

Caveats for the Use of Citation Indicators in Research and Journal Evaluations

Journal of the American Society for Information Science and Technology (forthcoming)

Loet Leydesdorff

Amsterdam School of Communications Research (ASCoR), University of Amsterdam,

Kloveniersburgwal 48, 1012 CX Amsterdam, The Netherlands

loet@leydesdorff.net ; <http://www.leydesdorff.net>

Abstract

Ageing of publications, percentage of self-citations, and impact vary from journal to journal within fields of science. The assumption that citation and publication practices are homogenous within specialties and fields of science is invalid. Furthermore, the delineation of fields and among specialties is fuzzy. Institutional units of analysis and persons may move between fields or span different specialties. The match between the citation index and institutional profiles varies among institutional units and nations. The respective matches may heavily affect the representation of the units. Non-ISI journals are increasingly cornered into “transdisciplinary” Mode-2 functions with the exception of specialist journals publishing in languages other than English. An “externally cited impact factor” can be calculated for these journals. The citation impact of non-ISI journals will be demonstrated using *Science and Public Policy* as the example.

Keywords: indicators, evaluation, research assessment, journal, impact

1. Introduction

The emergence of new search engines for citation analysis like *Google Scholar* and Elseviers's *Scopus* (Meho & Yang, forthcoming) and the increased emphasis on quality control using the list of journals included in the citation indices of the Institute of Scientific Information (Thomson-ISI) has led to the emergence of an evaluation industry. The US National Science Foundation recently launched a new program for the study of science and innovation policies (at <http://www.nsf.gov/pubs/2007/nsf07547/nsf07547.htm>). Science indicators also figure prominently in the Seventh Research Framework Program (FP7) of the European Commission. Furthermore, evaluations using publication and citation counts have become standard practices in national research assessment exercises and evaluations at university, faculty, and even departmental levels.

This increased interest in bibliometric indicators puts an onus on scientometricians. Since earning-power is increasingly important for university positions and tenure may depend on new contracts, relations with the clientele have become business-like and strategic in this field (Gibbons *et al.*, 1994; Nowotny *et al.* 2001). Entry barriers are based on existing relations of trust. Newcomers on the supply side profile themselves with new and additional ranking and search options. However, the results are sometimes conflicting and confusing because of different parameter choices (Anderson *et al.*, 1988; Leydesdorff & Wagner, 2007).

For example, the *Journal-ranking.com* at the Internet launched a new Center for Journal Ranking (CJR) in April 2006. This tool builds on the underlying information provided by the *Journal Citation Reports* (JCR) of the Institute of Scientific Information (ISI). While the traditional impact factors (Sher & Garfield, 1965; Garfield, 1972) use a two-year time window,¹ the CJR enables the user to search impact factors using different periods of time (Garfield, 1998a; Frandsen & Rousseau, 2005) and to rank journals within subject categories. Both the time-frames of the citation measures and the differences in citation practices among fields of science are major sources of variation. Thus, a parameter space is generated in which an analyst has to make reasoned and/or pragmatic choices.

When developing the impact factor, Garfield (1972) originally opted for a shorter time-frame because he thought that a two-year time limit provided a sufficient sample on the basis of work of Martyn & Gilchrist (1968). Garfield (2003) admitted that another reason for the two-year period was that it suited biochemistry and molecular biology—major areas of interest for the ISI. The impact factor normalizes for size because of the division by the number of publications in the denominator and thus gives small journals a better chance of selection. Bensman & Wilder (1998) provided evidence that total citations (i.e., accumulated impact) correlate with perceived quality of journals more than the impact factors (Leydesdorff, 2007a). Total citations can be considered as reflecting the prestige of a journal, while impact factors highlight a journal's current value at one or more research fronts. The two measures are correlated (Spearman's $\rho = 0.73$; $p < 0.01$; JCR

¹ The impact factor for a journal is calculated by dividing the number of current year citations to the source items published in that journal during the previous two years. It can be considered to be the average number of times published papers are cited up to two years after publication (Garfield, 1979, at p. 149).

2005).² Total citations, of course, are more stable over time—since accumulating—than impact factors.

Recently, Hirsch (2005) has proposed the *h*-index as a new indicator for evaluation purposes. This index is defined as the highest rank on a scientist's list of publications such that the first *h* publications received at least *h* citations. (The indicator follows naturally from the ordering of the listings using Google Scholar because the search results are approximately ranked in terms of the number of citations. In order to determine the *h*-index, one scrolls down to the sequence number which equals the number of citations of the author.) The *h*-index can be extended to any set of documents; for example, journals (Braun *et al.*, 2005). In the meantime, the *h*-index has also been included under the so-called "Citation Reports" available online at the ISI's Web-of-Science. For example, 162 documents can be retrieved as published in *JASIST* in 2005. The *h*-index of this set is six, while the index is ten for the 139 documents published by *JASIST* in 2004.³ Unlike the impact factor, the *h*-index cannot decrease for a given set and can thus be considered as an accumulating indicator for lifetime achievement in the case of individual scholars.

As in the case of the impact factor (Egghe & Rousseau, 1990; Moed, 2005; Rousseau, 2004; cf. Seglen, 1997), scientometricians have been prolific in proposing a series of technical improvements to the measurement of the new indicator. Egghe (2006a, 2006b) first developed the *g*-index, Jin (2006) followed with the *A*-index, and, more recently, Jin

² When the *Social Science Citation Index* is excluded from this set, the Spearman's $\rho = 0.71$; $p < 0.01$.

³ These measurements were done on July 27, 2007.

et al. (2007) suggested correcting the *h*-index for the aging of *papers* using the *AR*-index. Burrell (2007) proposed the *h*-rate which intends to normalize the *h*-index of an *author* for one's career length. For pragmatic reasons, Kosmulski (2006) proposed to use not the *h*, but h^2 for cutting the list of (*h*) cited papers in fields of science (like medicine and biology) where the number of citations per article is high.

As in the previous refinements of the impact factor, in my opinion, mathematical sophistication cannot increase the validity of an indicator used in processes of research evaluation. Most importantly, exogenous variation is caused by different citation and publication practices among fields of science (Leydesdorff & Amsterdamska, 1990), but also within fields of science among different publication channels such as reviews, research articles, and letters (Garfield, 1996, 1998b; Bensman, 2007). For example, Garfield (1980, at p. 1A) summarized this conclusion as follows: "I've often stressed the importance of limiting comparisons between journals to those in the same field." Most evaluators pay lip-service to this problem by stating it up front, but then propose a pragmatic solution and proceed to the measurement. Some major research teams have developed their own (in-house) classification schemes which they claim to be robust (Pinski & Narin, 1976; Moed *et al.*, 1985; Glänzel & Schubert, 2003).

For lack of an agreed-upon alternative, the ISI subject categories are often used for "comparing like with like" (Martin & Irvine, 1983). These categories are assigned by the ISI staff on the basis of a number of criteria, including the journal's title, its citation patterns, etc. (McVeigh, *personal communication*, 9 March 2006). The classifications,

however, match poorly with classifications derived from the database itself on the basis of analysis of the principal components of the networks generated by citations (Leydesdorff, 2006a, at pp. 611f.). Using a different methodology, Boyack, Klavans & Börner (2005) found that in somewhat more than 50% of the cases the ISI categories corresponded closely with the clusters based on inter-journal citation relations. These results accord with the expectation: many journals can be assigned unambiguous affiliations in one core set or another, but the remainder which is also a large group is very heterogeneous (Bradford, 1934; Garfield, 1972).

