call for "a theory of citation" (Cozzens, 1989; Cronin, B., 1998; Garfield, E., 1979; Leydesdorff, 1998; Nicolaisen, 2007). Wouters (1998) noted that in science and technology studies (STS), citations are studied as references in the context of "citing" practices, whereas the citation index inverts the directionality and studies "citedness" as a measure of impact. From the perspective of STS, the citation index thus would generate a semiotic artifact (Luukkonen, 1997).

References can have different functions in texts, such as legitimating research agendas, warranting knowledge claims, black-boxing discussions, or be perfunctory. In and among texts, references can

---

[20] http://altmetrics.org/, http://www.altmetric.com/
[21] http://blogs.nature.com/
[22] http://scienceseeker.org/
[23] http://f1000.com/prime

also be compared with the co-occurrences and co-absences of words in a network model of science (Braam, Moed, & van Raan, 1991a, 1991b) A network theory of science was formulated by Hesse (1980, p. 83) as "an account that was first explicit in Duhem and more recently reinforced in Quine. Neither in Duhem nor in Quine, however, is it quite clear that the netlike interrelations between more observable predicates and their laws are in principle just as subject to modifications from the rest of the network as are those that are relatively theoretical." A network can be visualized, but can also be formalized as a matrix. The eigenvectors of the matrix span the latent dimensions of the network.

There is thus a bifurcation within the discipline of scientometrics. On the one hand, and by far the dominant partner, we have the relatively positivistic, quantitative analysis of citations as they have happened, after the fact so to speak. And on the other, we have the sociological, and often constructivist theorising about the generation of citations – a theory of citing behaviour. Clearly the two sides are, and need to be linked. The citing behaviour, as a set of generative mechanisms (Bhaskar, 1978), produces the citation events but, at the same time, analyses of the patterns of citations as "demi-regularities" (Lawson, 1997) can provide insights into the processes of scientific communication which can stimulate or validate theories of behaviour.

Another interesting approach is to consider the overall process as a social communication system. One could use Luhmann's (1995, 1996) theory of society as being based on autopoietic communication (Leydesdorff, 2007; Mingers, J., 2002). Different functional subsystems within society, e.g., science, generate their own organizationally closed networks of recursive communications. A communicative event consists of a unity of information, utterance and understanding between senders and receivers. Within the scientometrics context, the paper, its content and its publication would be the information and utterance, and the future references to it in other papers would be the understanding that it generates. Such communication systems operate at their own emergent level distinct from the individual scientists who underlie them, and generate their own cognitive distinctions that can be revealed the visualisation procedures discussed above.

35

References

Aad, G., Abajyan, T., Abbott, B., Abdallah, J., Abdel Khalek, S., Abdelalim, A. A., et al. (2012). Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B, 716*, 1-29.

Abramo, G., & D'Angelo, C. (2011). Evaluating research: from informed peer review to bibliometric. *Scientometrics, 87*, 499-514.

Abramo, G., D'Angelo, C., & Di Costa, F. (2011). National research assessment exercises: A comparison of peer review and bibliometric rankings. *Scientometrics, 89*, 929-941.

Adriaanse, L., & Rensleigh, C. (2013). Web of Science, Scopus and Google Scholar. *The Electronic Library, 31*, 727-744.

Ahlgren, P., Jarneving, B., & Rousseau, R. (2003). Requirements for a cocitation similarity measure, with special reference to Pearson's correlation coefficient. *Journal of the American Society for Information Science and Technology, 54*, 550-560.

Aksnes, D. W., Schneider, J. W., & Gunnarsson, M. (2012). Ranking national research systems by citation indicators. A comparative analysis using whole and fractionalised counting methods. *Journal of Informetrics, 6*, 36-43.

Alonso, S., Cabrerizo, F. J., Herrera-Viedma, E., & Herrera, F. (2009). h-Index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of Informetrics, 3*, 273-289.

Amara, N., & Landry, R. (2012). Counting citations in the field of business and management: why use Google Scholar rather than the Web of Science. *Scientometrics, 93*, 553-581.

Archambault, É., Vignola-Gagné, É., Côté, G., Larivière, V., & Gingras, Y. (2006). Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. *Scientometrics, 68*, 329-342.

Batista, P., Campiteli, M., Kinouchi, O., & Martinez, A. (2006). Is it possible to compare researchers with different scientific interests? *Scientometrics, 68*, 179-189.

Bergstrom, C. (2007). Measuring the value and prestige of scholarly journals. *Coll Res Libr News, 68*, 3146.

Bhaskar, R. (1978). *A Realist Theory of Science*. Hemel Hempstead: Harvester.

Björneborn, L., & Ingwersen, P. (2004). Toward a basic framework for webometrics. *Journal of the American Society for Information Science and Technology, 55*, 1216-1227.

