



Review

# Informetrics at the beginning of the 21st century—A review

Judit Bar-Ilan\*

*Department of Information Science, Bar-Ilan University, Ramat Gan 52900, Israel*

Received 1 August 2007; received in revised form 18 November 2007; accepted 22 November 2007

## Abstract

This paper reviews developments in informetrics between 2000 and 2006. At the beginning of the 21st century we witness considerable growth in webometrics, mapping and visualization and open access. A new topic is comparison between citation databases, as a result of the introduction of two new citation databases Scopus and Google Scholar. There is renewed interest in indicators as a result of the introduction of the *h*-index. Traditional topics like citation analysis and informetric theory also continue to develop. The impact factor debate, especially outside the informetric literature continues to thrive. Ranked lists (of journal, highly cited papers or of educational institutions) are of great public interest.

© 2007 Elsevier Ltd. All rights reserved.

*Keywords:* Informetrics; Bibliometrics; Scientometrics; Webometrics

## Contents

1. Introduction . . . . .	2
2. Data collection . . . . .	3
2.1. General characteristics of the reviewed items . . . . .	5
3. General works on informetrics . . . . .	5
3.1. Structure of the field and relations to other fields . . . . .	6
3.2. History of bibliometrics . . . . .	7
3.3. Acknowledgements . . . . .	7
4. Theory . . . . .	7
4.1. Informetric laws, distributions and concentration . . . . .	7
4.2. Other theory related to time and growth . . . . .	8
5. Methods and techniques . . . . .	9
5.1. Mapping and visualization . . . . .	9
5.2. Text and data mining and other linguistic techniques . . . . .	10
5.3. Network analysis . . . . .	11
5.4. Classification . . . . .	12
5.5. Other methods . . . . .	12
6. Citation analysis . . . . .	12
6.1. Characterizing publications and fields based on citation analysis . . . . .	13
6.2. Cocitation analysis—theoretical and methodological aspects . . . . .	14

\* Tel.: +972 3 5318351; fax: +972 3 7384027.

*E-mail address:* [barilaj@mail.biu.ac.il](mailto:barilaj@mail.biu.ac.il).

6.3.	Self-citation	14
6.4.	Ego-centered citation studies	15
6.5.	Quality assessment	15
6.6.	Errors in citation analysis and statistics	15
7.	Indicators	16
7.1.	<i>h</i> -Index	16
7.2.	Impact factor	16
7.3.	Other bibliometric indicators	17
8.	Webometrics	19
8.1.	Web Impact Factors, other measures and methodological issues	19
8.2.	The ‘academic’ web	21
8.3.	Commercial sites	21
8.4.	Linking motivations and content analysis	22
8.5.	Journals and web citations	22
9.	Journals	22
9.1.	The ‘impact factor’ debate	23
9.2.	Journal quality	23
9.3.	Journal coverage and structure	24
9.4.	Journal profiles	24
9.5.	Non-English journals and internationalization	25
9.6.	Journal editors	26
10.	Open access and electronic publications	26
10.1.	Open access	26
10.2.	Electronic publications	27
11.	Productivity and publications	27
11.1.	Country level studies	27
11.2.	Field specific studies	28
11.3.	Media coverage of science	28
11.4.	Interdisciplinarity	29
11.5.	Other studies	29
12.	Collaborations	29
13.	Research policy	30
13.1.	Research evaluation	31
13.2.	University rankings	32
14.	Patent analysis	32
14.1.	Patent citations and science and technology links	32
14.2.	Other patent indicators	33
15.	Databases	33
16.	Summary	34
	Acknowledgements	35
	References	35

## 1. Introduction

This paper reviews recent developments in informetrics. [Wilson’s \(1999\)](#) excellent review article in ARIST covers the topic until the end of the 20th century and [Thelwall, Vaughan, and Björneborn \(2005\)](#) provide an even more recent review of the sub-topic, webometrics.

We adopt [Tague-Sutcliffe’s \(1992, p. 1\)](#) definition: “Informetrics is the study of the quantitative aspects of information in any form, not just records or bibliographies, and in any social group, not just scientists”. [Egghe \(2005a, p. 1311\)](#) uses the informetrics “as the broad term comprising all the—metrics studies related to information science, including bibliometrics (bibliographies, libraries, . . .), scientometrics (science policy, citation analysis, research evaluation, . . .), webometrics (metrics of the web, the Internet or other social networks such as citation or collaboration networks), . . .”.

Hood and Wilson (2001a) provide an extensive discussion of the terms informetrics, bibliometrics and scientometrics. Webometrics, one of the newest metrics is discussed by Björneborn and Ingwersen (2004). They consider webometrics (studies of the Web) as a subfield of cybermetrics (studies of all Internet applications), however often these two terms are used as synonyms. In this paper we will use the more widely used terminology—webometrics, in the general sense of cybermetrics.

A literature review of a field with such large number of indexed publications (see Section 2 for details) has to be selective and is inevitably subjective. In Wilson's (1999) literature review the two largest sections were dedicated to citation analysis and to informetric laws. In the current review we will also cover these areas, while emphasizing new developments like the emergence of webometrics (which was mentioned under “new trends in informetrics” by Wilson), new visualization and mapping techniques, multiple citation databases and the *h*-index.

## 2. Data collection

How does one start preparing a review article on such an extensive topic? A search carried out on the Web of Science, Scopus, Google Scholar, LISA and LISTA for “informetric\* OR bibliometric\* OR scientometric\* OR webometric\*” with time-span limited to 2000–2006, resulted in 1021, 1734, 1283, 1462, and 8350 items, respectively.<sup>1</sup> In addition, a search with so few keywords does not cover all the relevant literature. Thus we decided to construct a more specific query. This query was submitted to ISI (Thomson)'s Web of Science (WOS) and to Elsevier's Scopus. Google Scholar, LISA (Library and Information Science Abstracts) and LISTA (Library Information Science & Technology Abstracts) were also consulted, but WOS and Scopus were the main data sources. Data were collected in April 2007 using multiple methods:

### 1. Submitting a complex query:

Informetric\* OR bibliometric\* OR webometric\* OR “citation analysis” OR “citation analyses” OR “cocitation analysis” OR “cocitation analyses” OR “co-citation analysis” OR “co-citation analyses” OR “link analysis” OR “link analyses” OR “link structure” OR “self citation” OR “self citations” OR self-citation OR self-citations OR “S&T indicator” OR “S&T indicators” OR citation map\* OR citation visual\* OR “science policy” OR “research policy” OR “impact factor” OR “impact factors” OR *h*-index OR “h index” OR Hirsch index OR “patent analysis” OR “patent analyses” OR Zipf OR Bradford OR Lotka OR “collaboration network” OR “collaboration networks” OR “coauthorship network” OR “coauthorship networks” OR “co-authorship network” OR “co-authorship networks”.

2. Searching for articles published by authors who attended the 10<sup>th</sup> ISSI Conference (Stockholm, 2005) and the 9<sup>th</sup> International Conference on S&T Indicators (Leuven, 2006)—based on the lists of participants in these conferences.
3. Relevant books published during the period—searching library catalogues and based on personal knowledge.
4. Systematic coverage of the journals *Scientometrics* and *Cybermetrics*.
5. Additional items that were referenced in the reviewed documents.

Next the initial data set was cleansed based on the abstracts of the retrieved items. For example, one of the most cited items (cited 316 times) was “Chemokines: immunology's high impact factors”, however it had nothing to do with impact factors in the informetric sense. Next all non-English items were removed. This process still left us with too many items: thus it was necessary to further reduce the number of selected items. An effort was made to cover a representative sample of the informetric papers published during the period. However, it must be noted that any review article attempting to make a selection from a set of this size will be inevitably subjective.

Each selected item was submitted to Google Scholar in order to retrieve the citation counts reported by it. We found Google Scholar difficult to use for initial data collection, but it indexed all except a single item on our list, thus we collected the citation counts reported by Google Scholar as well. The next section describes the basic informetric characteristics of the 598 reviewed publications. The selected items were categorized according to the following major categories: general, theory, methods and techniques, citation analysis, indicators, webometrics, journals, open access & electronic publications, productivity and publications, collaboration, research policy, patents, and databases.

<sup>1</sup> Searches were carried out on April 15, 2007.