In summary, I listed a number of (partly well-known) problems which I will address below using empirical data. In the next section (section 2), I show that impact factors are different with an order of magnitude when comparing journals in mathematics with journals in genetics. In section 3, I address the other major source of variation in the database: the different time-frames of publication media. Journals which publish letters are compared with those that publish exclusively reviews. The two types of media exhibit significantly different cited journal half-life patterns. Therefore, even within a narrowly defined specialty, one cannot expect publications in these different media to have comparable long- and short-term citation impacts.

Elsewhere (Leydesdorff, 2006a), I have shown that journals cannot unambiguously be classified into journal categories on the basis of their aggregated citation patterns. All sets are overlapping and fuzzy (Bradford, 1934; Garfield, 1971; Bensman, 2007). When one evaluates research programs and research groups on the basis of articles, one always

needs a reference set (Studer & Chubin, 1980, at p. 269). However, decisions about how to cut the cake determine what is included and what is excluded. Furthermore, institutional research programs and even creative individuals can be expected to cross disciplinary boundaries (section 4; cf. Kreft & De Leeuw, 1988).

The preliminary delineation remains that of the selection of journals included in the ISI-databases versus the “non-ISI journals” (Garfield, 1972). This is the subject of section 5. What are the effects of the increasing pressures at the institutional levels to publish exclusively in “ISI journals”? Using *Science & Public Policy* as an example, I shall show what the inclusion of this journal adds to the network of journal-journal citation relations. The non-ISI journals can be retrieved within the ISI-set in terms of their being-cited patterns. What are the effects of exclusion of such a journal on the relevant definitions of specialties?

2. Variation in the impact factors among fields of science

Figures 1 and 2 show the distribution of impact factors for two subject categories of the ISI: “mathematics” ($N = 191$ journals) and “genetics” ($N = 133$). As noted, the journal selection and the ISI-categories are themselves debatable, but for the purpose of this argument the comparison between these two sets is revealing. The values of the x -axis differ not slightly, but by an order of magnitude!

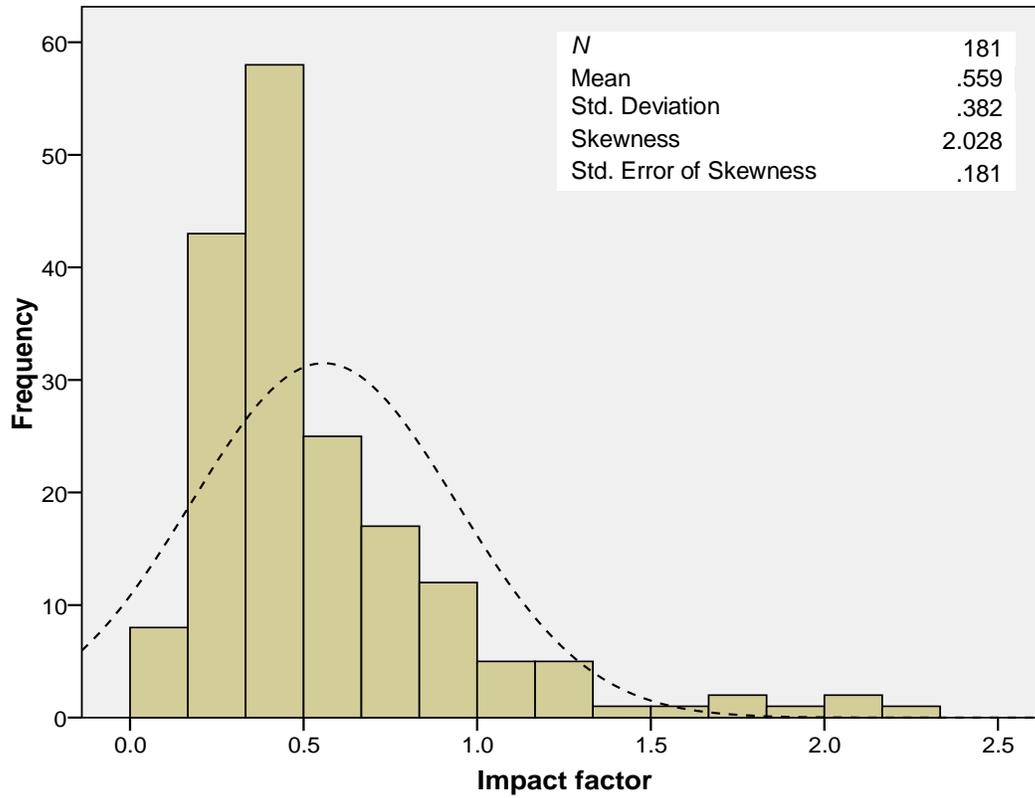


Figure 1: Impact factors of 181 journals in Mathematics according to the classification of the ISI (source: *Journal Citation Reports 2005*).

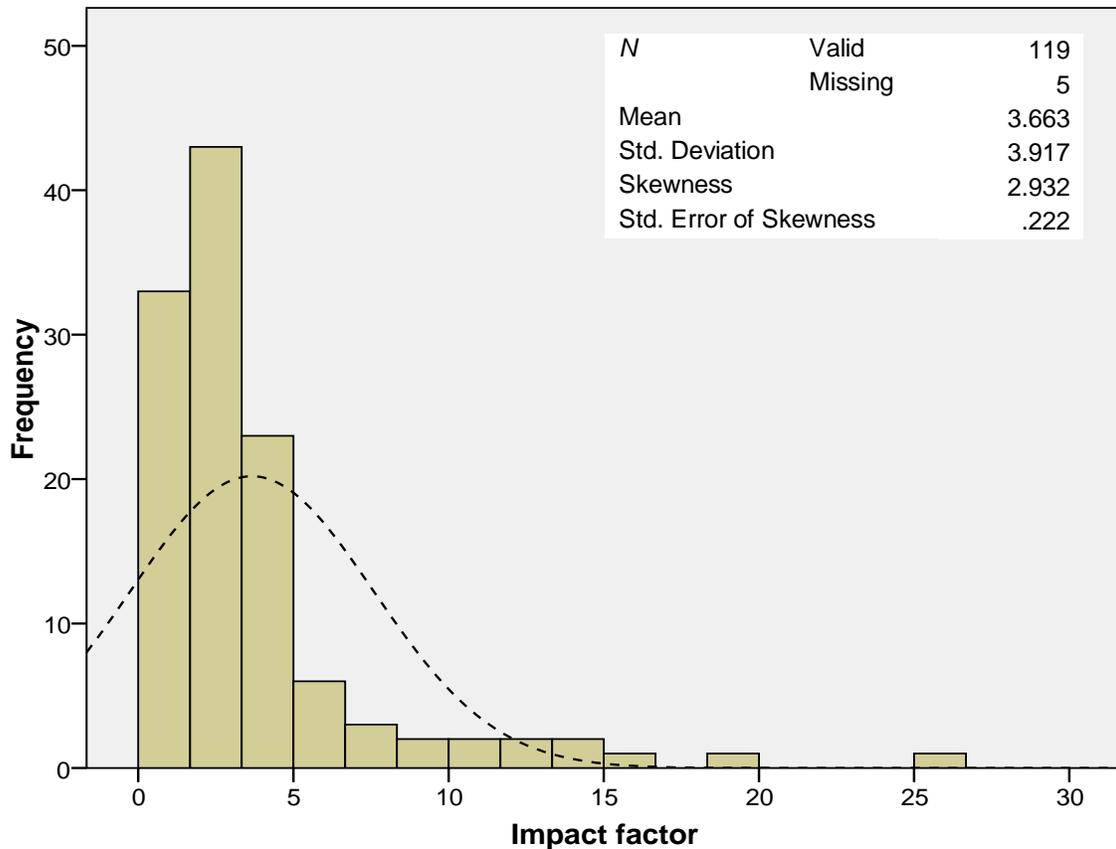


Figure 2: Impact factors of 119 journals in Genetics according to the classification of the ISI (source: *Journal Citation Reports 2005*).