Blondel, V., Guillaume, J.-L., Lambiotte, R., & Lefebvre, L. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, 2008*, P10008.

Bollen, J., Van de Sompel, H., Hagberg, A., Bettencourt, L., Chute, R., Rodriguez, M., et al. (2009). Clickstream data yields high-resolution maps of science. *PLoS ONE, 4*, e4803.

Bornman, L., & Daniel, H.-D. (2005). Does the h-index for ranking of scientists really work? *Scientometrics, 65*, 391-392.

Bornmann, L. (2010). Towards an ideal method of measuring research performance: Some comments to the Opthof and Leydesdorff (2010) paper. *Journal of Informetrics, 4*, 441-443.

Bornmann, L. (2014). Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. *Journal of Informetrics, 8*, 895-903.

Bornmann, L., & Daniel, H.-D. (2007). What do we know about the h index? *Journal of the American Society for Information Science and Technology, 58*, 1381-1385.

Bornmann, L., & Leydesdorff, L. (2014). Scientometrics in a changing research landscape. *EMBO reports, 15*, 1228-1232.

Bornmann, L., Leydesdorff, L., & Mutz, R. (2013). The use of percentiles and percentile rank classes in the analysis of bibliometric data: Opportunities and limits. *Journal of Informetrics, 7*, 158-165.

Bornmann, L., Leydesdorff, L., & Wang, J. (2013). How to improve the prediction based on citation impact percentiles for years shortly after the publication date? . *Journal of Informetrics, 8*, 175-180.

Bornmann, L., Leydesdorff, L., & Wang, J. (2013). Which percentile-based approach should be preferred for calculating normalized citation impact values? An empirical comparison of five approaches including a newly developed citation-rank approach (P100). *Journal of Informetrics, 7*, 933-944.

Bornmann, L., & Mutz, R. (2011). Further steps towards an ideal method of measuring citation performance: The avoidance of citation (ratio) averages in field-normalization. *Journal of Informetrics, 5*, 228-230.

Bornmann, L., Mutz, R., & Daniel, H. D. (2008). Are there better indices for evaluation purposes than the h index? A comparison of nine different variants of the h index using data from biomedicine. *Journal of the American Society for Information Science and Technology, 59*, 830-837.

Boyack, K., Klavans, R., & Börner, K. (2005). Mapping the Backbone of Science. *Scientometrics, 64*, 351-374.

Braam, R., Moed, H., & van Raan, A. (1991a). Mapping of science by combined co-citation and word analysis I: Structural aspects. *Journal of the American Society for Information Science, 42*, 233-251.

Braam, R., Moed, H., & van Raan, A. (1991b). Mapping of science by combined co-citation and word analysis. II: Dynamical aspects. *Journal of the American Society for Information Science, 42*, 252-266.

Bradford, S. (1934). Sources of information on specific subjects. *Engineering, 137*, 85-86.

Braun, T., Glänzel, W., & Schubert, A. (1985). *Scientometric Indicators: A 32-Country Comparison of Publication Productivity and Citation Impact*. Singapore: World Scientific Publishing.

Braun, T., Glänzel, W., & Schubert, A. (2006). A Hirsch-type index for journals. *Scientometrics, 69*, 169-173.

Brumback, R. (2008). Worshipping false idols: the impact factor dilemma. *Journal Child Neurology, 23*, 365-367.

Burrell, Q. (2001). Stochastic modelling of the first-citation distribution. *Scientometrics, 52*, 3-12.

Burrell, Q. (2002). The nth-citation distribution and obsolescence. *Scientometrics, 53*, 309-323.

Burrell, Q. (2003). Predicting future citation behaviour. *Journal of the American Society for Information Science, 54*, 372-378.

Burrell, Q. (2007). Hirsch index or Hirsch rate? Some thoughts arising from Liang's data. *Scientometrics, 73*, 19-28.

Butler, L., & McAllister, I. (2009). Metrics or peer review? Evaluating the 2001 UK Research assessment exercise in political science. *Political Studies Review, 7*, 3-17.

Callon, M., Courtial, J.-P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information, 22*, 191-235.

Cambell, D., Archambaulte, E., & Cote, G. (2008). Benchmarking of Canadian Genomics: 1996-2007. In.

Cameron, B. (2005). Trends in the usage of ISI bibliometric data: Uses, abuses and implications. *Libraries and the Academy, 5*, 105-125.

Campanario, J. M. (2011). Empirical study of journal impact factors obtained using the classical two-year citation window versus a five-year citation window. *Scientometrics, 87*, 189-204.

Chapman, A. (1989). Assessing research:Citation-count shortcomings. *The Psychologist: Bulletin of the British Psychological Society, 8*, 339-341.

Colliander, C. (2014). A novel approach to citation normalization: A similarity-based method for creating reference sets. *Journal of the Association for Information Science and Technology, online*.