Table 1  
The most highly cited reviewed publications

Author(s)	Title	Publication year	Source	Times cited Scopus	Times cited WOS	Times cited Scholar
Newman, MEJ	The structure of scientific collaboration networks	2001	PNAS	318	302	570
Girvan, M.; Newman MEJ	Community structure in social and biological networks	2002	PNAS	225	201	461
Newman, MEJ	Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality	2001	Physical Review E	205	156	240
Etzkowitz, H; Leydesdorff, L	The dynamics of innovation—From National Systems and “Mode 2” to a Triple Helix of university-industry-government relations	2000	Research Policy	171	114	413
Barabási, AL; Jeong, H; Neda, Z; Ravasz, E; Schubert, A; Vicsek, T	Evolution of the social network of scientific collaborations	2002	Physica A	134	122	260
Barrat, A; Barthelemy, M; Pastor-Satorras, R; Vespignani, A	The architecture of complex weighted networks	2004	PNAS	120	119	183
Mowery, DC; Nelson, RR; Sampat, BN; Ziedonis, AA	The growth of patenting and licensing by US universities—An assessment of the effects of the Bayh-Dole act of 1980	2001	Research Policy	107	95	261
King, DA	The scientific impact of nations	2004	Nature	84	71	136
Lawrence, S.	Free online availability substantially increases a paper's impact	2001	Nature	82	82	194
Adam, D.	The counting house	2002	Nature	74	64	129
Borgman, CL; Furner, J	Scholarly communication and bibliometrics	2002	ARIST	70	55	160
McMillan, GS; Narin, F; Deeds, DL	An analysis of the critical role of public science in innovation—the case of biotechnology	2000	Research Policy	67	54	132
Cronin, B	Bibliometrics and beyond—Some thoughts on web-based citation analysis	2001	J. Inf. Sci.	65	61	97
Thelwall, M	Extracting macroscopic information from Web links	2001	JASIST	65	63	103
Björneborn, L; Ingwersen, P	Perspectives of webometrics	2001	Scientometrics	63	58	119
Börner, K; Chen, CM; Boyack, KW	Visualizing knowledge domains	2003	ARIST	62	30	106
Lawrence, PA	The politics of publication—Authors, reviewers and editors must act to protect the quality of research	2003	Nature	61	56	88
Glänzel, W; Moed, HF	Journal impact measures in bibliometric research	2002	Scientometrics	50	38	6
Lee, KP; Schotland, M; Bacchetti, P; Bero, LA	Association of journal quality indicators with methodological quality of clinical research articles	2002	JAMA	49	41	75
Thelwall, M	Conceptualizing documentation on the Web—An evaluation of different heuristic-based models for counting links between university Web sites	2002	JASIST	49	38	70
Ioannidis, JPA	Contradicted and initially stronger effects in highly cited clinical research	2005	JAMA	49	47	52
Meyer, M	Does science push technology? Patents citing scientific literature	2000	Research Policy	49	43	90
Saha, S; Saint, S; Christakis, DA	Impact factor—a valid measure of journal quality?	2003	JMLA	49	44	56
Callaham, M; Wears, RL; Weber, E	Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals	2002	JAMA	47	42	5
Newman, MEJ	Clustering and preferential attachment in growing networks	2001	Physical Review E	47	39	108
Patel, V., Sumathipala, A.	International representation in psychiatric literature—Survey of six leading journals	2001	British Journal of Psychiatry	46	33	58
Wilkinson, D; Harries, G; Thelwall, M; Price, L	Motivations for academic web site interlinking—evidence for the Web as a novel source of information on informal scholarly communication	2003	JInfSci	46	15	67
Bar-Ilan, J	Data collection methods on the Web for informetric purposes—A review and analysis	2001	Scientometrics	43	43	73
Vaughan, L; Thelwall, M	Scholarly use of the Web—What are the key inducers of links to journal Web sites?	2003	JASIST	43	41	65
Fassoulaki, A; Paraskeva, A; Papilas, K; Karabinis, G	Self-citations in six anaesthesia journals and their significance in determining the impact factor	2000	British Journal of Anaesthesia	43	39	44
Smith, A; Thelwall, M	Web impact factors for Australasian universities	2002	Scientometrics	43	38	60
Egghe, L	New informetric aspects of the Internet – some reflections – many problems	2000	J.Inf. Sci.	42	40	69
Moed, HF	The impact-factors debate—The ISI's uses and limits	2002	Nature	42	37	55
Landry, R; Amara, N; Lamari, M	Utilization of social science research knowledge in Canada	2001	Research Policy	42	32	63
Ho, YS	Citation review of Lagergren kinetic rate equation on adsorption reactions	2004	Scientometrics	42	40	37

### 2.1. General characteristics of the reviewed items

In Table 1 the 35 most highly cited reviewed items are displayed. For each item citation counts reported by WOS, Scopus and Google Scholar are displayed. The publications are arranged in decreasing order of the Scopus citation count, which is similar but slightly higher than the WOS citation count. The citation counts reported by Google Scholar are usually much higher, but there are a few exceptions. Note that some of the highly cited items are not informetric papers *per se*, but discuss implications of informetric indicators (e.g., citation counts and impact factors). A number of highly cited papers are on collaboration networks, where these networks often serve as test cases of complex and/or social networks and these papers are cited by other communities as well.

Table 2 displays the list of the most frequently occurring journals in the list. Except for the top two journals (Scientometrics and JASIST), this list is considerably different from the list of major journals in Wilson's review (1999) of informetrics for the period 1990–1999. There can be a number of possible reasons for the differences: (1) Wilson considered the total list of publications retrieved for “bibliometric? OR informetric?” from a combination of databases. (2) These databases seemingly had considerable coverage of non-English journals. In the current case non-English items were removed from the set of selected items. (3) Wilson used only two general search terms and excluded scientometrics, while our list is based on a complex query and additional data collection techniques. The last column of Table 2 contains the rank of the journal in Wilson's list.

Finally, Table 3 displays the top list of authors, i.e., the authors with the five or more publications in the list. Altogether 683 authors were identified. Tables 1 and 3 both show the emergence of webometrics in the reviewed period.

### 3. General works on informetrics

Hood and Wilson's (2001a) explain the usage and the history of the terms scientometrics, bibliometrics and informetrics. Their paper is also the first paper in the recently published first volume of the Scientometrics Guidebook Series, edited by Braun (2006). The aim of the guidebook series is to provide a theoretical and practical guide to the quantitative studies of science. The book is a collection of previously published articles. Another book with somewhat similar aims was edited by Moed, Glänzel, and Schmoch (2004). In this case the chapters of the book are original

Table 2  
The most frequently occurring journals

Journal name	No. of docs	Rank	Rank in Wilson (1999)
Scientometrics	183	1	1
Journal of the American Society for Information Science and Technology (incl. JASIS)	93	2	2
Research Policy	26	3	12
Information Processing & Management	22	4	4
Journal of Information Science	19	5	8–9
Journal of Documentation	17	6	13
Cybermetrics	10	7–9	n/a
Research Evaluation	10	7–9	n/a
Nature	10	7–9	n/a
Proceedings of the National Academy of Sciences of the United States of America	9	10	n/a
Online Information Review	7	11	n/a
College & Research Libraries	6	12–13	n/a
JAMA: Journal of the American Medical Association	6	12–13	n/a
Journal of the Medical Library Association	5	14	8–9
Annual Review of Information Science and Technology	4	15–19	n/a
ASLIB Proceedings	4	15–19	n/a
Library Trends	4	15–19	n/a
ASIST Monograph Series	4	15–19	n/a
Physical Review E	4	15–19	n/a
Current Science	3	20–22	n/a
D-Lib Magazine	3	20–22	n/a
Physica A	3	20–22	n/a

Table 3  
Top authors

Rank	Author	No. of publications in list	Rank	Author	No. of publications in list
1	Thelwall, M	45	18–20	Wilkinson, D	7
2	Rousseau, R	25	21–29	Aksnes, DW	6
3	Egghe, L	24	21–29	Bassecoulard, E	6
4–5	Glänzel, W	22	21–29	Garfield, E	6
4–5	Leydesdorff, L	22	21–29	Godin, B	6
6–7	van Raan, AFJ	18	21–29	Harries, G	6
6–7	Vaughan, L	18	21–29	Kostoff, RN	6
8	Bar-Ilan, J	17	21–29	Persson, O	6
9	Moed, HF	14	21–29	Shaw, D	6
10	Schubert, A	12	21–29	Zitt, M	6
11	van Leeuwen, TN	11	30–37	Björneborn, L	5
12–14	Braun, T	10	30–37	Börner, K	5
12–14	Cronin, B	10	30–37	Boyack, KW	5
14–14	Meyer, MS	10	30–37	Burrell, QL	5
15	Chen, CM	9	30–37	Carr, L	5
16–17	Oppenheim, C	8	30–37	Lewison, G	5
16–17	White, HD	8	30–37	Newman, MEJ	5
18–20	Harnad, S	7	30–37	Visser, MS	5
18–20	Tijssen, RJW	7			

contributions by a long list of well-known informetricians. It concentrates mainly on indicators derived from scholarly publications and patent literature. During this period a new edition of *Leydesdorff's Challenge of Scientometrics (2001a)* was published as well.

In addition the Wilson's ARIST chapter on informetrics in 1999, and Thelwall, Vaughan and Björneborn's chapter on webometrics in 2004, *Borgman and Furner (2002)* published a chapter on scholarly communication and bibliometrics in 2002. They highlight the emergence of electronic scholarly communication, and the chapter focuses on how bibliometric methods can reflect on scholarly communicative behavior. Specifically, they examine scholarly communicative behavior in the contexts of linking, writing, submitting and collaborating, with an emphasis on linking that includes both citing in the classical sense and linking between hypertext documents. *van Raan (2001a)* also discusses the implications of the Internet of scholarly communication. In his view, the major implication is increased access to scholarly publications and to research data. Increased and early access are major issues for open access as well—discussed in Section 11.

### 3.1. Structure of the field and relations to other fields

The social and cognitive structure of Science and Technology Studies (STS)—including qualitative, quantitative and policy aspects, was studied by *van den Besselaar (2001)*. In a previous study by *van den Besselaar (2000)*, he has already found that there are very few cognitive links between these subfields based on journal-journal citation analysis. The author cocitation analysis revealed a similar split in the social dimension as well, showing the isolation of the different subfields. *Schoepflin and Glänzel (2001)* classified the articles published in *Scientometrics* in the years 1980, 1989 and 1997. The results of the classification show a decrease in the percentages of both the articles related to science policy and to the sociology of science over the years, as pointed out in (*Schubert, 2002*) as well. On the other hand, *Peritz and Bar-Ilan (2002)* found, based on the analysis of papers published in the journal *Scientometrics* in 1990 and in 2000, that Research Policy (the major publication venue for policy studies) and Social Studies of Science (one of the two major journals publishing qualitative studies) are the third and fourth most frequently referenced journals in articles published in the journal *Scientometrics*.

Based on factor analysis, *van den Besselaar (2001)* defined the following subfields of the quantitative studies of science: policy oriented scientometrics, empirical studies, cword studies, scientometric distributions, critique of scientometrics, patent studies and economics of technical change—based on data for the period 1986–1997 (which explains the absence of webometrics as a subfield). The classification used by *Schoepflin and Glänzel (2001)* included the following categories: bibliometric theory, case studies and empirical papers, methodological papers including appli-

cations, indicator engineering and data presentation, sociological approach to bibliometrics and research policy. In a later paper (Glenisson, Glänzel, Janssens, & De Moor, 2005) the categories were revised to reflect recent developments: advances in scientometrics, empirical papers/case studies, mathematical models, political issues, sociological approaches and informetrics/webometrics (their definition of “informetrics” is obviously different from the definition used in this review).

Janssens, Leta, Glänzel, and De Moor (2006) applied cword analysis to map the structure of library and information science based on the full text of articles published during period 2000–2004 in five LIS journals: Information Processing and Management, JASIST, Journal of Documentation, Journal of Information Science and Scientometrics. They identified six clusters: two bibliometrics clusters and one cluster each on patents, information retrieval, webometrics and social aspects of information science. Leydesdorff (2002a) studied the cited and citing position of JASIS and later JASIST at several points in time and was able to classify the major citing/cited journals as journals in information science, library science or science studies and new technologies.

### 3.2. History of bibliometrics

Paul Wouters (1999) PhD thesis, the Citation Culture is the only item in this review that was published before 2000, because even though it was accepted as a PhD thesis in 1999, it became available online only in the 21st century. Eugene Garfield (no date) describes the book as: “This is undoubtedly one of the most significant works to be produced in this field since its advent over forty years ago.”