The problem is well-known among scientometricians (e.g., Martin & Irvine, 1983). However, it has an important policy implication: one is not allowed to compare evaluations based on citation rates across disciplines. Consequently, it is not legitimate to allocate funds across fields of science on the basis of comparisons among citation rates for different disciplines (National Research Council, 2007). In other words, one has to disaggregate. The rules for the disaggregation, however, are far from obvious.

3. Variation in the cited journal half-life among journal types

In addition to the static variation in publication and citation behavior among fields of science, the impact factor and the recently proposed modifications of the *h*-index (Jin *et al.*, 2007; Burrell, 2007) incorporate the time dimension. All indicators imply a choice for a time-frame, but sometimes this choice is left implicit. Scientometricians often use a so-called citation window for the standardization (Moed, 2005). As a standard for the aging effect of journal publications, the ISI provides the so-called “cited half-life” of a journal. The cited half-life for the journal is the median age of its articles cited in the current year. In other words, half of the citations to the journal are to articles published within the cited half-life (Rousseau, 2006b). One can expect these measures also to differ among fields of science (Price, 1970). In the humanities and the social sciences, research fronts are often virtually absent (Nederhof *et al.*, 1989).

Independently of the differences among fields of science, publications come in different types: articles, letters, and reviews. This source of variation can be expected to influence the measurement of citation rates across fields of science. By using a relatively short time window, one can expect to favor letters as compared with other publications. Similarly, the ranking of reviews and review journals is negatively affected by the decision to focus on the dynamics of the last two years at research fronts. Accumulating indicators like measures for life-long achievements tend to work the other way round. The various types of publications have both different impacts and life-cycles.

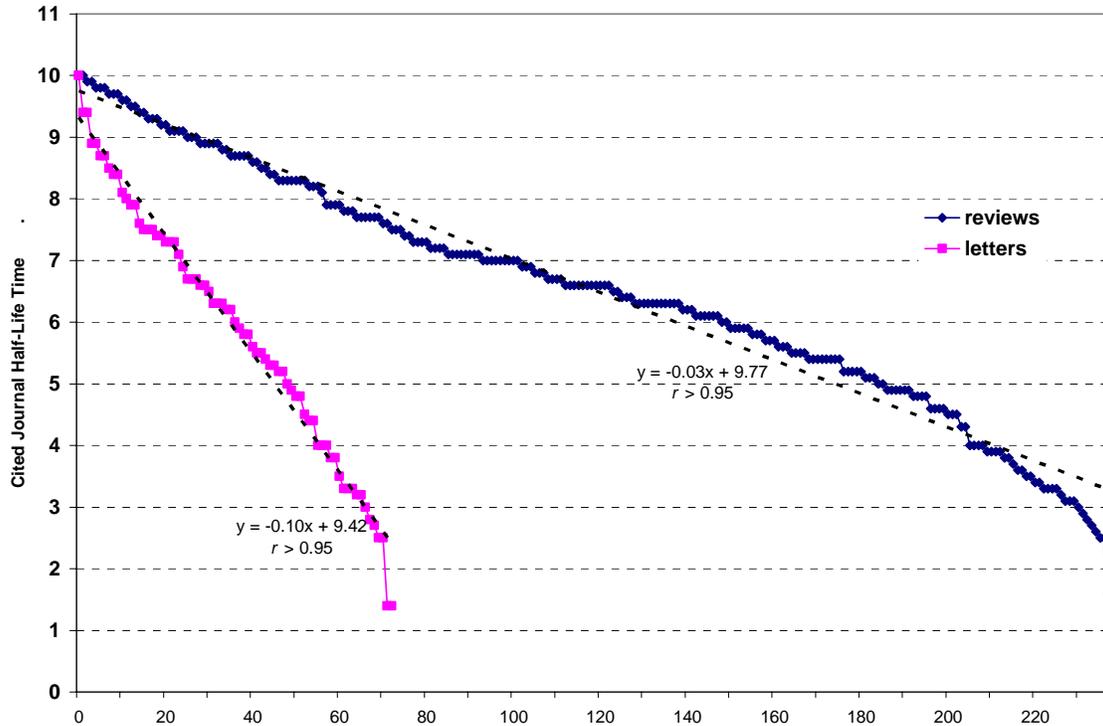


Figure 3: Descending Cited Journal Half-Life of 238 journals with “Review” and 73 with “Letters” among their title words (source: *Journal Citation Reports*, 2005).⁴

Figure 3 substantiates this point empirically by showing the effects of sorting letters and review journals in terms of their cited half-life in descending order (using the *Journal Citation Reports* of 2005). A one-way analysis of variance using the two subsets of journals with the words “Reviews” or “Letters” in their titles, versus the third subset of remaining journals shows that the impact factors of review journals is significantly higher ($p < 0.01$) than those of other journals (Table 1). The cited half-life times of review journals are not significantly higher than those of other journals at the one-percent level, but the cited half-life times of letters are on average more than twice as low. As can be

⁴ Additionally, 90 journals with “Review” in their title and three with “Letters” have no cited half-life times listed in the *JCR* 2005 because they extend beyond 10 years. Four more journals with “Letters” in their title have no cited half-times listed.

expected, the correlation between impact factors and cited half-life times over the file is negative: $\rho = -0.239$ ($p < 0.01$; *JCR2005*).

		N	Mean	Standard Deviation	Standard Error	95% Confidence Interval for Mean	
						Lower Bound	Upper Bound
Impact Factor	“Reviews”	326	4.0081	6.324776	0.350297	3.31897	4.69724
	“Letters”	79	1.79813	1.902339	0.21403	1.37203	2.22423
	Other journals	7386	1.47353	2.022984	0.023539	1.42739	1.51967
	Total	7791	1.58287	2.41719	0.027385	1.52919	1.63656
Cited Half-life	“Reviews”	294	25.24932	37.507056	2.187456	20.9442	29.55444
	“Letters”	76	9.57632	18.5418	2.12689	5.33933	13.8133
	Other journals	6790	24.22607	36.954594	0.44847	23.34693	25.10521
	Total	7160	24.11258	36.858479	0.435593	23.25869	24.96647

Table 1: One-way Analysis of Variance of the impact factors and cited half-life times for journals with the words “Reviews” or “Letters” in their titles, compared with the remainder of the *JCR*-set.

In summary, at the level of the ISI-set, the major problems in evaluation are the delineation and the assumption of homogeneity in the sets to be evaluated. All sets are overlapping and fuzzy (Bradford, 1934; Garfield, 1971; Bensman, 2007). Within relatively homogenous sets, different media of communication (articles, letters, and reviews) can be expected to generate different citation dynamics. For analytical purposes, one may wish to confine studies in the sociology of science to homogenous sets like only research articles in a restricted set of core journals, but for the practical purpose of evaluation such analytical restrictions may be counterproductive.

4. Institutions and individuals

Unlike analytical units of analysis like disciplines and specialties, social institutions have the function to integrate different perspectives in research programs. If one wishes to compare—or even benchmark—institutions in terms of citation rates, the above figures

suggest that a university could be advised to close down its mathematics department in order to increase its standing, for example, on the national ranking (Leydesdorff & Zhou, 2005). In summary, institutions are the wrong units of analysis for scientometric comparisons in terms of delineated journal sets because they are intellectually heterogeneous (Collins, 1985).