Costas, R., & Bordons, M. (2007). The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. *Journal of Informetrics, 1*, 193-203.

Cozzens, S. (1989). What do citations count? The rhetoric-first model. *Scientometrics, 15*, 437-447.

Crane, D. (1972). *Invisible Colleges*. Chicago: University of Chicago Press.

Cronin, B. (1998). Metatheorizing citation. *Scientometrics, 43*, 45-55.

Cronin, B. (2001). Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices? *Journal of the American Society for Information Science and Technology, 52*, 558-569.

Cuellar, M. J., Takeda, H., & Truex, D. P. (2008). The Hirsch family of bibliometric indices as an improved measure of IS academic journal impact.

Davis, P. M. (2008). Eigenfactor: Does the principle of repeated improvement result in better estimates than raw citation counts? *Journal of the American Society for Information Science and Technology, 59*, 2186-2188.

Day, R. (2014). "The data - It is me!". In B. Cronin & C. Sugimoto (Eds.), *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact* (pp. 47-66). London: MIT Press.

de Bellis, N. (2014). History and evolution of (biblio)metrics. In B. Cronin & C. Sugimoto (Eds.), *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact* (pp. 23-44). London: MIT Press.

De Moya-Anegón, F., Vargas-Quesada, B., Chinchilla-Rodríguez, Z., Corera-Álvarez, E., Munoz-Fernández, F., & Herrero-Solana, V. (2007). Visualizing the marrow of science. *Journal of the American Society for Information Science and Technology, 58*, 2167-2179.

de Nooy, W., Mrvar, A., & Batgelj, V. (2011). *Exploratory Social Network Analysis with Pajek* (2nd ed.). New York, NY: Cambridge University Press.

Delgado-López-Cózar, E., & Cabezas-Clavijo, Á. (2012). Google Scholar Metrics: an unreliable tool for assessing scientific journals. *El profesional de la información, 21*, 419-427.

DeMaria, A., Bax, J., Ben-Yehuda, O., Clopton, P., Feld, G. K., Ginsburg, G., et al. (2008). Highlights of the Year in JACC 2007. *Journal of the American College of Cardiology, 51*, 490-512.

Doreian, P., & Fararo, T. (1985). Structural Equivalence in a Journal Network. *Journal of the American Society for Information Science, 36*, 28-37.

Egghe, L. (2000). A heuristic study of the first citation distribution. *Scientometrics, 48*, 345-359.

Egghe, L. (2005). *Power Laws in the Information Production Process: Lotkaian Informetrics*. Amsterdam: Elsevier.

Egghe, L. (2006). Theory and practice of the g-index. *Scientometrics, 69*, 131-152.

Egghe, L., & Rousseau, R. (1988). Informetrics 87/88. In. Amsterdam: Elsevier.

Elkana, Y., Lederberg, J., Merton, R., Thackray, A., & Zuckerman, H. (1978). Toward a Metric of Science: The Advent of Science Indicators. In. New York: Wiley.

Eysenbach, G. (2011). Can Tweets Predict Citations? Metrics of Social Impact Based on Twitter and Correlation with Traditional Metrics of Scientific Impact. *J Med Internet Res, 13*, e123.

Falagas, M. E., Kouranos, V. D., Arencibia-Jorge, R., & Karageorgopoulos, D. E. (2008). Comparison of SCImago journal rank indicator with journal impact factor. *The FASEB Journal, 22*, 2623-2628.

Fersht, A. (2009). The most influential journals: Impact Factor and Eigenfactor. *Proceedings of the National Academy of Sciences, 106*, 6883-6884.

Franceschet, M. (2010). A comparison of bibliometric indicators for computer science scholars and journals on Web of Science and Google Scholar. *Scientometrics, 83*, 243-258.

Franceschini, F., & Maisano, D. A. (2010). Analysis of the Hirsch index's operational properties. *European Journal of Operational Research, 203*, 494-504.

Furner, J. (2014). The ethics of evaluative bibliometrics. In B. Cronin & C. Sugimoto (Eds.), *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact* (pp. 85-107). London: MIT Press.

García-Pérez, M. A. (2010). Accuracy and completeness of publication and citation records in the Web of Science, PsycINFO, and Google Scholar: A case study for the computation of h indices in Psychology. *Journal of the American Society for Information Science and Technology, 61*, 2070-2085.

Garfield, E. (1955). Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas. *Science, 122*, 108-111.

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science, 178*, 471-479.

Garfield, E. (1979). *Citation Indexing: Its Theory and Application in Science, Technology and Humanities*. New York: Wiley.

Garfield, E. (1998). The impact factor and using it correctly. *Der Unfallchirug, 48*, 413.

Garfield, E., Sher, I., & Thorpie, R. (1964). *The Use of Citation Data in Writing the History of Science*. Philadelphia: Institute for Scientific Information.