Godin (2005) published a book on the development of S&T statistics. The historical events are followed from 1920 to the present, emphasizing governmental statistics, and measures and manuals developed by and for organizations like UNESCO, the OECD and the NSF. In the second part of the book the use of these statistics are discussed.

In a recent paper, Godin (2006a) placed the origins of bibliometrics with Cattell and other psychologists at the beginning of the 20th century; while Bensman (2005) emphasized the role of Donald Urquhart in the development of bibliometrics. In another paper, Bensman (2001) discussed the controversy in Britain about the validity of Urquhart’s and Garfield’s laws. Line’s contributions to bibliometrics are described in (Meadows, 2005). Rousseau (2005a) considered the influence of Robert Faithorne’s article on empirical hyperbolic distributions on the development of informetrics. Schubert (2002) analyzed the shorter term “history” of the journal, Scientometrics; while Small (2003) provides his personal account on the development of maps of science.

### 3.3. Acknowledgements

Cronin, Shaw, and Le Barre (2003) also mention the psychologist Cattell, but in a different context—as one of the founders of the Psychological Review. They manually examined every issue of 100 volumes of the journals Psychological Review and Mind to identify the number of authors and the number of acknowledgements for each article. The acknowledgements were classified as: conceptual, editorial, financial, instrumental/technical, moral and reader. The findings show that acknowledgements have become established elements of academic writing, in some cases expressing “subauthorship”. A similar study was carried out by the same authors (Cronin, Shaw, & Le Barre, 2004) in chemistry. Cronin (2001a) also discusses the relation between the list of authors and the acknowledgements in scholarly publications, especially in the cases when the number of authors is huge (“hyperauthorship”).

While Cronin et al. (2003, 2004) extracted the acknowledgements form the articles manually, Giles and Council (2004) developed automatic methods for acknowledgment extraction and analysis and applied their method to 335,000 computer science papers. They tabulated the most frequently acknowledged funding agencies, companies, institutions and individuals in the above-mentioned set of papers.

## 4. Theory

### 4.1. Informetric laws, distributions and concentration

Wilson (1999) dedicated a large part of her review to informetric laws. Our review of this area will be much less extensive. In this section, as in the previous one, we start the review with the relevant books (or book in this case) on the topic. Egghe’s (2005b) book on Lotkaian informetrics studies information production processes, where sources produce

items (in some parts of the book this two dimensional case is extended). The production quantities of the sources are described by a size-frequency function,  $f$ , where  $f(n)$  is the number of sources with  $n$  items. When  $f$  is function of the form  $f(n) = C/n^\alpha$ , where  $\alpha \geq 0$ , the function  $f$  is said to be Lotkaian. Other informetric laws and phenomena can be described in terms of appropriate Lotkaian functions, resulting in a unified theory based on Lotkaian informetrics. The book incorporates results from several previously published articles by Egghe (e.g., Egghe, 2000a, 2000b, 2002, 2003, 2004a, 2004b, 2005c, 2005d; Egghe & Rao, 2001, 2002; Egghe & Rousseau, 2002a). Burrell (2005a) suggests an alternative formulation to what Egghe (2003) calls type/token-taken informetrics (i.e., where not only sources and items are considered, but also the use that is made of these items).

Tsallis and de Albuquerque (2000) suggest a more complex power law type function to fit the curve  $N(x)$ —the number of papers with  $x$  citations. A similar idea is used by Montemurro (2001) to provide a better fit to the frequency-rank distribution of words in texts.

In contrast to the power-law like distributions suggested by Lotkaian informetrics, van Raan (2001b) showed that the distribution of citations over publications does not follow a power law, but is represented by a modified Bessel function.

Lotka's law was reexamined by several authors in the period under review; Huber (2002) suggested a new model that generates Lotka's law and generates good fits to several samples. The parameters of the model are based on a few, plausible assumptions. Goldstein, Morris, and Yen (2004) discuss problems with fitting to a power law distribution using graphical methods. They suggest that the maximum likelihood estimate is much better than the linear fit on a log-log scale. They demonstrate this for Lotka's law for a set of 900 papers on complex systems. The Goldstein et al. (2004) paper has already been cited extensively (24, 28 and 57 citations according to WOS, Scopus and GS, respectively), even though their findings about the usage of the maximum likelihood estimator is already known for quite some time (see, for example Nicholls, 1989). Rousseau and Rousseau (2000) even published a publicly available computer program to calculate the best parameters. Goldstein et al. (2004) claim, that Nicholls did not correct for the dependency of the parameters on the hypothesized distribution function. Bailón-Moreno, Jurado-Alameda, Ruiz-Baños, and Courtial (2005) claim that power laws are not adequate to describe informetric phenomena, since to achieve reasonable fit exponential terms need to be included as well.

Rousseau (2002a) discusses the lack of standardization in informetric research, and as an example he shows again the superiority of the maximum likelihood estimator over the least square method for estimating the parameters of Lotka's law. Vinkler (2001) also calls for standardization of scientometric research in general.

Lorenz curves and concentration measures are used in (Egghe & Rousseau, 2002b, 2006a) for identifying cores and for measuring the similarity between two items for information retrieval. For IR (information retrieval) applications, the appropriate theory is symmetric relative concentration theory (see Egghe & Rousseau, 2001). Burrell (2005b) introduces a concentration measure that extends the Gini index and allows measuring the inequality of productivity between two population of sources. Yoshikane, Kageura, and Tsuji (2003) propose to take into account not only the relative concentration measures (like the Gini index), but also the total number of authors in the specific field.

Lafouge and Michel (2001) discuss the connections between bibliometric distributions and entropy—extending previous results based on the least effort principle.

#### 4.2. Other theory related to time and growth

Börner, Maru, and Goldstone (2004) propose a novel simulation where both author and paper networks evolve simultaneously. Even though the assumptions for the simulation seem to be rather unrealistic (e.g. all authors remain active throughout the 20 year period, the number of coauthors and the number of references for each paper is fixed, each author publishes in one fixed category only), the simulation achieves surprisingly good results compared to a 20 year dataset of articles published in PNAS. Goldstein, Morris, and Yen (2005) introduced a new model for author-paper networks that is able to reproduce several informetric characteristics, e.g., Lotka's law, whereas such characteristics are not discussed by Börner, Maru and Goldstone for their model.

First citation distributions were studied by Egghe (2000a) and by Burrell (2001). They provide different methods to model the distribution of the time an item is cited for the first time. Both models account for items that are never cited. In a follow-up paper, Burrell (2002) studied the  $n$ th citation distribution. Egghe and Rousseau (2000a) also studied the effect of publication delays on the observed age distribution of references. Yu, Wang, and Yu (2005) developed these ideas further, to analyze the effect of publication delays on impact factors.



Huber and Wagner-Döbler (2001) describe scientific production in terms of the rate of production and career duration. They show that taking into account dates of publication can greatly improve data analysis.

Egghe and Rousseau (2000b) studied the effects of growth on the citing distribution: with growth there are more items to be referenced, and the faster a field grows there is more “competition” among the older papers to get into the reference list of the newer ones. van Raan (2000a) proposed a model for the aging of scientific literature. The model describes well the growth and ageing processes for long periods of time: the whole system is made up of subsystems characterized by a power law distribution. The empirical data span nearly 200 years.

The terminology “sleeping beauties in science” – publications that are not cited (or cited at a very low rate) and then suddenly become highly cited – was introduced by van Raan (2004). He provides several examples of such items. Burrell (2005c) investigated whether these sleeping beauties can be explained by the existing models of the citation process. Burrell’s conclusion is that moderate sleeping beauties can be expected, but the extreme cases are not, and “informetrics/scientometrics is not just a matter of mathematical/statistical modelling (or analysis) but should always be directly addressed at a particular problem by taking full account of its context” (Burrell, 2005c, p. 386). Examples for the “context” of early uncitedness are provided by Glänzel, Schlemmer, and Thijs (2003).

## 5. Methods and techniques

### 5.1. Mapping and visualization

With the advancement of display capabilities of computers, there has been growing interest in mapping and visualization. The ARIST chapter on visualization (Börner, Chen, & Boyack, 2003) provides an excellent overview of the techniques used in the process of mapping and visualizing knowledge domains. In addition they demonstrate the different visualization techniques on what they call the “ARIST set”, a large set of relevant publications from the period 1977–2001. In order to view good quality figures, I suggest using the copies of the chapter available at the authors’ websites: <http://ella.slis.indiana.edu/~katy/paper/arist02.pdf> or <http://www.cs.sandia.gov/projects/VxInsight/pubs/arist03.pdf>.

Two books by Chen (2003, 2004a) provide further information on visualization techniques and demonstrations of their use. One of the examples (Chen, 2004a) is the visualization of the scientific revolutions in the field of “string theory” (Chen & Kuljis, 2003). In a follow-up paper (Chen, 2004b) the progressive evolution of the string theory network is visualized. Information visualization is also useful for the development of scientific debates and tracking the movement of paradigms. This is illustrated (Chen, Cribbin, Macredie, & Morar, 2002) for mass-extinction and for the possible relation between mad-cow disease and a variant of the Creutzfeld-Jacobs disease in humans. Chen and Paul (2001) describe the usage of Pathfinder networks for the creation of author cocitation maps; in (Chen, Paul, & O’Keefe, 2001) they illustrate their use for the domain of computer graphics. Howard White (2003a), one of the pioneers of author cocitation analysis, emphasizes the advantages of using Pathfinder networks for author cocitation mapping over the classical techniques: multidimensional scaling, hierarchic clustering and factor analysis. In a most-recent paper, Chen (2006) describes the newly developed system CiteSpaceII. In this system, a specialty is visualized based on its research fronts (emergent and transient grouping of concepts and research issues) and intellectual base (citations and cocitations).

Garfield and colleagues (Garfield, 2004; Garfield, Pudovkin, & Istomin, 2003) discuss the need for algorithmic historiography, and illustrate the use of HistCite<sup>TM</sup> in a number of cases, e.g., bibliographic coupling, gene flow and small world. Algorithmic historiography allows studying the origins of a topic and its development over time.

Boyack, Wylie, and Davidson (2002) describe the data visualization tool, VxInsight<sup>®</sup> and its use for domain analysis. Their placement algorithm uses a technique similar to simulated annealing. Landauer, Laham, and Derr (2004) discuss the usefulness of latent semantic analysis (a method for dimension reduction, see Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990) in visualization. They claim that this method has good results when the number of dimensions is about 300, but of course so many dimensions cannot be visualized. Thus they suggest the use of a high-dimensional dynamic viewer with user control.