Recently, Braam (2007) proposed to use “journal scope change” in the publication profile of research programs and/or research groups as an indicator for intellectual growth. However, this begs the question of how to calibrate change at the level of a group against change at the level of relevant fields (Whitley, 1984; Bonaccorsi, 2005). If a group rows in terms of its publications against the tide at the journal level, chances of success are different from when one has an opportunity to ride the wave of new developments (Leydesdorff & Van der Schaar, 1987; Rafols & Meyer, 2007). In an established specialty, it may be a more rewarding strategy to focus on core journals or, in other words, to be less innovative in terms of exploring interdisciplinary relations. Furthermore, the trade-off may be different for individuals, research groups and/or (national) research programs.

Nations can be considered as aggregated institutional units of analysis. Thus, what holds for institutions holds also for comparisons among nations (King, 2004). For example, the research portfolios of East Asian countries are less developed in the bio-sciences than those of the USA and most Western European countries (Park *et al.*, 2005). The ISI database, however, is more strongly developed in the life sciences than in the physical

sciences. Thus, the match of East Asian countries with the database is weaker than for some Western countries. The latter will for this reason enjoy an advantage in terms of their visibility (and thus, their chances of being cited) within the database.

While nations are macro-units of institutional analysis, individuals can perhaps be considered as minimal units of analysis. Individuals may also be heterogeneous in terms of their scientific output (Hellsten *et al.*, forthcoming). In a certain sense, creative researchers are supposed to generate new variations or, in other words, knowledge claims that reach beyond existing borderlines. Bar-Ilan (2006) found that Google Scholar reflected this heterogeneity more than the ISI-databases in the case of the information sciences. However, one is not allowed to apply statistics as a predictor in individual cases. Individuals almost by definition deviate from the mean.

Finally, a theoretical reflection is here in place. The prevailing conceptualization in the sociology of science in terms of levels (group versus field; context of discovery versus context of justification) misses the communication dynamics between these two dimensions of the scientific enterprise. Publications contain knowledge claims that compete for proving their value at the level of (one or more) scientific discourses. Citations can be considered indicators of diffusion at the network level and cannot inform us about the intrinsic quality of research at the site of production. Knowledge claims in publications provide the variation.

Publication counts can inform us about the *quantity* of the production. Citation counts reflect the usefulness of the publication in the construction of new knowledge claims at *other* places in the network (Fujigaki, 1998). However, networks of citations among journals are structural (Price, 1965). They are reproduced from year to year and can thus be expected to exhibit their own dynamics. It is not only the intrinsic quality of specific publications, but also their position and timing in the distribution that can be expected to determine how the networks will absorb new variants (Leydesdorff, 1998).

5. Non-ISI journals

Many editors and publishers feel a sense of frustration if their journal is not included in the ISI dataset. The ISI provides a number of criteria for inclusion such as (1) three consecutive issues must be published on time; (2) how many citations the journal has had in the previous two years in other journals of the *Science Citation Index* or the *Social Science Citation Index*, respectively; and (3) special factors, such as journals which appeal especially to decision-makers. The latter users are considered influential, but they do not write academic papers with systematic references (Page, *personal communication*, 16 February 2007). The criteria are weighted and this procedure is a company secret of the ISI (Garfield, 1990; Testa, 1997). As can be expected, the ISI is under pressure from lobbying agencies (Moravcsik, 1985; Arvinitis & Gaillard, 1992) and publishing houses to include journals on various grounds (Tijssen & Van Leeuwen, 1995; Hicks, 1999; Maricic, 1997).

Inclusion in this database has become increasingly important for journals in recent years because of the strong pressures on institutions and authors, particularly in Asia, to publish in journals with impact factors that are included in the ISI-databases. Using the ISI-database, however, it is possible to calculate a quasi impact factor for journals not included in the ISI set because all the citations—including those to non-source journals—are part of the database (Stegmann, 1997; 1999). This so-called “externally-cited impact factor” (Christensen *et al.*, 1997, at p. 536) underestimates the real impact factor because in this case one does not include “within-journal self-citations” (Leydesdorff, 2007a).⁵ This can lead to a considerable under-representation: 124 of the 299 citations (41.5%) in 2005 to the 2003 and 2004 volumes of *Scientometrics*, for example, were provided by authors publishing in this same journal. For *JASIST*, this percentage was 20.8% in 2005.

In order to construct this externally-cited impact factor, one can search on the Web of Science, for example, with “Sci Publ Policy”—that is, the abbreviation for *Science and Public Policy*—as the cited work, a specific year (e.g., 2005) for the citations, and the two preceding years for the cited years. This provides not the citation score, but the number of unique article linkages generated by citations. However, each citing article may cite different sources from the same journal. By aggregating all the citations from the citing journals to the cited journal using the JCR, one can then construct an impact factor. This routine can also be automated. The externally-cited impact factor 2005 for *Science and Public Policy* is $(25/79 =) 0.316$.⁶ (Let me follow the standard practice of

⁵ Within-journal self-citations are not necessarily self-citations by authors, but they provide us with a measure of how much a journal is used by a scientific community for an inward-directed discourse (Leydesdorff, 2007a and b).

⁶ The number of citable issues was kindly provided by the publisher of *SPP*.

providing impact factors with three decimals although one should be aware that this impressive precision is part of the codification process surrounding impact factors.)

The citation impact environment of *Science & Public Policy* is visualized in Figure 4 using techniques developed in other contexts (Leydesdorff, 2007a and b). *SPP* is cited in 2005 by articles in 29 other journals. Twenty journals are included in its citation environment above the threshold of cosine = 0.2.⁷ The figure is interesting because it provides a more integrated view of the field of Science & Technology Studies (STS) than one is able to generate by using one of the leading journals of the field as a seed journal.

⁷ The cosine is increasingly used as a similarity criteria because unlike the Pearson correlation the normalization is not to the arithmetic, but to the geometrical mean. On the one hand, this solves the problem of highly skewed (i.e., non-normal) distributions, and, on the other hand, the cosine values can be taken as input to the vector-space-model for the visualization (Salton & McGill, 1983; Ahlgren *et al.*, 2003; Leydesdorff, 2007a).