Garfield, E., & Sher, I. H. (1963). New factors in the evaluation of scientific literature through citation indexing. *American Documentation, 14*, 195-201.

Gingras, Y. (2014). Criteria for evaluating indicators. In B. Cronin & C. Sugimoto (Eds.), *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact* (pp. 110-125). London: MIT Press.

Gingras, Y., & Larivière, V. (2011). There are neither "king" nor "crown" in scientometrics: Comments on a supposed "alternative" method of normalization. *Journal of Informetrics, 5*, 226-227.

Glänzel, W. (2006). On the h-index - a mathematical approach to a new measure of publication activity and citation impact. *Scientometrics, 67*, 315-321.

Glänzel, W., & Schoepflin, U. (1995). A bibliometric study of ageing and reception processes of scientific literature. *Journal of Information Science, 21*, 37-53.

Glänzel, W., & Schubert, A. (1985). Price distribution: An exact formulation of Price's "square root law". *Scientometrics, 7*, 211-219.

Glänzel, W., Schubert, A., Thijs, B., & Debackere, K. (2011). A priori vs. a posteriori normalisation of citation indicators. The case of journal ranking. *Scientometrics, 87*, 415-424.

González-Pereira, B., Guerrero-Bote, V. P., & Moya-Anegón, F. (2010). A new approach to the metric of journals' scientific prestige: The SJR indicator. *Journal of Informetrics, 4*, 379-391.

Guerrero-Bote, V. P., & Moya-Anegón, F. (2012). A further step forward in measuring journals' scientific prestige: The SJR2 indicator. *Journal of Informetrics, 6*, 674-688.

Guns, R., & Rousseau, D. (2009). Real and rational variants of the h-index and g-index. *Journal of Informatrics, 3*, 64-71.

Harzing, A.-W., Alakangas, S., & Adams, D. (2014). hIa: an individual annual h-index to accommodate disciplinary and career length differences. *Scientometrics, 99*, 811-821.

Harzing, A.-W., & van der Wal, R. (2008). Google Scholar as a new source for citation analysis. *Ethics in Science and Environmental Politics, 8*, 61-73.

Harzing, A.-W., & Van der Wal, R. (2009). A Google Scholar h-index for journals: An alternative metric to measure journal impact in economics and business? *J. Am. Soc. for Information Science and Technology, 60*, 41-46.

HEFCE. (2008). Counting what is measured or measuring what counts. In: HEFCE.

HEFCE. (2009). Report on the pilot exercise to develop bibliometric indicators for the Research Excellence Framework. In. London: HEFCE.

Hesse, M. (1980). *Revolutions and Reconstructions in the Philosophy of Science*. London: Harvester Press.

Hirsch, J. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America, 102*, 16569-16572.

Hirsch, J. (2007). Does the h-index have predictive power? *Proceedings of the National Academy of Sciences of the United States of America, 104*, 19193–19198.

Holmes, R. (2014). How to win citations and rise in the rankings. In *University Ranking Watch* (Vol. 2014).

Horrobin, D. (1990). The philosophical basis of peer review and the suppression of innovation. *Journal of the American Medical Association, 263*, 1438-1441.

Huutoniemi, K., Klein, J. T., Bruun, H., & Hukkinen, J. (2010). Analyzing interdisciplinarity: Typology and indicators. *Research Policy, 39*, 79-88.

Iglesias, J., & Pecharromán, C. (2007). Scaling the h-index for different scientific ISI fields. *Scientometrics, 73*, 303-320.

Irvine, J., Martin, B., Peacock, T., & Turner, R. (1985). Charting the decline of British science. *Nature, 316*, 587-590.

Jacso, P. (2008). Savvy searching: Google Scholar revisited. *Online Information Review, 32*, 102-114.

Jin, B. (2006). The h-index: An evaluation indicator proposed by scientist. *Science Focus, 1*, 8-9.

Jin, B. (2007). The AR-index: Complementing the h-index. *ISSI Newsletter, 3*, 6.

Jun, B., Liang, L., Rousseau, D., & Egghe, L. (2007). The R- and AR-indices: Complementing the h-index. *Chinese Science Bulletin, 52*, 855-863.

Kamada, T., & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters, 31*, 7-15.

Kay, L., Newman, N., Youtie, J., Porter, A., & Rafols, I. (2014). Patent overlay mapping: Visualizing technological distance. *Journal of the Association for Information Science and Technology, 65*, 2432-2443.

Konkiel, S., & Scherer, D. (2013). New opportunities for repositories in the age of altmetrics. *Bulletin of the American Society for Information Science and Technology, 39*, 22-26.

Kuhn, T. (1970). *The Structure of Scientific Revolutions*. Chicago: Chicage University Press.