Leydesdorff (2004a, 2004b) suggests the use of biconnected components in order to produce clusters of journals appearing in JCR. Morris, Yen, Wu, and Asnake (2003) prefer bibliographic coupling over cocitations for identifying a time line of research fronts, where a research front according to their definition is any cluster of ten or more documents, such that at least 40% of the documents in the cluster reference one or more specific items. They illustrate their method for Anthrax literature. Lin, White, and Buzydlowski (2003) built a very nice application that works in conjunction with

the ISI Citation Indices and produces interactive, real-time author maps based on cocitation data that allow the users to easily focus on specific sub-areas (e.g., for authors working in many areas or for several authors with the same name). He and Hui (2002) also visualized related authors (based on cocitation analysis) for a set of information retrieval papers indexed by the Social Science Citation Index. In their application there is also a hypertext link to the full text of the papers in the set.

Boyack and Börner (2003) created visualizations where grants and publications appear on the same map—enhancing the usability of the map for science policy makers. In a recent paper, Boyack, Klavans, and Börner (2005) experimented with several similarity measures in order to create a global map of science. In contrast, Mahlck and Persson (2000) mapped author networks at the departmental level, based on co-authorship and cocitation.

Noyons (2001) discusses the use of bibliometric maps as a research management and science policy tool. Noyons advocates the use of co-word analysis instead of co-citation analysis for the creation of bibliometric maps and emphasizes the advantages of dynamic maps and interfaces accessible through the Web. In order to be effective, the maps need to be validated by the scientometrician (who created them), by the evaluated scientists (of whom the maps are) and by the policy makers (the users). Cahlik (2000) compares the two main methods used for creating science maps: co-word and co-citation analysis. Cahlik claims that in case there are maps based both on co-word and on co-citation analysis it is easier to interpret the maps.

Persson (2001) discusses the shortcomings of first-author vs. all author citation counts for visualization. First author citation analyses were natural when data on all authors were hard to find, but now when bibliographic databases index all authors (although currently Scopus has an upper limit on the number of indexed authors), there is no need to consider the first author only, especially since the role of the first author is field dependent. Egghe, Rousseau, and Van Hooydonk (2000) reviewed different accrediting methods for countries, institutes, research groups or authors, based on their publication or citation output. Rousseau and Zuccala (2004) define several methods for counting cocitations (e.g. first authors only, all authors, cocitations only and cocitations and coauthorships). Zhao (2006) experimented with all author versus first author cocitation analysis.

## 5.2. Text and data mining and other linguistic techniques

Losiewicz, Oard, and Kostoff (2000) provide a detailed review of the current techniques applicable for data mining in the scientometric setting. Kostoff, Braun, Schubert, Toothman, and Humenik (2000) apply the technique called “database tomography” patented by Kostoff, Miles and Eberhardt (1995) and supplement it with more standard bibliometric techniques to analyze the research literature of fullerenes. A similar study was carried out in the areas of aircraft science and technology (Kostoff, Green, Toothman, & Humenik, 2000). In a third study (Kostoff, del Rió, Humenik, García, & Ramírez, 2001) the database of the analyzed items was defined as the set of articles citing a given highly cited article. In this case too, text mining techniques (from the abstracts of the citing articles) and more traditional bibliometric methods were used for characterization of the set. Zitt and Bassecoulard (2006) used similar methods to delineate complex scientific fields. The method was applied to nanoscience.

Porter, Kongthon, and Lui (2002) advocate the use of “research profiling” when working on a literature review. Using data mining techniques, the regular literature review can be improved by providing additional relevant information, like identifying a range of information sources, examining trends, finding intersecting interests and mapping topical relationships.

Swanson, Smalheiser, and Bookstein (2001) generated a list of viruses that may have the potential of becoming biological weapons based on data retrieved from Medline. They identified two relevant sub-literatures (with very few articles in common): one on the genetic aspects of virulence and one on virus transmission. Virus terms common to both literatures served as the basis for generating the list of potential viruses. Stegmann and Grohmann (2003) showed that cword analysis applied to Medline keywords (MESH term) is able to reveal paths between complimentary but disjoint literature.

Leydesdorff and Hellsten (2005) used co-word analysis of words extracted from titles to study the structure of items on “stem cells” in the scientific, patent, newspaper and Internet domains. The main finding of the study was that the domains differ both in the way the words are organized and in terms of the underlying processes that generate structures in the communication. Onyancha and Ocholla (2005) studied the associations between HIV/AIDS and “opportunistic infections” using co-word analysis. The more common infections in HIV infected person exhibited stronger associations with HIV in the cword analysis.

Bhattacharya, Kretschmer, and Meyer (2003) studied the interactions between the scientific and patent literatures in the study of “thin films” using patent citation and co-word analysis. They also found that different intellectual spaces are associated with patent, non-patent references in patents and in the scientific literature. Co-word analysis was used to study the intellectual structure of the literature on information retrieval (Ding, Chowdhury, & Foo, 2001)—in this study only research articles were consulted, but unlike the previously mentioned studies, their data spanned ten years, and thus they were able to study the dynamic changes in the field. They (Ding, Chowdhury, & Foo, 2000) have also conducted a journal cocitation analysis of the same topic and the same time period.

Glenisson, Glänzel and Persson (2005) compared full-text and title and abstract only co-word analyses for a set of 19 articles: although full text analysis provided somewhat better results (compared to manual classification of the articles), they saw some indications that when considering the fulltext, relevant keywords are “drowned” in methodological terms. The results were complemented by traditional bibliometric methods, resulting in what they called a “hybrid methodology”.

### 5.3. Network analysis

Network analysis is a huge topic which includes complex networks and social network analysis. In this review the scope is limited to analysis of collaboration networks. Hypertext network analysis is reviewed in the section of webometrics. The basics of social network analysis (SNA) basics are clearly explained by Otte and Rousseau (2002), and its use is demonstrated for a network of authors in the field of social network analysis.

Barabási et al. (2002) and Farkas et al. (2002) studied co-authorship networks as complex systems (for an extensive review of complex systems see Albert & Barabási, 2002 or Newman, 2003a), and specifically analyzed co-authorship networks and their evolution over time in the areas of mathematics and neuroscience. Barrat, Barthélémy, Pastor-Satorras, and Vespignani (2004) studied a weighted variant of the co-authorship network, where the weight assigned to an edge between author  $i$  and author  $j$  is the sum of their coauthored papers, where each co-authorship is normalized by the number of authors collaborating on the given paper—a measure suggested by Newman (2001a). The co-authorship data for the Barrat et al. study were taken from the Physics e-print archive. Newman (2001b) studied dynamic aspects of collaboration networks, and showed that for the networks based on Medline and the Physics e-print data the probability of collaboration is correlated with the number of mutual acquaintances and their previous collaboration pattern.

Wagner and Leydesdorff (2005) viewed international collaboration as a self-organizing network based on the principle of preferential attachment. A similar approach was taken by Ramasco, Dorogovtsev, and Pastor-Satorras (2004).

While the previous studies viewed the co-authorship networks as complex systems, Newman (2001c) studied such networks from the point of view of social network analysis. Specifically co-authorship networks based on data from Medline, the Physics e-print archive, SPIRES (high energy physics) and NCSTRL (computer science) were considered. Girvan and Newman (2002) proposed a method to identify communities within a social network, and their method was tested on the co-authorship network of the faculty of the Santa Fe Institute. Newman (2003b) developed estimates of the number of “friends-of-friends” a person has (for co-authorship networks, this is the number of authors that co-authored papers with the co-authors of the specific authors). He demonstrated the concept on two co-authorship networks, one based on Medline data and the other on data from the American Mathematical Society. Newman and Girvan (2004) devised algorithms to discover communities within social networks and demonstrated the applicability of these algorithms for physicists who conduct research on networks. Moody (2004) applied SNA techniques to study the collaboration network arising from data from the Sociological Abstracts.

Collaboration is not always modeled in terms of coauthorship, for example White, Wellman, and Nazer (2004) analyzed the social ties, cocitations and inter citations of Globenet, an international and interdisciplinary group of researchers studying human development. Zuccala (2006b) examined the cocitation, coauthorship and collegial interaction patterns of a group of mathematicians involved in Singularity Theory. Kretschmer and Aguillo (2004) examined the offline and online collaboration patterns of the COLLNET group, and found by applying techniques from social network analysis that the structures of the offline and online networks are similar. Balconi, Breschi, and Lissoni (2004) studied a network of patent inventors and the special position of academic inventors within this network.

There are a large number of social network analysis tools, e.g. Pajek (<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>), see (de Nooy, Mrvar, & Batagelj, 2005) and UCINET (<http://www.analytictech.com/ucinet/ucinet.htm>). These tools are able to calculate various SNA measures and create visualizations of the networks.

#### 5.4. Classification

In this section, we review a few studies that were carried out during the reviewed period on classification. Glänzel and Schubert (2003) introduced a new two-level hierarchical system of scientific fields using a three-step approach: setting the categories, classifying the journals and classifying the articles. Pudovkin and Garfield (2002) proposed a new classification of journals, based on what they call the “relatedness factor”: the number of citations journal A received from journal B during the current year (all years for A) divided by the number of papers published by A in the current year times the number of references in papers published in journal B during the current year. The result is multiplied by one million to give the “relatedness factor”  $R_{B>A}$  (note that the measure is not symmetric). This measure was used to identify journals with genetic content—some of these journals were not included in the JCR category “genetics and heredity”. Leydesdorff (2006) used factor analysis based on aggregated citation relations to define journal categories. The categories are dependent on the rotation, thus unambiguous classification is not possible using this method.

#### 5.5. Other methods

van Raan (2005a) created a network based on bibliographic coupling. He found that the network has different statistical properties depending on the age of the references used to build the network.

Dietz, Chompalov, Bozeman, Lane, and Park (2000) discuss the utility of the CV as a data source for studying career paths of scientists and engineers. They discuss the methodological issues related to data collection and to coding and show the applicability of the method on a set of 281 CVs. Gaughan and Bozeman (2002) also examined CVs in order to try to understand how NSF Center funding affects researchers’ publication rates and their obtaining industry grants. Funded centers showed an increase in obtaining industry grants, but there was no improvement in publication rates.