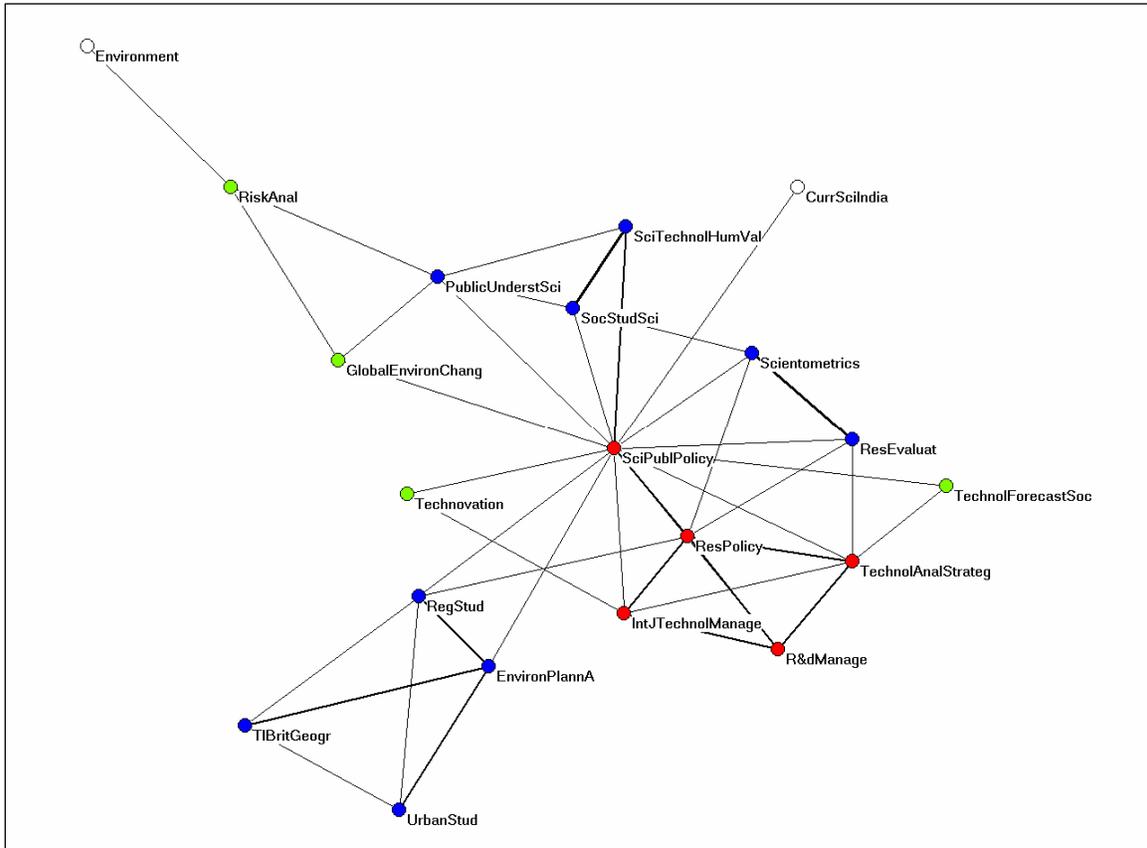


Figure 4: The citation impact environment of *Science and Public Policy*. (JCR 2005; $N = 20$; cosine ≥ 0.2).

During the last decade, the journal *Scientometrics* has become increasingly embedded in the information sciences (Figure 5). Unlike *JASIST* which is extensively cited in both information and computer science journals (to a total of 134 journals in 2005), *Scientometrics* has remained a specialist journal at the interface between information science and STS (cited only in 34 other journals in 2005). Two other core journals of STS—*Social Studies of Science* and *Research Policy*—have hardly any citation traffic between them. As increasingly an *interdiscipline*, STS is oriented towards and cited by journals in a variety of other disciplines (Leydesdorff & Van den Besselaar, 1997; Van den Besselaar, 2001; Leydesdorff, 2007b).

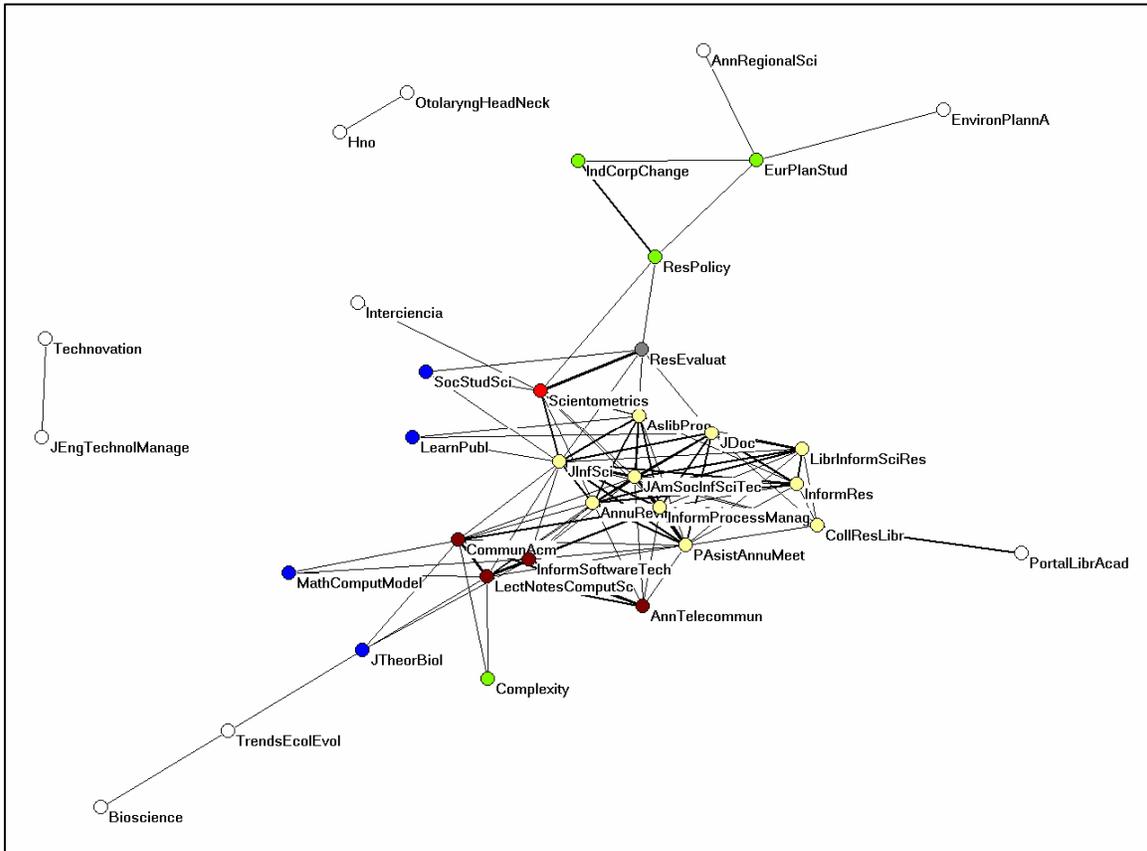


Figure 5: The citation impact environment of *Scientometrics*. (JCR 2005; $N = 34$; cosine ≥ 0.2).

Table 2 provides the values of the impact factors and the externally-cited impact factors—corrected for within-journal self-citations—for the 20 journals in the citation impact environment of *SPP*. The impact factor of *SPP* remains the lowest among them also after correction for “within-journal self-citations.” This is a consequence of being a central node in an otherwise weakly connected network. However, the “betweenness centrality” of *SPP* in this environment is 69.9%, while this value is only 4.4% for *Scientometrics* in the ego-network of this journal as depicted in Figure 5. In other words, *SPP* fulfills an important function in this interdisciplinary network of journals, but this is

not reflected in its impact factor because of the thinness of this network (Freeman, 1977; Hanneman & Riddle, 2005; Leydesdorff, 2007c).

2005	Impact factor (ISI)	Externally-cited impact factor
<i>Current Science</i>	0.728	0.492
<i>Environment and Planning A</i>	1.367	1.222
<i>Environment</i>	1.020	0.571
<i>Global Environmental Change--Human and Policy Dimensions</i>	1.952	1.629
<i>International Journal of Technology Management</i>	0.240	0.194
<i>Public Understanding of Science</i>	0.913	0.739
<i>R & D Management</i>	0.506	0.454
<i>Regional Studies</i>	1.525	1.258
<i>Research Evaluation</i>	0.474	0.421
<i>Research Policy</i>	1.835	1.125
<i>Risk Analysis</i>	1.510	1.217
<i>Science & Public Policy</i>	<i>n.a.</i>	0.316
<i>Science Technology & Human Values</i>	1.439	1.366
<i>Scientometrics</i>	1.738	1.017
<i>Social Studies of Science</i>	0.929	0.768
<i>Transactions of The Institute of British Geographers</i>	2.218	1.927
<i>Technology Analysis & Strategic Management</i>	0.446	0.410
<i>Technological Forecasting and Social Change</i>	0.811	0.444
<i>Technovation</i>	0.497	0.377
<i>Urban Studies</i>	0.988	0.752

Table 2: ISI-Impact factors and quasi impact factors of journals in the citation environment of *Science & Public Policy*.