Lancho-Barrantes, B. S., Guerrero-Bote, V. P., & Moya-Anegón, F. (2010). What lies behind the averages and significance of citation indicators in different disciplines? *Journal of Information Science, 36*, 371-382.

Larivière, V., Archambault, É., Gingras, Y., & Vignola-Gagné, É. (2006). The place of serials in referencing practices: Comparing natural sciences and engineering with social sciences and humanities. *Journal of the American Society for Information Science and Technology, 57*, 997-1004.

Larivière, V., & Gingras, Y. (2010). On the relationship between interdisciplinarity and scientific impact. *Journal of the American Society for Information Science and Technology, 61*, 126-131.

Lawson, T. (1997). *Economics and Reality*. London: Routledge.

Lehmann, S., Jackson, A. D., & Lautrup, B. E. (2006). Measures for measures. *Nature, 444*, 1003-1004.

Leydesdorff, L. (1986). The development of frames of references. *Scientometrics, 9*, 103-125.

Leydesdorff, L. (1998). Theories of citation? *Scientometrics, 43*, 5-25.

Leydesdorff, L. (2007). Scientific communication and cognitive codification: Social systems theory and the sociology of scientific knowledge. *European Journal of Social Theory, 10*, 375-388.

Leydesdorff, L. (2008). Caveats for the use of citation indicators in research and journal evaluation. *Journal of the American Society for Information Science and Technology, 59*, 278-287.

Leydesdorff, L. (2009). How are New Citation-Based Journal Indicators Adding to the Bibliometric Toolbox? *Journal of the American Society for Information Science and Technology, 60*, 1327-1336.

Leydesdorff, L. (2012). Alternatives to the journal impact factor: I3 and the top-10% (or top-25%?) of the most highly cited papers. *Scientometrics, 92*, 355-365.

Leydesdorff, L. (2013). The revised SNIP indicator of Elsevier's Scopus. *Journal of Informetrics, 7*, 859-860.

Leydesdorff, L., & Bornmann, L. (2011a). How fractional counting of citations affects the impact factor: Normalization in terms of differences in citation potentials among fields of science. *Journal of the American Society for Information Science and Technology, 62*, 217-229.

Leydesdorff, L., & Bornmann, L. (2011b). Integrated impact indicators compared with impact factors: An alternative research design with policy implications. *Journal of the American Society for Information Science and Technology, 62*, 2133-2146.

Leydesdorff, L., & Bornmann, L. (2014). The Operationalization of" Fields" as WoS Subject Categories (WCs) in Evaluative Bibliometrics: The cases of" Library and Information Science" and" Science & Technology Studies". In *arXiv preprint arXiv:1407.7849*: arXiv.

Leydesdorff, L., Bornmann, L., Mutz, R., & Opthof, T. (2011). Turning the tables on citation analysis one more time: Principles for comparing sets of documents. *Journal of the American Society for Information Science and Technology, 62*, 1370-1381.

Leydesdorff, L., Bornmann, L., Opthof, T., & Mutz, R. (2011). Normalizing the measurement of citation performance: principles for comparing sets of documents. In (Vol. 1101.3863. 2011.): arXiv.

Leydesdorff, L., & Felt, U. (2012). "Books" and "book chapters" in the book citation index (BKCI) and science citation index (SCI, SoSCI, A&HCI). *Proceedings of the American Society for Information Science and Technology, 49*, 1-7.

Leydesdorff, L., & Milojevic, S. (2015). Scientometrics. In M. Lynch (Ed.), *International Encyclopedia of Social and Behavioral Sciences*. New York: Elsevier.

Leydesdorff, L., & Opthof, T. (2010). Scopus's source normalized impact per paper (SNIP) versus a journal impact factor based on fractional counting of citations. *Journal of the American Society for Information Science and Technology, 61*, 2365-2369.

Leydesdorff, L., & Opthof, T. (2011). Remaining problems with the "New Crown Indicator" (MNCS) of the CWTS. *Journal of Informetrics, 5*, 224-225.

Leydesdorff, L., Radicchi, F., Bornmann, L., Castellano, C., & Nooy, W. (2013). Field-normalized impact factors (IFs): A comparison of rescaling and fractionally counted IFs. *Journal of the American Society for Information Science and Technology, 64*, 2299-2309.

Leydesdorff, L., & Rafols, I. (2011). Indicators of the interdisciplinarity of journals: Diversity, centrality, and citations. *Journal of Informetrics, 5*, 87-100.

Leydesdorff, L., Rafols, I., & Chen, C. (2013). Interactive Overlays of Journals and the Measurement of Interdisciplinarity on the basis of Aggregated Journal-Journal Citations. *Journal of the American Society for Information Science and Technology, 64*, 2573-2586.

Liang, L. (2006). h-index sequence and h-index matrix: constructions and applications. *Scientometrics, 69*, 153-159.