Wormell (2000a) used issue tracking – a method that allows understanding how a concept is moving through various publication forms – to study how the Welfare State concept was moving from being a theoretical to a practical issue in Denmark. Several data sources, both scientific and popular (newspapers) were examined. The bibliometric analysis of the topic appears in (Wormell, 2000b). Recently, Thelwall, Barjak, and Kretchmer (2006) and Thelwall, Vann, and Fairclough (2006) introduced “Web issue analysis” that uses both text and link analysis of Web pages in order to identify key issues of a given topic, in this case, integrated water resource management.

Kleijnen and Van Groenendaal (2000) introduce a new method for evaluating the quality of publications. The method uses citations, but the citation window is infinite, and it also takes into account (through sampling) proceedings and books. The number of citation is not normalized by the number of publications. They apply their method in the area of Information Systems, a rather problematic field for evaluating quality, probably because of its multidisciplinary, publication delays and proceedings publications.

Jarvelin, Ingwersen, and Niemi (2000) propose a user-oriented interface for informetric analysis. They model bibliographic data as complex objects in a relational data structure. The query interface resembles SQL. Wolfram (2006) also suggests the use of SQL for processing informetric data. Ding, Chowdhury, Foo and Qian (2000) developed a Web interface that allows retrieving and visualizing information on authors, journals and keywords.

## 6. Citation analysis

Again, we start with reviewing the relevant books in the subarea. We start with “The Web of Knowledge – A Festschrift in honor of Eugene Garfield” (Cronin & Atkins, 2000). The chapters of the book were written for Eugene Garfield’s 75th birthday. Although the chapters encompass more than citation analysis and citation indexing, these two are major themes of the Festschrift.

In 2005, Moed (2005) published a book on citation analysis. The book studies the evaluation of scholarly research performance of individuals, groups and institutions, and aims to show that citation analysis is a useful tool for such assessment. The book explores the uses and limits of citation analysis; and is intended for both the scholarly community and for research policy officials. The book presents a number of studies carried out by Moed and his colleagues and builds upon the works of many other scholars.

Davenport and Cronin (2000) look at citations networks from a different point of view. They consider them as prototypes for representing trust in virtual environments. When A cites B, A assumes that B's claims are supported, and that works cited by B have been consulted and influenced B's thinking. Thus citation networks can be viewed as networks of trust. Citation indexes may be viewed as a recommender system (recommending highly cited items). Davenport and Cronin propose to apply this theory to Web, because they claim that "[t]hese affinities exists because 'citing' and 'linking', in cases where linking is not trivial, arise from a common concept, 'trust'" (Davenport & Cronin, 2000, p. 518). Cronin (2000) views references, citations and acknowledgements as signs, and discusses the reciprocal relations between references and citations. Small (2004) explains the relation between Merton's theoretical view on the social system of science and the theory of citations. Citations are viewed both as symbolic payments of intellectual debt and symbols of meaning. White (2004) discusses the different uses of citation analysis in applied linguistics (English for research purposes), history and sociology of science and information science. Perkel (2005) interviewed several information scientists on the future of citation analysis in light of recent developments on the Web.

Fang and Rousseau (2001) view citation networks as lattices. In these lattices diamonds, corresponding to bibliographic coupling and to cocitation are basis structures. Adams (2005) demonstrated that for a set of science papers, early citation counts correlate with later citation counts; possibly allowing policy makers to reach decisions based on early citation counts, say within a year of publication.

Lange (2001) discusses the issue of counting first authors versus all authors, and demonstrates the differences for two psychology journals. The finding shows that first-author indexing provides "roughly valid results".

In a case study, van Leeuwen, Moed, Tijssen, Visser, and van Raan (2000) showed that when considering the publication output of a large, German medical faculty's publication impact, removing the papers published in German language journals indexed by ISI from the list of papers considered increases the impact scores for this faculty increase by 15%, indicating what they call "a serious language-bias" (p. 155). Zitt, Ramanana-Rahary, and Bassecoulard (2003) also observed that the inclusion of nationally oriented journals may decrease the impact of these countries. The initial study by van Leeuwen et al. was extended to study the medical output of Germany, France, Switzerland, Italy, Spain and Japan when analyzing all ISI indexed publications versus analysis of only English language publications indexed by ISI (van Leeuwen, Moed, Tijssen, Visser, & van Raan, 2001).

### 6.1. Characterizing publications and fields based on citation analysis

Patsopoulos, Analatos, and Ioannidis (2005) examined whether the medical study design (e.g., meta analysis, randomized controlled design, case study) influences citation impact. Based on a set of articles from 1991 and 2001 they reached the conclusion that meta-analyses receive the largest number of citations. Kjaergard and Gluud (2002) showed that articles describing randomized medical trials with positive outcomes are cited significantly more than similar articles reporting a negative outcome. Ioannidis (2005) studied a set of highly cited publications (publications that received more than 1000 citations) on clinical trials, and found that for about 32% of the items, the results in these publications were later contradicted or were shown to have a lesser effect than what was originally published, with significantly higher percentage for non-randomized studies.

Citation impact in the social sciences has been studied during this period as well. van Dalen and Henkens (2001) constructed a model to explain what makes a demography article influential. They found that author reputation has some affect and the journal reputation has a major affect. In addition, focusing on the US or Europe in empirical studies increases citation frequency and publishing in French is a clear disadvantage. Holden, Rosenberg, and Barker (2005) reviewed citation studies in the area of social work and concluded that social workers publish in scholarly journals. Griggs and Proctor (2002) analyzed the referencing patterns in introductory psychology textbooks. Freud and Piaget are the most frequently referenced psychologists. Although they analyzed references, they call their study a citation analysis. It seems that a number of authors do not differentiate between references and citations. Price clearly differentiates between the two terms: "It seems to me a great pity to waste a good technical term by using the words citation and reference interchangeably. I therefore propose and adopt the convention that if Paper R contains a bibliographic footnote using and describing Paper C, then R contains a reference to C, and C has a citation from R. The number of references a paper has is measured by the number of items in its bibliography as endnotes, footnotes, etc., while the number of citations a paper has is found by looking it up [in a] citation index and seeing how many others papers mention it." (Price, 1986, p. 284).

As for the humanities, Thompson (2002) investigated whether monographs are still a major reference source in the humanities, specifically for studying 19<sup>th</sup> century British and American literature. Her findings show that monographs are still a major source both for primary and secondary materials, as nearly 80% of the references were to books in the analyzed set. Thomson speaks of “citation patterns”, even though in reality she analyzed “referencing patterns”.

Publications and patents generated by the pharmaceuticals industry were analyzed as well (McMillan & Hamilton, 2000). Rather interestingly, these companies publish a lot, and the publications receive a considerable number of citations as well.

Aksnes (2003a) examined the characteristics of highly cited papers, based on a set of highly cited Norwegian papers. He found that typically such articles are multi-authored, often involving non-Norwegian authors as well; and review articles are over-represented in the set. In another paper (Aksnes, 2004) the effect of a few highly cited papers on national indicators was studied. He found that these papers have considerable influence and cause great annual fluctuations in the average citation impact. He suggests using the median instead of the mean and in any case to consider the citation distribution and not just the mean or the median. On the other hand, Tijssen, Visser, and van Leeuwen (2002) found highly cited papers a useful national indicator.

Ramos-Rodríguez and Ruiz-Navarro (2004) based on the references of papers appearing in the Strategic Management Journal, carried out a citation and cocitation analysis, in order to identify the structure of the field.

### 6.2. Cocitation analysis—theoretical and methodological aspects

In the recent years there has been considerable discussion on the appropriate similarity measure to be used for cocitation analysis, there are colleagues who favor Pearson's  $r$  (White, 2003b; Bensman, 2004), and those who criticize its use (Ahlgren, Jarneving, and Rousseau, 2003). Leydesdorff (2005) suggests using a logarithmic transformation on the data. The idea of applying transformations to the citation data is further developed by Leydesdorff and Bensman (2006). Leydesdorff and Vaughan (2006) claim that the use of Pearson's  $r$  is not appropriate for symmetrical matrices (cocitation matrices), while it can be used in the asymmetrical case (citation matrices).

Gmur (2003) experimented with different cluster formation methods, and demonstrated them for the field of organization science. An example of a cocitation study conducted during the reviewed period is an author cocitation analysis in the field of human behavioral ecology (Sandstrom, 2001).

Small (2006) applied cocitation analysis of highly cited papers to track the emergence and growth of research areas. Co-citation clusters were created for overlapping periods of time, and the technique is illustrated for some emerging fields, e.g., webometrics, Alzheimer's disease and SARS.

### 6.3. Self-citation

Rather interestingly, the most highly cited item in this category is an article by Fassoulaki, Paraskeva, Papilas, and Karabinis (2000) that examined journal self-citing and self-cited rates in six anaesthesia journals and was published in the British Journal of Anaesthesia. The article reaches the conclusion that journal self-citations have a positive influence on the journal's impact factor. In spite of this, none of the citations of this paper listed in WoS are journal self-citations. Six anaesthesia journals were examined in 1995 and 1996, and the highest self-citing rate (number of references to the given journal divided by the total number of references in the journal) was 57%. Peritz and Bar-Ilan (2002) calculated the self-citing rate of Scientometrics for 1990 and 2000, and the results were 12.1% and 20%, respectively. Nisonger (2000) examined the effect of removing self-citations from all information and library science and genetics journals on the journal impact factor. He concluded that librarians can use the impact factor without removing journal self-citations. In a later study, Fassoulaki, Papilas, Paraskeva, and Patris (2002) demonstrated negative correlations between the impact factor and the self-citing and the self-cited rates.

An interesting linguistic analysis of author self-citation and self-mention in different disciplines was published by Hyland (2001, 2003). The analysis presented is both qualitative and quantitative. Although the 2001 and 2003 articles are almost certainly based on the same data (although there are slight differences in the results presented), Hyland decided not to self-cite the 2001 paper in the 2003 publication, perhaps operationalizing the title of his 2001 paper: “Humble servants of the discipline?”

Aksnes (2003b) carried out a large-scale author self-citation study of Norwegian authors for articles published between 1981 and 1996. He calculated diachronous self-citation rates (called self-cited rates by Fassoulaki et al.,

2000), i.e., the number of self-citations of a paper divided by the total number of citations that paper received in the given period). The findings of the paper are, that during the initial period after publications self-citation rates are higher, the number of authors of the paper influences self-citation rates, and that self-citation rates differ among disciplines. Glänzel, Thijs, and Schlemmer's (2004) results for the whole set of indexed items by WoS for 1992 support Aksnes' (2003) findings that self-citation rates decrease over time. Another interesting finding (Glänzel et al., 2004) is that self-citations and non-self citations (called foreign citations) are not independent—indicating that self-citations are an organic part of the citation process. In a later study, Glänzel and Thijs (2004) showed that at the macro level there is no need to exclude self-citations when calculating bibliometric indicators. Pichappan and Sarasvady (2002) discuss the theoretical aspects of self-citation.