It is not possible to generalize to all non-ISI journals because this is a very heterogeneous set, including among others also newspapers. The *Journal Citation Reports* 2005 contain 282,955 references to non-ISI sources. The analysis was hitherto only done on a case-by-case basis. On the basis of a large number of case studies of these non-ISI journals (Bornmann et al., 2007; Zhou & Leydesdorff, 2007a and b), I conjecture that scientific journals which are not included in the ISI database—that is, the so-called B-journals in evaluation studies—fall increasingly into two major groups. First, there are journals which like *SPP* fulfill transdisciplinary functions, but—unlike general journals such as

Science and *Nature*—at the micro-level and on the margins of different specialties.

Among the medical journals, these can, for example, be journals which focus on specific diseases. For example, the *Journal of Alzheimer's Disease* has recently been included in the *SCI-Expanded* at the web, but is not yet included in the JCR-set.

The second group consists of disciplinary journals which were selected away in favour of other journals in the same cluster. These may include journals publishing in languages other than English,⁸ since virtually all papers in the *Science Citation Index* nowadays are in English (Table 3).

⁸ Journals publishing in both English and other languages can be overrepresented in the database because authors sometimes cite both editions. Bornmann *et al.* (2007), for example, found an overestimation of impact factor of *Angewandte Chemie* of 21.5% because of this effect.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	English	998635	98.7	98.7	98.7
	German	4941	.5	.5	99.2
	French	2368	.2	.2	99.5
	Chinese	2074	.2	.2	99.7
	Spanish	1156	.1	.1	99.8
	Japanese	926	.1	.1	99.9
	Russian	861	.1	.1	100.0
	Czech	165	.0	.0	100.0
	Multi-Lang	83	.0	.0	100.0
	Finnish	62	.0	.0	100.0
	Portuguese	24	.0	.0	100.0
	Romanian	18	.0	.0	100.0
	Latvian	14	.0	.0	100.0
	Welsh	12	.0	.0	100.0
	Italian	10	.0	.0	100.0
	Slovak	6	.0	.0	100.0
	Afrikaans	3	.0	.0	100.0
	Dutch	2	.0	.0	100.0
	Danish	1	.0	.0	100.0
	Gaelic	1	.0	.0	100.0
	Serbian	1	.0	.0	100.0
	Total	1011363	100.0	100.0	

Table 3: language distribution of papers included in the CD-Rom version of the *Science Citation Index 2005*.

Perhaps, it may be increasingly difficult for journals that publish in English and are not included in the ISI database to maintain a position within a disciplinary setting. A division between disciplinary (“Mode-1”) and transdisciplinary (“Mode-2”) journals might thus further be reinforced (Gibbons *et al.*, 1994; Leydesdorff & Jin, 2005).

6. Discussion and conclusions

Can journals and journal rankings be used for the evaluation of research? Garfield's (1972 and 1979) original purpose when creating the impact factor and the *Journal Citation Reports* was not to evaluate research, but journals! Based on Bradford's (1934) Law of Scattering, Garfield (1971) formulated his Law of Concentration which states that the tail of the literature of one discipline consists, in a large part, of the cores of the literature of other disciplines (Garfield, 1979, at p. 23): "So large is the overlap between disciplines, in fact, that the core literature for all scientific disciplines involves a group of no more than 1000 journals, and may involve as few as 500."

Thus, the ISI journal selection is based on its purposefulness for information retrieval. A journal which links literatures together, like the above example as *SPP*, may demonstrate the "strength of weak ties" (Granovetter, 1973), but the intellectual and social organization of the sciences is not the primary concern of the providers of this database. The impact factors provide a summary statistics which conveniently allows for the ranking of journals. However, these statistics are based on the means of a highly skewed distribution. Elsewhere, Leydesdorff & Bensman (2006) explained how the organization of the journal set in subsets can be expected to lead to a compounded distribution: each of the subsets of a scientific journal set has different underlying probabilities and therefore a different expected value or arithmetic mean. I showed this empirically for the cases of mathematics and genetics: the means differed with an order of magnitude. In the case of compounded distributions, heterogeneity and contagion act multiplicatively instead of

additively, creating exponential and curvilinear relationships instead of the assumed additive, linear ones. In other words, an impact factor of two is not twice as good as an impact factor of one. Much depends on the context(s) of the measurement.

Incorrect journal abbreviations and non-ISI sources	Citations
<i>J Phys Chem-US</i>	54,139
<i>Phys Rev</i>	32,352
<i>Biochim Biophys Acta</i>	26,108
<i>Communication</i>	22,062
<i>Am J Physiol</i>	14,716
<i>Unpub</i>	14,020
<i>Am J Med Genet</i>	13,467
<i>J Bone Joint Surg</i>	13,405
<i>J Biomed Mater Res</i>	12,962
<i>J Chem Soc Perk T 1</i>	11,870
<i>Am Rev Respir Dis</i>	11,033
<i>P Soc Photo-Opt Ins</i>	10,817
<i>Acta Metall Mater</i>	10,310
<i>Mmwr-Morbid Mortal W</i>	10,208

Table 4: Non-ISI sources and incorrect journal abbreviations with more than 10,000 citations in the *JCR* 2005.

In addition to the effects of the selection process on the deselected journals, the otherwise impressively rich database contains also a lot of error. Table 4 shows the fourteen “journal abbreviations” among the non-ISI sources which obtained more than 10,000 citations in 2005. With its 54,139 citations, the *J Phys Chem-US* would belong to the top-50 journals of the database if it were included. However, this journal is included in the ISI-database under the abbreviations *J Phys Chem A* and *J Phys Chem B* with 32,086 and 59,826 citations, respectively. For some journals, however, the different spellings in the references may have large implications. Bornman *et al.* (2007, at p. 105) found 21.5% overestimation of the impact factor of *Angewandte Chemie* in 2005 because of authors

providing references to both the German and international editions of this journal (Marx, 2001). Increasing a journal's impact factor has become an industry in itself (Li, 2006; Park & Leydesdorff, forthcoming).

While it may nevertheless be feasible (in some cases more than in others) to delineate journal sets which provide a fair representation of a specialty, institutional units of analysis are almost never confined to a single and relatively homogenous journal set. One might even put a question mark in terms of a unit's longer-term perspective if it were publishing exclusively in a highly specialized set. Policy-makers may wish to encourage researchers to take more risks by funding "interdisciplinary" programs, but an average researcher has only a single chance of making a career in science. Trade-offs between funding opportunities and intellectual perspectives remain difficult for individual scholars, research groups, and societal stakeholders.

Where does this leave us with respect to political and managerial incentives for evaluations? Scientometric evaluations have obtained a function in public policy and the management of R&D. Policy makers and management may be inclined to opt for choices based on relevant information—and probably one should. Indicators can function to inform the various discourses. However, one should be very careful not to throw out the baby with the bathwater on the basis of normative assumptions like the expected scale-effects of large-scale concentrations of R&D on productivity (Adams & Smith, 2003; Von Tunzelman *et al.*, 2003). More often than not, one is able to generate evidence which points in another direction. The emerging knowledge-based economy may have more

need to stimulate variation than to increase selection pressures (Ashby, 1958; Bruckner *et al.*, 1994; Leydesdorff, 2006b).

Acknowledgement

I am grateful to Bill Page, the Editor of *Science & Public Policy*, for providing me with relevant data. I am also grateful to three anonymous referees.