Lin, J., & Fenner, M. (2013). Altmetrics in Evolution: Defining and Redefining the Ontology of Article-Level Metrics. *Information Standards Quarterly, 25*.

Lotka, A. (1926). Statistics: The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences, 16*, 317-325.

Luhmann, N. (1995). *Social Systems*. Stanford: Stanford University Press.

Luhmann, N. (1996). On the scientific context of the concept of communication. *Social Science Information, 35*, 257-267.

Lundberg, J. (2007). Lifting the crown—citation z-score. *Journal of Informetrics, 1*, 145–154.

.

Luukkonen, T. (1997). Why has Latour's theory of citations been ignored by the bibliometric community? discussion of sociological interpretations of citation analysis. *Scientometrics, 38*, 27-37.

Ma, N., Guan, J., & Zhao, Y. (2008). Bringing PageRank to the citation analysis. *Information Processing and Management, 44*, 800-810.

Mahdi, S., D'Este, P., & Neely, A. (2008). Citation counts: Are they good predictors of RAE scores? In. London: AIM Research.

Marshakova, I. (1973). Bibliographic coupling system based on references. *Nauchno-Tekhnicheskaya Informatsiya Seriya, Ser, 2*, 3-8.

Meade, N., & Islam, T. (1998). Technological forecasting - model selection, model stability and combining models. *Management Science, 44*, 1115-1130.

Meho, L., & Rogers, Y. (2008). Citation counting, citation ranking, and h-index of human-computer interaction researchers: A comparison of Scopus and Web of Science. *Journal American Society for Information Science and Technology, 59*, 1711-1726.

Meho, L., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science, Scopus and Google Scholar. *Journal American Society for Information Science and Technology, 58*, 2105-2125.

Merton, R. (1968). The Matthew effect in science. *Science, 159*, 56-63.

Merton, R. (1988). The Matthew effect in science II: Cumulative advantage and the symbolism of intellectual property. *ISIS, 79*, 606-623.

Michels, C., & Acsmoch, U. (2014). Impact of bibliometric studies on publication behaviour of authors. *Scientometrics, 98*, 369-385.

Mingers, J. (2002). Can social systems be autopoietic? Assessing Luhmann's social theory. *Sociological Review, 50*, 278-299.

Mingers, J. (2008a). Exploring the dynamics of journal citations: modelling with S-curves. *Journal Operational Research Society, 59*, 1013-1025.

Mingers, J. (2008b). Measuring the research contribution of management academics using the Hirsch-index. *Journal Operational Research Society, 60*, 1143-1153.

Mingers, J. (2014). Problems with SNIP. *Journal of Informatics, 8*, 890-894.

Mingers, J., & Burrell, Q. (2006). Modelling citation behavior in Management Science journals. *Information Processing and Management, 42*, 1451-1464.

Mingers, J., & Leydesdorff, L. (2014). Identifying research fields within business and management: A journal cross-citation analysis. *Journal of the Operational Research Society, Online*.

Mingers, J., & Lipitakis, E. (2010). Counting the citations: A comparison of Web of Science and Google Scholar in the field of management. *Scientometrics, 85*, 613-625.

Mingers, J., Macri, F., & Petrovici, D. (2012). Using the h-index to measure the quality of journals in the field of Business and Management. *Information Processing & Management, 48*, 234-241.

Mingers, J., & Willmott, H. (2013). Taylorizing business school research: On the "one best way" performative effects of journal ranking lists. *Human Relations, 66*, 1051-1073.

Mingers, J., & Xu, F. (2010). The drivers of citations in management science journals. *European Journal of Operational Research, 205*, 422-430.

Moed, H. (2007). The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review. *Science and Public Policy, 34*, 575-583.

Moed, H. (2008). UK Research Assessment Exercises: Informed judgements on research quality or quantity? *Scientometrics, 74*, 153-161.

Moed, H. (2010a). CWTS crown indicator measures citation impact of a research group's publication oeuvre. *Journal of Informetrics, 4*, 436-438.

Moed, H. (2010b). Measuring contextual citation impact of scientific journals. *Journal of Informetrics, 4*, 265-277.

Moed, H. (2010c). The Source-Normalized Impact per Paper (SNIP) is a valid and sophisticated indicator of journal citation impact. In *arXiv preprint*: arxiv.org.

Moed, H., & Visser, M. (2008). Appraisal of Citation Data Sources. In. Leiden: Centre for Science and Technology Studies, Leiden University.

Moed, H., Visser, M., & Buter, R. (2008). Development of bibliometric indicators of research quality. In: Centre for Science and Technology Studies, Leiden University.

Moussa, S., & Touzani, M. (2010). Ranking marketing journals using the Google Scholar-based hg-index. *Journal of Informatrics, 4*, 107-117.

Moxham, H., & Anderson, J. (1992). Peer review: A view from the inside. *Science and Technology Policy, 5*, 7-15.