#### 6.4. *Ego-centered citation studies*

Ego centered citation studies start with an individual author. In this case it is possible to study: the chosen authors' coauthors, citation identity (authors the chosen author cited), citation image makers (authors citing the chosen author), and citation image (authors co-cited with the chosen author). White (2000) first applied this technique for the works of Eugene Garfield and in a later paper presented the citation identity of eight information scientists (White, 2001a). Cronin and Shaw (2002a) generated the citation identities and images of another set of information scientists. In White (2001b), the appropriate data collection methods for generating personal profiles (including the ones mentioned above) are explained. All the above-mentioned papers gathered citation data from the ISI citation indices, while Bar-Ilan (2006a) relied on WOS, Citeseer and Google Scholar for generating the list of collaborators, the citation identity and the citation image makers of computer scientist, Michael Rabin.

#### 6.5. *Quality assessment*

The top one hundred most cited articles in surgery for the period 1945–1995 were tabulated by Paladugu, Schein, Gardezi, and Wise (2002). The article was cited 11 times according to WOS and 14 times according to Scopus – all the listed citations came from medical journals, showing the interest of medical researchers in quality assessment. Redman, Willett, Allen, and Taylor (2001) analyzed the citations to the ten most highly cited papers of the Cambridge Crystallographic Data Centre. This data center houses the Cambridge Structural Database and it is widely used by industrial and academic researchers in the field. Meho and Sonnenwald (2000) supplemented citation data with peer evaluation and content analysis of citations and book reviews. They found good correlations between citation counts and peer evaluation for high or low ranking researchers, but the results were inconclusive when considering medium-ranked researchers. Aksnes (2006) compared citation counts and the opinions of the authors about their papers. Here too, like in the case of peer assessment, there results were highly similar for major and minor contributions.

Kostoff (2002) discusses the challenges involved in evaluating research teams based on citation data. He suggests gathering data from multiple subject experts, collecting a set of papers on the topic for comparison purposes and evaluating the set of comparison papers manually, making sure that they are actually similar to the output of the research team to be evaluated.

#### 6.6. *Errors in citation analysis and statistics*

In 2002 (Anon., 2002), Nature published a short opinion paper, where it alerts to errors in citation statistics published by ISI. This paper was cited 16 times since then according to Scopus, it is also indexed by WOS, but has not received any citations.

Ho (2004) studied the references to a paper of Lagergren from 1989. Based on the errors in the references, he tries to identify the source from which the reference was copied. Fenton, Brazier, De Souza, Hughes, and McShane (2000) examined what they called the “accuracy of citations”—actually they examined the accuracy of the references in otolaryngology/head and neck surgery journals. They found errors in 37.5% of the references. A similar study was carried out for the nursing literature by Lok, Chan, and Martinson (2001). Wetterer (2006) provides a concrete example of how quotation errors and citation copying led to the mistaken story about the extinction of ants on Madeira.

## 7. Indicators

### 7.1. *h-Index*

We start this section with reviewing developments related to the *h*-index. The *h*-index was defined in 2005 by Hirsch (2005): “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  $\leq h$  citations each.” (Hirsch, 2005, p. 16569). The obvious advantage of the *h*-index is that it is a single indicator that characterizes the scientific output of a researcher. Batista, Campitelli, Kinouchi, and Martinez (2006) showed that the *h*-index is field dependent. They propose to correct the measure to account for multi-authorship.

Hirsch demonstrated the use of the *h*-index for physicists. Of course *h*-index calculations for informetricians (Bar-Ilan, 2006b; Glänzel & Persson, 2005) and information scientists in general followed soon (Cronin & Meho, 2006). Saad (2006) compared data obtained from the Web of Science (WoS) and Google Scholar for consumer scholars, Bornmann and Daniel (2005a) studied the relation between the *h*-index and the acceptance of post-doctoral grants.

The *h*-index is applicable not only to individuals. Braun, Glänzel, and Schubert (2005) and Braun, Glänzel, and Schubert (2006) applied the definition to journals instead of individuals, while van Raan (2006) studied the applicability of the measure to research groups. Banks (2006) studied the *h*-index of scientific topics and compounds (called the *hb*-index).

On the theoretical side, it was quickly noticed that the *h*-index is time dependent (i.e. the longer a researcher is active the better chance he has for a higher *h*-index; see Kelly & Jennions, 2006; Liang, 2006). There are some suggestions for models (Egghe & Rousseau, 2006b; Glänzel, 2006) and for improvements (Batista et al., 2006; Egghe, 2006).

The research in this area is still in its “infancy”, thus one can expect future developments to be covered in a follow-up review.

### 7.2. *Impact factor*

Glänzel and Moed (2002) review the background and history of the impact factor, explain its strengths and problems and analyze technical and methodological aspects of the indicator. A very clear review of the types of impacts factors and the major issues surrounding them was written by Rousseau (2002b). Ingwersen, Larsen, Rousseau, and Russell (2001) explain the difference between synchronous and diachronous impact factors. Garfield (2000, 2006) also explains the history and the proper uses of Journal Impact Factors.

Bordons, Fernández, and Gómez (2002) discuss the advantages and the disadvantages of using journal impact factors for evaluations in a non-English speaking country whose national journals are very poorly covered in the JCR. Winkmann, Schlutius, and Schweim (2002) alert to the considerable English language bias of the journals listed in the JCR.

Jacsó (2001) discusses the distortion in the calculation of the impact factor caused by ISI’s definition of “citable” items. In fact, Jacsó’s paper discusses issues already well-documented (Moed & van Leeuwen, 1995, 1996). Based on the considerable number of citations to Jacsó’s article, it seems that researchers were not aware of this problem before. van Leeuwen and Moed (2002) discuss additional shortcomings of JIF (journal impact factor): no field sensitiveness and inadequate citation window. They propose a new measure called Journal to Field Impact Score (JFIS) that takes into account the above-mentioned shortcomings. The JFIS is normalized per field, and takes into account all four types of publications: articles, letter, notes and reviews and is based on flexible citation windows. Ramírez, García, and Del Río (2000) and Sombatsompop and Markpin (2005) also suggest to normalize the IF according to the specific field. Egghe and Rousseau (2002c) define a relative impact factor, which is more sensitive to changes of relative contributions of journals than the Ramírez-García-Del Río normalization. Coelho et al. (2003) propose to adjust the JIF according to the specific field as well.

Buela-Casal (2002) proposes to take into account the reputation of the journal from which the citation emanates. Actually, the idea of assigning weights to citations was already proposed by Pinski and Narin (1976) and is successfully applied for Web link analysis, and was introduced to the Web by Brin and Page (1998). Bollen, Rodriguez, and van de Sompel (2006) also propose a weighted measure of citations, one that is based both on popularity (the number of citations) and prestige (expert appreciation).



Longer citation windows, under generally accepted assumptions (Rousseau, Jin, Yang, and Liu, 2001) result in higher impact factors. However, surprisingly, empirical results based on the Chinese Science Citation Database show that for 42% of the journals this was not the case.

Sombatsompop, Markpin, and Premkamolnetr (2004) propose the following modification to the impact factor: to calculate the ratio of the number of current year citations to items published in the previous  $X$  years and the number of articles published in the previous  $X$  years in the journal, where  $X$  is the cited half-life of the journal. This new measure was tested for journals in polymer science and it seems to be more stable to abrupt changes in the number of citations. Rousseau (2005b) provides a mathematical explanation of the proposed modification, extends it and defines a new set of impact factors, called “percentile impact factors”. Frandsen and Rousseau (2005) generalize the definition of the JIF to allow different publication and citation periods.

Weale, Bailey, and Lear (2004) propose to use the level of non-citedness as a basis for ranking journals. The ranks based on this measure are similar to the ranks based on JIF for the ISI categories immunology and surgery. van Leeuwen and Moed (2005) also studied the relation between IF and other measures, including non-citedness.

Whereas the previous studies emphasized the shortcomings of the IF, Vinkler (2004) is satisfied with the IF (but even he slightly modifies it). He defines Garfield Impact Factor (GF) as the regular IF, but takes into account all publication types. Vinkler states that the GF is an appropriate measure for characterizing the relative eminence of journals within a set of journals appropriately selected. Vinkler (2002) also discusses the value of using the Subfield Factor and the Specific Impact Contribution—these factors are normalized GFs, which take into account field specific differences.

Solari and Magri (2000) classified the journals appearing in the JCR for 1996 into five groups based on the journals’ impact factors. The tool they developed displays all the categories to which the journal belongs and its relative rank in each of the categories.

As can be seen from the above review, researchers continue to try to overcome various shortcomings of the JIF. Still, the JIF remains one of the widely used measures for assessment of journals, groups and individuals.

### 7.3. Other bibliometric indicators

The two most cited articles in this subsection were both published in JAMA. It seems that the medical community has great interest in bibliometric evaluations. Lee, Schotland, Bacchetti, and Bero (2002) set out to find indicators assessing the methodological quality of clinical research articles. The best indicators they found were citation rates, impact factors of the publishing journal, circulation rates, low manuscript acceptance rates and being listed on the Brandon/Hill Library List (this is a list for small medical libraries that was last published in 2001 and will not be updated anymore—see Sinai, 2007). In a study of 204 journal articles published based on proceedings abstracts in the field of emergency medicine, Callaham, Wears, and Weber (2002) found that the best predictor of the citation count after 3.5 years was the impact factor of the publishing journal.

van Raan and van Leeuwen (2002) report on a study of an interdisciplinary research institute concerned both with basic and applied research. In this paper, like in several previous studies (e.g., Nederhof & van Raan, 1987), they advocate the use of a combination of bibliometric indicators and peer review as the preferred method of evaluation. Indicators used were: average number of citations per publication (with and without self-citations), average citation rate of all journals in which the institute published and average citation rate of all journals in the specific ISI category. Morillo, Bordons, and Gómez (2003) are also concerned with interdisciplinary research—their aim was to establish a typology of disciplines and research areas according to their degree of interdisciplinarity. In an earlier paper, Morillo, Bordons, and Gómez (2001) proposed a set of indicators for studying interdisciplinarity. These indicators include: the percentage of multi-assigned journals in the ISI subject categories and pattern of multi-assignment (within the field or outside the field). Rinia, van Leeuwen, Bruins, van Vuren, and van Raan (2002) introduced indicators of interdisciplinary impact of journals: openness to articles stemming from other disciplines and external citation (citations given to the field by other fields) average and import/export ratio (the number of external citations divided by the number of external references).