References

- Adams, J., & Smith, D. (2003). Funding research diversity. Leeds: Evidence Ltd, available at http://www.evidence.co.uk/downloads/funding_tech.pdf, last accessed on 27 March 2007.
- Ahlgren, P., Jarneving, B., & Rousseau, R. (2003). Requirement for a Cocitation Similarity Measure, with Special Reference to Pearson's Correlation Coefficient. *Journal of the American Society for Information Science and Technology*, 54(6), 550-560.
- Anderson, J., Collins, P. M. D., Irvine, J., Isard, P. A., Martin, B. R., Narin, F., et al. (1988). On-line approaches to measuring national scientific output-- A cautionary tale, *Science and Public Policy* 15, 153-161.
- Arvanitis, R., & Gaillard, J. (1992). Science Indicators for Developing Countries. Proceedings of the International Conference on Science indicators for Developing Countries. UNESCO: Colloques et Séminaires, Paris, 15-19 October 1990.
- Ashby, R. (1958). Requisite variety and its implications for the control of complex systems. *Cybernetica*, 1(2), 1-17.
- Bar-Ilan, J. (2006). H-index for Price medalists revisited. *ISSI Newsletter*, 2(1), 3-5.
- Bensman, S. J. (2007). Garfield and the Impact Factor. *Annual Review of Information Science and Technology*, 41, 93-155.
- Bensman, S. J., & Wilder, S. J. (1998). Scientific and Technical Serials Holdings Optimization in an Inefficient Market: A LSU Serials Redesign Project Exercise. *Library Resources and Technical Services*, 42(3), 147-242.
- Bonaccorsi, A. (2005). Search regimes and the industrial dynamics of science. Paper presented at the PRIME General Assembly, Manchester, 7-9 January 2005.
- Boyack, K. W., Klavans, R., & Börner, K. (2005). Mapping the Backbone of Science. *Scientometrics*, 64(3), 351-374.
- Bornmann, L., Leydesdorff, L., & Marx, W. (2007). Citation Environment of *Angewandte Chemie*. *CHIMIA*, 61(3), 104-109.
- Braam, R. (2007). Everything about Genes: Some Results on the Dynamics of Genomics Research. Proceedings of the 11th International Conference of Scientometrics and

- Informetrics, D. Torres-Salinas & H. Moed (Eds.), CSIC, Madrid, 21-25 June 2007, Vol. 1, 136-148.
- Bradford, S. C. (1934). Sources of information on specific subjects. *Engineering*, 137, 85-86.
- Braun, T., Glänzel, W., & Schubert, A. (2006). A Hirsch-type index for journals. *Scientometrics*, 69(1), 169-173.
- Bruckner, E., Ebeling, W., Montaña, M. A. J., & Scharnhorst, A. (1994). Hyperselection and Innovation Described by a Stochastic Model of Technological Evolution. In L. Leydesdorff & P. Van den Besselaar (Eds.), *Evolutionary Economics and Chaos Theory: New Directions in Technology Studies* (pp. 79-90). London: Pinter.
- Burrell, Q. L. (2007). Hirsch index or Hirsch rate? Some thoughts arising from Liang's data. *Scientometrics*, 73(1), 19-28.
- Christensen, F. H., Ingwersen, P., & Wormell, I. (1997). Online determination of the journal impact factor and its international properties. *Scientometrics*, 40(3), 529-540.
- Collins, H. M. (1985). The Possibilities of Science Policy. *Social Studies of Science*, 15, 554-558.
- Egghe, L. (2006a). How to improve the h-index. *The Scientist*, 20(3), 315-321.
- Egghe, L. (2006b). An improvement of the H-index: the G-index. *ISSI Newsletter*, 2(1), 8-9.
- Egghe, L., & Rousseau, R. (1990). *Introduction to Informetrics*. Amsterdam: Elsevier.
- Frandsen, T. F., & Rousseau, R. (2005). Article impact calculated over arbitrary periods. *Journal of the American Society for Information Science and Technology*, 56(1), 58-62.
- Freeman, L. C. (1977). A Set of Measures of Centrality Based on Betweenness. *Sociometry*, 40(1), 35-41.
- Fujigaki, Y. (1998). Filling the Gap Between Discussions on Science and Scientists' Everyday Activities: Applying the Autopoiesis System Theory to Scientific Knowledge. *Social Science Information*, 37(1), 5-22.
- Garfield, E. (1971). The mystery of the transposed journal lists—wherein Bradford's Law of Scattering is generalized according to Garfield's Law of Concentration. *Current Contents*, 3(33), 5-6.
- Garfield, E. (1972). Citation Analysis as a Tool in Journal Evaluation. *Science* 178(Number 4060), 471-479.
- Garfield, E. (1979). *Citation Indexing: Its Theory and Application in Science, Technology, and Humanities*. New York: John Wiley.
- Garfield, E. (Ed.) (1980). *SCI journal citation reports: A bibliometric analysis of science journals in the ISI data base. Science Citation Index 1979 annual*, v. 14. Philadelphia: Institute for Scientific Information.
- Garfield, E. (1990). How ISI selects Journals for Coverage: Quantitative and Qualitative Considerations. *Current Contents* (28 May), 5-13.
- Garfield, E. (1996). How can impact factors be improved? *British Medical Journal*, 313(7054), 411-413.
- Garfield, E. (1998a). Long-term vs. short-term journal impact: does it matter? *The Scientist*, 12(3), 11-12.

- Garfield, E. (1998b). The Multiple Meanings of Impact Factors. *Journal of the American Society for Information Science*, 49(8), 768.
- Garfield, E. (2003). The Meaning of the Impact Factor. *Revista Internacional de Psicologia Clinica y de la Salud* 3(2), 363-369.
- Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). *The new production of knowledge: the dynamics of science and research in contemporary societies*. London: Sage.
- Glänzel, W., & Schubert, A. (2003). A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. *Scientometrics*, 56(3), 357-367.
- Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360-1380.
- Hanneman, R. A., & Riddle, M. (2005). *Introduction to social network methods*. Riverside, CA: University of California, Riverside; available at <http://faculty.ucr.edu/~hanneman/>, last accessed on 19 February 2007.
- Hellsten, I., Lambiotte, R., Ausloss, M. & Scharnhorst, A. (forthcoming). Self-Citations, Co-Authorships and Keywords: A New Method for Detecting Scientists' Field Mobility? *Scientometrics*, 72(3) (2007; forthcoming).
- Hicks, D. (1999). The difficulty of achieving full coverage of international social science literature and the bibliometric consequences. *Scientometrics*, 44(2), 193-215.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the USA*, 102(46), 16569-16572.
- Jin, B. H. (2006). H-index: an evaluation indicator proposed by scientist. *Science Focus*, 1(1), 8-9.
- Jin, B., Liang, L., Rousseau, R., & Egghe, L. (2007). The R- and AR-indices: complementing the h-index. *Chinese Science Bulletin*, 52(6), 855-863.
- King, D. A. (2004). The scientific impact of nations. *Nature*, 430(15 July 2004), 311-316.
- Kosmulski, M. (2006). A new Hirsch-type index saves time and works equally well as the original h-index. *ISSI Newsletter*, 2(3), 4-6.
- Kreft, G. G., & De Leeuw, E. D. (1988). The See-Saw Effect: A multilevel problem. *Quality and Quantity*, 22(2), 127-137.
- Leydesdorff, L. (1998). Theories of Citation? *Scientometrics*, 43(1), 5-25.
- Leydesdorff, L. (2006a). Can Scientific Journals be Classified in Terms of Aggregated Journal-Journal Citation Relations using the Journal Citation Reports? *Journal of the American Society for Information Science & Technology*, 57(5), 601-613.
- Leydesdorff, L. (2006b). *The Knowledge-Based Economy: Modeled, Measured, Simulated*. Boca Rota, FL: Universal Publishers.
- Leydesdorff, L. (2007a). Visualization of the Citation Impact Environments of Scientific Journals: An online mapping exercise. *Journal of the American Society of Information Science and Technology*, 58(1), 207-222.
- Leydesdorff, L. (2007b). Mapping Interdisciplinarity at the Interfaces between the Science Citation Index and the Social Science Citation Index. *Scientometrics*, 71(3), 391-405.