Nacke, O. (1979). Informetrie: Ein neuer Name fur eine neue Disziplin. *Nachrichten fur Dokumentation, 30*, 219-226.

Nalimov, V., & Mulcjenko, B. (1971). *Measurement of Science: Study of the Development of Science as an Information Process*. Washington DC: Foreign Technology Division.

Nederhof, A. (2006). Bibliometric monitoring of research performance in the social sciences and humanities: A review. *Scientometrics, 66*, 81-100.

Newman, M., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E, 69*, 026113.

Nicolaisen, J. (2007). Citation analysis. *Annual Review of Information Science and Technology, 41*, 609-641.

Oppenheim, C. (2007). Using the h-index to rank influential British researchers in Information Science and Librarianship. *Journal of the American Society for Information Science and Technology, 58*, 297-301.

Opthof, T., & Leydesdorff, L. (2010). Caveats for the journal and field normalizations in the CWTS ("Leiden") evaluations of research performance. *Journal of Informetrics, 4*, 423-430.

Oswald, A. (2007). An Examination of the Reliability of Prestigious Scholarly Journals: Evidence and Implications for Decision-Makers. *Economica, 74*, 21-31.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank citation ranking: Bringing order to the web.

Perneger, T. V. (2004). *Relation between online "hit counts" and subsequent citations: prospective study of research papers in the BMJ* (Vol. 329).

Pinski, G., & Narin, F. (1976). Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics. *Information Processing & Management, 12*, 297-312.

Pislyakov, V. (2009). Comparing two "thermometers": Impact factors of 20 leading economic journals according to Journal Citation Reports and Scopus. *Scientometrics, 79*, 541-550.

Price, D. J. d. S. (1963). *Little Science, Big Science*. New York: Columbia University Press.

Price, D. J. d. S. (1965). Networks of scientific papers. *Science, 149*, 510-515.

Price, D. J. d. S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science, 27*, 292-306.

Priem, J. (2014). Altmetrics. In B. Cronin & C. Sugimoto (Eds.), *Beyond Bibliometrics: Harnessing Multidimensional Indicators nof Scholarly Impact* (pp. 263-288). London: MIT Press.

Pritchard, A. (1969). Statistical bibliography or bibliometrics? *Journal of Documentation, 25*, 348-349.

Radicchi, F., & Castellano, C. (2012). Testing the fairness of citation indicators for comparison across scientific domains: The case of fractional citation counts. *Journal of Informetrics, 6*, 121-130.

Rafols, I., Porter, A. L., & Leydesdorff, L. (2010). Science overlay maps: A new tool for research policy and library management. *Journal of the American Society for Information Science and Technology, 61*, 1871-1887.

Rao, C. (1982). Diversity and dissimilarity coefficients: A unified approach. *Theoretical Population Biology, 21*, 24-43.

Reale, E., Barbara, A., & Costantini, A. (2007). Peer review for the evaluation of academic research: lessons from the Italian experience. *Research Evaluation, 16*, 216-228.

Rehn, C., & Kronman, U. (2008). Bibliometric handbook for Karolinska Institutet,. In.

Roemer, R. C., & Borchardt, R. (2012). From bibliometrics to altmetrics A changing scholarly landscape. *College & Research Libraries News, 73*, 596-600.

Rossner, M., Van Epps, H., & Hill, E. (2007). Show me the data. *The Journal of cell biology, 179*, 1091-1092.

Rosvall, M., & Bergstrom, C. (2010). Mapping Change in Large Networks. *PLoS ONE, 5*, e8694.

Rousseau, D. (2006). New developments related to the Hirsch index. *Science Focus, 1*, 23-25.

Ruiz-Castillo, J., & Waltman, L. (2014). Field-normalized citation impact indicators using algorithmically constructed classification systems of science. In.

Saad, G. (2006). Exploring the h-index at the author and journal levels using bibliometric data of productive consumer scholars and business-related journals respectively. *Scientometrics, 69*, 117-120.

Schubert, A., & Glänzel, W. (2007). A systematic analysis of Hirsch-type indices for journals. *Journal of Informetrics, 1*, 179-184.

Seglen, P. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal, 314*, 498-502.

Seglen, P. O. (1992). The skewness of science. *Journal of the American Society for Information Science, 43*, 628-638.

Senge, P. (1990). *The Fifth Discipline: the Art and Practice of the Learning Organization*. London: Century Books.

Sichel, H. (1985). A bibliometric distribution which really works. *Journal of the American Society for Information Science, 36*, 314-321.

Silva, F. N., Rodrigues, F. A., Oliveira Jr, O. N., & da F. Costa, L. (2013). Quantifying the interdisciplinarity of scientific journals and fields. *Journal of Informetrics, 7*, 469-477.