Zitt, Ramanana-Rahary and Bassecoulard (2003) examined the influence of the choice of the journal set on classic bibliometric indicators. Nationally oriented journals introduce irregularities in the dataset. After removal of these journals, the authors were able to model the influence of extending the journal set. For the USA and Switzerland, and to a much smaller extent for France, Germany and the Netherlands, the extension of the journal set results in a decrease of the publication share of these countries. For all other countries the trend is the opposite.

Godin and Gingras (2000a) discuss the meaning of “scientific culture” and propose the following definition: “scientific and technological culture is the expression of all the modes through which individuals and society appropriate science and technology” (Godin & Gingras, 2000a, p. 44). They defined a set of indicators that can measure the scientific culture (input, activity and output indicators).

Leydesdorff (2003a) compared journal–journal citation patterns based on the Social Sciences JCR in 1998 and 1999. He used entropy statistics as indicators of change. The findings show that specialty formation is a mechanism of change in the Social Sciences. A similar study was carried out by him based on the Science JCR (Leydesdorff, 2002b).

Journal diffusion factor – the average number of citing journals of a journal within a given citation window – was introduced by Rowlands (2002). This measure describes the dynamics of citation reception and is intended to be a complimentary indicator. Frandsen, Rousseau, and Rowlands (2006) further extended these ideas. Lewison, Rippon, and Wooding (2005) also study citation diffusion, although from a different perspective – they consider several citation generations. Diffusion of medical research results to newspapers was studied by Lewison (2002a).

In recent years, there is increasing interest in “scientific excellence” instead of average performance of research groups. The CWTS developed several indicators to measure scientific excellence (Moed, Burger, Frankfort, & van Raan, 1985; van Raan, 1996, 2000b; van Leeuwen, Visser, Moed, Nederhof, & Van Raan, 2003). The use of these indicators is demonstrated through several concrete examples. A single indicator cannot provide a complete picture, thus the use of several indicators is advocated, including: number of papers per period, citations per paper, percentage of uncited articles during the period. The values are compared to international mean journal citation score for journals in which the group published and mean field citation score, from which further indicators are derived.

Moed, Luwel, and Nederhof (2002) consider the special problems when evaluating research in the humanities. They outline the methodology for such evaluations, which includes defining the boundary of the field, questionnaires created with the help of an expert group and analysis of publication output (with an emphasis on monographs).

Zitt, Ramanana-Rahary, and Bassecoulard (2005) experimented with different levels of aggregation (all-science, large-discipline, sub-discipline, sub-specialty and journal) in order to produce normalized ranked lists of the items based on citation data. The findings show that the rankings change considerably with a change in the level of aggregation. This result has serious implications for bibliometric benchmarking.

Bonitz and Scharnhorst (2001) look at “Matthew core journals”—these are journals for which the number of Matthew citations are the largest, where the number of Matthew citations for a journal is the sum of the actual minus the expected number of citations for a country in the journal, where the sum is only over the case where the above-mentioned difference is positive. They develop a typology of journals based on the number of publications, the number of citations, the number of participating countries and the number of Matthew citations in the journals. The authors suggest that for country level assessments, it is enough to consider Matthew core journals. Liang (2005) also looks at the observed versus the expected number of citations, but in this case these are taken into account in order to compute relative indicators of the “rhythm of science”.

The number of Nobel prizes per country is often considered an indicator of excellence (at the country level). Often, Nobel price winners were born in one country (and countries appear and disappear over the course of time), affiliated with another country at the time of receiving the award and are residents of a third country. Braun, Szabadi-Peresztegi, and Kovacs-Németh (2003) show that different attributions to countries based on this information result in considerably different rankings of the countries.

Aksnes and Taxt (2004) calculated the correlations between peer review and bibliometric indicators for the University of Bergen. Although the correlations were positive for all examined indicators, they were weak (in the range of 0.24–0.48).

Another issue of interest is usage in libraries. The traditional measure is re-shelving data, but with the emergence of electronic journals this measure can be substituted by the number of downloads. Both these measures correlate well with local citations (citations in publications of the users of the specific library), but not with impact factors, as was shown by Duy and Vaughan (2006). While Duy and Vaughan looked at the Internet as the hosting environment of electronic journals, Leydesdorff (2001b) and Leydesdorff and Curran (2000) used the Internet as the data source for studying scientometric phenomena. In the 2000 study, they reached the conclusion that the Internet can be used for studying triple helix relationships between universities, government and industries; while in the 2001 paper, Leydesdorff concludes: “the Internet is nowadays so overwhelmingly commercial that it seems no longer useful as an indicator of ‘user’ interests”.

## 8. Webometrics

Webometrics is an emerging subfield of informetrics. It was mentioned as one of the new trends in informetrics in Wilson's ARIST review (1999). Since then the topic developed considerably, it was reviewed in two ARIST chapters (Bar-Ilan, 2004a; Thelwall et al., 2005), in two articles in *Scientometrics* (Bar-Ilan, 2001; Björneborn & Ingwersen, 2001). A book on the topic of webometrics, entitled "Link analysis: An information science approach", has been written by Thelwall (2004a). It introduces link analysis which is one of the main webometric methods (just like citation analysis is one of the major informetric methods) and reviews applications.

Webometrics, or more exactly "Web impact factors, Web links, Journal Web sites, university Web sites and counting links" is also defined as one of the research fronts in ISI (Thomson)'s Essential Science Indicators. A research front is defined as "a group of highly cited papers, referred to as *core papers*, in a specialized topic defined by a cluster analysis" (Essential Science Indicators, 2005). As of July 2007 the only other research fronts related to informetrics that we were able to locate are on self-citations ("Core papers in author self-citations, bibliometric approach, scientific communication, macro and role"), on the *h*-index ("Individual's scientific research output, 147 chemistry research groups, standard bibliometric indicators, peer judgment and work") and bibliometrics in social work ("Social work, bibliometrics, bibliometric analysis, potential decision, tracing thought"), however the webometrics research front includes the largest number of core papers and all the papers in this research front have greater citation counts than the papers in the other research fronts mentioned here.

Cronin (2001b) wrote about the opportunities the Web presents for bibliometrics, while Egghe (2001) discusses the challenges in applying informetric analysis to the Internet. Among other topics, the differences between citation and link analysis are explained. Park and Thelwall (2003) review two major approaches to hyperlink studies: hyperlink network analysis, based on social network analysis, and webometrics.

The Web and the Internet influence research activities as well. Results of a survey (Barjak, 2006a, 2006b) indicate that there is a positive relationship between Internet use and research productivity.

### 8.1. Web Impact Factors, other measures and methodological issues

Informetrics, until the emergence of the Web dealt mainly with printed publications. One of the major sources of informetric data are the bibliographical databases, especially the citation databases. The Web and Web data are quite different, both because of the size of the Web, the dynamic nature of Web pages and Web links and because of the available tools to study the Web (crawlers and commercial search engines). Thus it is of utmost importance to understand the challenges, the features and the limitations of this new medium and the tools for studying it. Because of the problems with direct application of informetric analysis—see, for example Egghe (2001), it is necessary to develop new measures and indicators to study the Web. Obviously, the Web cannot be ignored in the informetric setting, since it has become a major information and communication source in both the academic and everyday life.

At first, some studies showed that informetric techniques can be applied to the Web, for example Larson (1996) showed that cocitation analysis can be applied when citations are substituted with links; Rousseau (1997) showed that Lotka functions (power laws) can describe the distribution of domains for a given query, and also the distribution of links to the retrieved Web pages; Bar-Ilan (1997) showed the applicability of Bradford's law for postings in newsgroups and Ingwersen (1998) introduced the concept of WIF (Web Impact Factor)—the analogue of IF and showed how the WIFs can be computed using the AltaVista search engine. Bar-Ilan and Peritz (2002) reviewed the application of informetric theories and methods for the Internet. Björneborn and Ingwersen (2004) define the position of webometrics in relation to other metrics, like cybermetrics and informetrics and define the basic webometric terminology, like inlinks, outlinks, self-links and external links and introduce Web node diagrams. Thelwall (2006) discusses the problems of both direct and indirect methods for link analysis interpretation and recommends the use of triangulation.

Ingwersen (1998) defined the WIF as the natural analogue of the IF: the number of pages linking to a site divided by the number of pages in the site. External-WIF (excluding linking pages that reside on the given site) was also defined. Later these measures were refined by Thelwall (2001a), who showed that when considering links between university websites within the UK a WIF, where only pages linking to research oriented pages within the Web site divided by the number of full-time equivalent faculty members correlated best with the results of the UK Research Assessment Exercise (RAE). This kind of WIF correlated well with Australian research output as well (Smith & Thelwall, 2002). In a subsequent study, Thelwall (2003a) experimented with the WUF (Web Use Factors—outlinks divided by the number

of full time equivalent faculty members). All the WIFs and WUFs mentioned in this paragraph were based on links from other university websites from the same country only. Thelwall (2002a) also experimented with link counts from other sources (e.g. all domains, .edu, .org, .uk) and found that the correlations are high and significant for almost all the choices, when the denominator is the number of full time equivalent faculty members. Li (2003) reviewed the different approaches to WIFs. Garfield (2001) discusses the future of Impact Factors (the traditional ones, not Web Impact Factors) in the Internet age.

Early studies, (e.g., Thomas & Willet, 2000) showed little correlation between Web link counts and research evaluation. In order to overcome anomalies in the data, Thelwall suggested different models, called ADMs that allow different granularities of counting besides counting the number of Web pages, e.g. directories or domains (Thelwall, 2002b; Thelwall & Wilkinson, 2003a). The correlation between research productivity and link counts were higher under these models at least when interlinking between UK universities is considered. ADMs for outlinks were used as well (Thelwall, 2004b), and restriction to research-oriented pages together with ADMs improved the correlations slightly (Thelwall & Harries, 2003). Correlations on the departmental level were checked as well (Li, Thelwall, Musgrove, and Wilkinson, 2003; Li, Thelwall, Wilkinson, & Musgrove, 2005; Tang & Thelwall, 2003) for several disciplines in the UK the US, Canada and Australia. Bar-Ilan (2004b) found a good correlation between the number of incoming links to each Israeli university Web site from other Israeli universities and the number of items indexed by the Web of Science having an address from that university.