- Leydesdorff, L. (2007c). "Betweenness Centrality" as an Indicator of the "Interdisciplinarity" of Scientific Journals. *Journal of the American Society for Information Science and Technology*, 58(9), 1303-1309.
- Leydesdorff, L., & Van der Schaar, P. (1987). The Use of Scientometric Indicators for Evaluating National Research Programmes. *Science & Technology Studies*, 5, 22-31.
- Leydesdorff, L., & Amsterdamska, O. (1990). Dimensions of Citation Analysis. *Science, Technology & Human Values*, 15, 305-335.
- Leydesdorff, L., & Van den Besselaar, P. (1997). Scientometrics and Communication Theory: Towards Theoretically Informed Indicators. *Scientometrics*, 38, 155-174.
- Leydesdorff, L., & Jin, B. (2005). Mapping the Chinese Science Citation Database in terms of aggregated journal-journal citation relations. *Journal of the American Society for Information Science and Technology*, 56(14), 1469-1479.
- Leydesdorff, L., & Wagner, C. S. (forthcoming). Is the United States losing ground in science? A global perspective on the world science system *Scientometrics*, In press.
- Leydesdorff, L., & Zhou, P. (2005). Are the contributions of China and Korea upsetting the world system of science? *Scientometrics*, 63(3), 617-630.
- Li, Z. X. (2006). How to establish a first-class international scientific journal in China? *World Journal of Gastroenterology*, 12(43), 6905-6908.
- Maricic, S. (1997). The mainstream-peripheral science communication. *Technoscience*, 10 (Winter 1997), available at <http://cybermetrics.cindoc.csic.es/cybermetrics/pdf/394.pdf>, last accessed 19 February 2007.
- Martin, B., & Irvine, J. (1983). Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio Astronomy. *Research Policy*, 12, 61-90.
- Martyn, J., & Gilchrist, A. (1968). An Evaluation of British Scientific Periodicals. Aslib Occasional Publication No. 1. London: Aslib.
- Marx, W. (2001). Angewandte Chemie in Light of the Science Citation Index. *Angewandte Chemie International Edition* 40(1), 139-143.
- Meho, L. I., & Yang, K. (forthcoming). A New Era in Citation and Bibliometric Analyses: Web of Science, Scopus, and Google Scholar. *Journal of the American Society for Information Science and Technology*.
- Moed, H. F. (2005). *Citation Analysis in Research Evaluation*. Dordrecht: Springer.
- Moed, H. F., Burger, W. J. M., Frankfort, J. G., & Raan, A. F. J. V. (1985). The Use of Bibliometric Data for the Measurement of University Research Performance. *Research Policy* 14, 131-149.
- Moravcsik, M. J. (1985). Applied scientometrics: An assessment methodology for developing countries. *Scientometrics*, 7(3), 165-176.
- National Research Council (2007). *Research Doctorate Programs*. The National Academies, Washington, DC; available at <http://www7.nationalacademies.org/resdoc/index.html>, last accessed 10 Feb. 2007.
- Nederhof, A. J., Zwaan, R. A., Bruin, R. E., & Dekker, P. J. (1989). Assessing the usefulness of bibliometric indicators for the humanities and the social sciences: A comparative study. *Scientometrics*, 15, 423-436.

- Nowotny, H., Scott, P., & Gibbons, M. (2001). *Re-Thinking Science: Knowledge and the Public in an Age of Uncertainty*. Cambridge, etc: Polity.
- Park, H. W., Hong, H. D., & Leydesdorff, L. (2005). A Comparison of the Knowledge-based Innovation Systems in the Economies of South Korea and the Netherlands using Triple Helix Indicators. *Scientometrics*, 65(1), 3-27.
- Park, H. W., & Leydesdorff, L. (forthcoming). Korean journals in the Science Citation Index: What do they reveal about the intellectual structure of S&T in Korea? *Scientometrics* (forthcoming).
- Pinski, G., & Narin, F. (1976). Citation Influence for Journal Aggregates of Scientific Publications: Theory, with Application to the Literature of Physics. *Information Processing and Management*, 12(5), 297-312.
- Price, D. J. de Solla (1965). Networks of scientific papers. *Science*, 149, 510- 515.
- Price, D. J. de Solla (1970). Citation Measures of Hard Science, Soft Science, Technology, and Nonscience. In C. E. Nelson & D. K. Pollock (Eds.), *Communication among Scientists and Engineers* (pp. 3-22). Lexington, MA: Heath.
- Rafols, I., & Meyer, M. (2007). How cross-disciplinary is bionanotechnology? Explorations in the specialty of molecular motors. *Scientometrics*, 70(3), 633-650.
- Rousseau, R. (2004). Impact factors and databases as instruments for research evaluation. Paper presented at the 4th International Conference on University Evaluation and Research Evaluation, Wuhan University (China), September 27-October 1, 2004, Available at http://www.vub.ac.be/BIBLIO/itp/lecturers/ronald_rousseau/ronald_rousseau_stim_4_impact_factors_as_instruments.pdf, last accessed 19 March 2007.
- Rousseau, R. (2006a). Simple models and the corresponding h- and g-index. *Science Technology Development*, 2006 (forthcoming).
- Rousseau, R. (2006b). Timelines in Citation Research. *Journal of the American Society for Information Science and Technology*, 57(10), 1404-1405.
- Salton, G., & McGill, M. J. (1983). *Introduction to Modern Information Retrieval*. Auckland, etc.: McGraw-Hill.
- Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal*, 314, 498-502.
- Sher, I. H., & Garfield, E. (1965). New tools for improving and evaluating the effectiveness of research. Second conference on Research Program Effectiveness, Washinton, DC July, 27-29.
- Stegmann, J. (1997). How to evaluate journal impact factors. *Nature*, 390, 550.
- Stegmann, J. (1999). Building a list of journals with constructed impact factors. *Journal of Documentation*, 55(3), 310-324.
- Studer, K. E., & Chubin, D. E. (1980). *The Cancer Mission. Social Contexts of Biomedical Research*. Beverly Hills, etc.: Sage.
- Testa, J. (1997). *The ISI database: The journal selection process* (Philadelphia: Thomson- ISI). Available at <http://scientific.thomson.com/free/essays/selectionofmaterial/journalselection/>, last accessed 19 March 2007.

- Tijssen, R. J. W., & Van Leeuwen, T. N. (1995). On generalising scientometric journal mapping beyond ISI's journal and citation databases. *Scientometrics*, 33(1), 93-116.
- Van den Besselaar, P. (2001). The cognitive and the social structure of Science & technology studies. *Scientometrics*, 51, 441-460.
- Von Tunzelmann, N., Ranga, M., Martin, B., & Geuna, A. (2003). The Effects of Size on Research Performance: A SPRU Review. Available at <http://www.sussex.ac.uk/spru/documents/geunaost.pdf> , last accessed 27 March 2007.
- Whitley, R. D. (1984). *The Intellectual and Social Organization of the Sciences*. Oxford: Oxford University Press.
- Zhou, P., & Leydesdorff, L. (2007a). A Comparison between the China Scientific and Technical Papers and Citations Database and the Science Citation Index in terms of journal hierarchies and inter-journal citation relations. *Journal of the American Society for Information Science and Technology* 58(2), 223-236.
- Zhou, P., & Leydesdorff, L. (2007b). The Citation Impacts and Citation Environments of Chinese Journals in Mathematics. *Scientometrics* 72(2), 185-200.