Small, H. (1973). Co-citation in the Scientific Literature: A New Measure of the Relationship between Two Documents. *Journal of the American Society for Information Science, 24*, 265-269.

Thelwall, M., & Vaughan, L. (2004). Webometrics: An introduction to the special issue. *Journal of the American Society for Information Science and Technology, 55*, 1213-1215.

Tijssen, R., de Leeuw, J., & van Raan, A. (1987). Quasi-Correspondence Analysis on Square Scientometric Transaction Matrices. *Scientometrics, 11*, 347-361.

Todeschini, R. (2011). The j-index: a new bibliometric index and mulitvariate comparisons between other bibliometric indices. *Scientometrics, 87*, 621-639.

Truex III, D., Cuellar, M., & Takeda, H. (2009). Assessing scholarly influence: Using the Hirsch indices to reframe the discourse. *Journal of the Association for Information Systems, 10*, 560-594.

Van Eck, N., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84*, 523-538.

van Leeuwen, T. (2006). The application of bibliometric analyses in the evaluation of social science research. Who benefits from it, and why it is still feasible. *Scientometrics, 66*, 133-154.

van Raan, A. (2005a). Comparison of the hirsch-index with standard bibliometric indicators and with peer judgement for 147 chemistry research groups. *Scientometrics, 67*, 491-502.

van Raan, A. (2005b). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics, 62*, 133-143.

van Raan, A. (2005c). Measuring science: Capita selectaof current main issues. In H. Moed, W. Glenzel & U. Schmoch (Eds.), *Handbook of Quantitative Science and Technology Research* (Vol. 19-50). New York: Springer.

van Raan, A., van Leeuwen, T., Visser, M., van Eck, N., & Waltman, L. (2011). Rivals for the crown: Reply to Opthof and Leydesdorff. *Journal of Informetrics, 4*, 431-435.

van Raan, A. J. (2004). Sleeping beauties in science. *Scientometrics, 59*, 467-472.

Van Veller, M., Gerritsma, W., Van der Togt, P., Leon, C., & Van Zeist, C. (2009). Bibliometric analyses on repository contents for the evaluation of research at Wageningen UR. In A. Katsirikou & C. Skiadas (Eds.), *Qualitative and Quantitative Methods in Libraries: Theory and Applications* (pp. 19-26,): World Scientific.

Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., et al. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics, 5*, 14-26.

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E., Tijssen, R. J., Eck, N. J., et al. (2012). The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology, 63*, 2419-2432.

Waltman, L., & van Eck, N. (2013a). Source normalized indicators of citation impact: An overview of different approaches and an empirical comparison. *Scientometrics, 96*, 699-716.

Waltman, L., & van Eck, N. (2013b). A systematic empirical comparison of different approaches for normalizing citation impact indicators. *Journal of Informetrics, 7*, 833-849.

Waltman, L., van Eck, N., van Leeuwen, T., & Visser, M. (2013). Some modifications to the SNIP journal impact indicator. *Journal of Informetrics, 7*, 272-285.

Waltman, L., van Eck, N., van Leeuwen, T., Visser, M., & van Raan, A. (2010). Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics, 5*, 37-47.

Waltman, L., van Eck, N., van Leeuwen, T., Visser, M., & van Raan, A. (2011). Towards a new crown indicator: an empirical analysis. *Scientometrics*, 1-15.

Weinstock, N. (1971). Citation indexes. In A. Kent (Ed.), *Encyclopedia of Library and Information Science* (pp. 16-41). New York: Marcel Dekker.

West, J., Bergstrom, T., & Bergstrom, C. (2010). Big Macs and Eigenfactor scores: Don't let correlation coefficients fool you. *Journal of the American Society for Information Science and Technology, 61*, 1800-1807.

Wilhite, A., & Fong, E. (2012). Coercive citation in academic publishing. *Science, 335*, 542-543.

Wouters, P. (1998). The signs of science. *Scientometrics, 41*, 225-241.

Wouters, P. (2014). The citation: From culture to infrastructure. In B. Cronin & C. Sugimoto (Eds.), *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact* (pp. 47-66). London: MIT Press.

Xu, F., Liu, W. B., & Mingers, J. (2015). New journal classification methods based on the global h-index. *Information Processing & Management, 51*, 50-61.

Zheng, C. (2009). The e-index, complementing the h-index for excess citations. *PLoS ONE, 4*, e5429.

Zipf, G. (1936). *The Psycho-Biology of Language: An Introduction to Dynamic Philology*. London: Routledge.

Zitt, M. (2011). Behind citing-side normalization of citations: some properties of the journal impact factor. *Scientometrics, 89*, 329-344.

Zitt, M., & Small, H. (2008). Modifying the journal impact factor by fractional citation weighting: The audience factor. *Journal of the American Society for Information Science and Technology, 59*, 1856-1860.