Thelwall (2003b) experimented with PageRank (taking into account only links in the academic Web) instead of straight link counts, but the results were disappointing. Aguillo, Granadino, Ortega, and Prieto (2006) also use the number of inlinks to a site as a major component in an indicator for measuring the presence of universities on the Web, the other components are size and the number of rich files (pdf, ps, doc, ppt and xls)—more details and the specific results can be found at <http://www.webometrics.info/>. Bar-Ilan (2004c) introduced self-linked and self-linking rates (analogues of self-citing and self-cited), and calculated these for a number of university web sites. Thelwall (2002g) proposed the concept of “Web Invocation Portfolio” that includes link data, data from the site’s logfile and non-link invocations on the Web.

There are two major data collection methods for webometric research: crawlers and search engines. Commercial search engines are not primarily intended for webometric research, thus they have many limitations for these kinds of applications. On the other hand for building, running and maintaining crawlers, technical and programming knowledge, computing and storage resources and time are needed. Thelwall (2001b) explains the basics of crawler design, presents results from the first crawl of the “WIF crawler” (Thelwall, 2001c) and explains some of the difficulties encountered in that crawl (Thelwall, 2002c). The crawler is freely available from <http://socscibot.wlv.ac.uk/>, the existing crawls and tools for analyzing the data are available from <http://cybermetrics.wlv.ac.uk/database/>. Cothey (2004) discusses the effects of different crawling policies, not for the SocSciBot crawler developed by Thelwall, but for crawling in general.

Commercial search engines are not the ideal tools for collecting data for webometric research, but they are freely available. Bar-Ilan (2005a) provided a “wish-list” of features search engines should have for carrying out webometric research and compared the list with what search engines offered as of 2005. Search engines are not always reliable, they do not always report or retrieve all the items indexed by them (e.g., Bar-Ilan, 2000a, 2002a; Mettrop & Nieuwenhuysen, 2001), search engine coverage is uneven (Thelwall, 2000; Vaughan & Thelwall, 2004) and some search engines are (or were) not responsive enough to the appearance of new web pages (Thelwall, 2001d). More recently, search engines released APIs that allow automatic retrieval of search results—Mayr and Tosques (2005) discuss the use of the Google API.

Bar-Ilan (2002b) introduced a whole set of measures to assess search engine performance over time. Vaughan (2004a) proposed some new measures to evaluate search engine performance in place of the traditional precision and recall, and also suggests measuring the stability of search results.

As pointed out before, the Web is dynamic; Scharnhorst and Wouters (2006) propose a classification of temporal changes that occur to the web graph. Bar-Ilan and Peritz (2004) demonstrate some of these changes on Web pages containing the term ‘informetrics’ for a long period of time. Koehler (2002) studied the changes that occurred to a fixed set of web pages and sites over time. Ortega, Aguillo, and Prieto (2006) analyzed the changes that occur to web sites over time in terms of web elements. In a study conducted in 2000 by Lawrence et al. it was shown that for computer science there is a steady increase in the number of URL references, but by 2000, 54% of the URL references created in 1994 have become invalid. Invalid URL references can serve as an additional indicator of the dynamic nature of the

Web. Davis and Cohen (2001) showed that 82% of the Web references that appeared in student papers in 1996 were inaccessible by 2000.

Although Larson (1996) applied co-link analysis, the Web analogue of co-citation analysis early on, Prime, Bassecouard, and Zitt (2002) warned against a straightforward application of co-citation techniques to the Web: before applying co-link analysis, data have to be cleaned carefully. Zuccala (2006a) discussed the problems related to the interpretation of co-link data.

A final methodological issue we consider here relates to the structure of the Web. Broder et al. (2000) introduced the bow-tie model for the web graph. Thelwall and Wilkinson (2003b) and Baeza-Yates, Castillo, and López (2005) studied the structure of smaller sets: national academic Webs and the Web of a country respectively applying the model developed by Broder et al. Björneborn (2001, 2004) studied the small-world structure of the Web through so-called “co-linkage chains”, where transversal links make the Web more strongly connected. Björneborn (2004) provided an improved representation of the Web graph: instead of the bow-tie model, he introduced the corona model.

Thelwall (2001e) used network diagrams to visualize the strength of interconnections between different areas of the Web. Cothey, Aguillo, and Arroyo (2006) suggest several possibilities for operationalizing the concept of a “website”.

### 8.2. The ‘academic’ web

Informetricians mainly concentrated on academic web sites and academic use of the Web. Thelwall and colleagues published a large number of studies on UK universities (e.g., Thelwall, 2001a, 2002d, 2002e). Quantitative studies of academic interlinking in other countries and geographical areas were carried out as well, e.g. for the Asia-Pacific region (Thelwall & Smith, 2002), for EU countries (Thelwall et al., 2002), for China (Tang & Thelwall, 2002) and for Canada (Vaughan & Thelwall, 2005). Vreeland (2000) studied the inlinks and the outlinks of 156 law libraries in the US. Heimeriks and van den Bessalar (2006) also studied academic interlinking at several levels: level of countries, universities, departments and individuals.

Thelwall and Harries (2004a) showed that links from personal web sites with links to universities correlate well with university interlinking. A classification of some of these links showed that pages from academic sites are linked from a wide range of contexts.

In another study, Thelwall and Harries (2004b) investigated whether websites of higher rated scholar have more online impact. The results show that for the UK, higher rated universities attract more links from other universities in the UK, but these universities also produce more web pages, and thus there are no considerable differences in the average impact. Thelwall, Barjak and Kretchmer (2006) investigated whether gender of a research group leader influences the number of inlinks to the site of the research group. The authors conclude that “hyperlinks are not a promising source of quantitative information about gender differences in communication strategies or online visibility” (p. 373), however, personally I am glad that in terms of receiving inlinks female researchers are not doing worse than their male colleagues.

Vaughan (2006) applied co-link analysis to Canadian university Websites. The results of the analysis clearly reflected the linguistic and cultural differences within the Canadian university system, showing that co-link data contain useful information. In the academic web of Europe the major interlinking language is English (Thelwall, Tang, & Price, 2003), while in China and Taiwan most link pages are in Chinese (Thelwall & Tang, 2003).

### 8.3. Commercial sites

Vaughan and colleagues investigated in a number of papers the relationships between the number of inlinks to a commercial site of a company and the business performance of the same company. Significant and considerable correlations were found for the top-100 companies in China (Vaughan & Wu, 2004), and for the top-51 US companies (Vaughan, 2004b). In a follow-up study (Vaughan, 2004c) all Canadian and US IT companies were included. In this case too the correlations between the number of inlinks and business performance was significant. In a more recent study, Vaughan and You (2006) applied co-link analysis to map business competitive positions of 32 telecommunication companies. Companies from the same telecommunication sector were mapped together, showing the applicability of the method. A content analysis of a sample of links to North American IT companies (Vaughan, Gao, & Kipp, 2006) showed that most links were created for business purposes.

#### 8.4. Linking motivations and content analysis

Quantitative studies showed the usefulness of link analysis, but qualitative studies are also needed in order to interpret and understand the linking process. When trying to interpret links, one has to try to understand all the components: characteristics of the source page, the target page and the relation between the source and the target (the link). Bar-Ilan (2005b) proposed a framework for qualitative studies of links, and applied it to the Israeli academic web. Zuccala (2006a) carried out a co-link analysis and suggested that the framework needs to be extended for interpreting co-links.

An early study by Thelwall (2002f) concluded that “high inlink counts are not usually associated with quality scholarly content” (p. 483) when looking at links between academic web sites. Wilkinson, Harries, Thelwall, & Price (2003) found that out of the 414 links that were classified only two were equivalent to journal citations, but 90% of the links were created for “broadly scholarly reasons” (p. 49). Thelwall (2003c) identified four major types of motivations for linking in the academic environment: ownership, social, general-navigational and gratuitous. Harries, Wilkinson, Price, Fairclough, and Thelwall (2004) characterized the sources and the targets of a set of links, where the source page was from mathematics, physics or sociology. Links both to targets within the same discipline and outside the discipline were analyzed. The results show that there are disciplinary differences in linking characteristics and these differences will have to be taken into account when creating maps based on links covering multiple disciplines. Chu (2005) created a detailed taxonomy of inlinked academic pages and identified reasons for hyperlinking, based on an analysis of links to US library school websites.

Thelwall, Vaughan, Cothey, Li and Smith (2003) used content analysis in order to answer the question: which academic subjects have most online impact? The data came from the Australian and Taiwanese academic webs. Unsurprisingly computing was most visible in both countries, with technology a close second in Taiwan (but rather low in Australia).

Content analysis is applicable also for characterizing web pages not in the context of link analysis. For example, Bar-Ilan (2000b, 2000c) characterized pages on informetrics and on S&T indicators. The analysis of pages on informetrics showed that even as early as 1999, most of the bibliographic references (not the full text) that were available in bibliographic databases could be located on the Web as well.

#### 8.5. Journals and web citations

Are journal impact factors and the Web visibility of the journal website related? This question has a positive answer both for LIS (Vaughan & Hysen, 2002) and for law (Vaughan & Thelwall, 2003) journals. In addition it was found that other factors like the age of the website and the type of information available on the website (Vaughan & Thelwall, 2003) also influence the number of links the journal website attracts.

Vaughan and Shaw (2003) introduced the concept of “Web citation”, which is roughly an appearance of the title of a publication within a webpage (not necessarily as a link). They found that for library and information science research papers there is a significant correlation of around 0.6 on the average between the number of Web citations and the number of citations recorded by ISI, and usually the number of Web citations is considerably higher. This is not surprising, since their study showed that Web citations can appear in class reading lists, authors’ lists of publications, etc. In a follow-up study (Vaughan & Shaw, 2005) the study was extended to four other areas: biology, genetics, medicine and multidisciplinary sciences. Nisonger also (2005) compared citations indexed by ISI to Web citations.

Kousha and Thelwall (2006) examined so-called “URL citations”—occurrences of the URL of an open access article, which appears in the text of a Web page, not necessarily as a link. The subject area chosen was library and information science, and they identified motivations for formal and informal URL citations.

### 9. Journals

In this section we cover informetric studies related to journals – open access journals are covered in the following section. Rather interestingly, the impact factor debate refuses to disappear – among the ten most frequently cited papers in this category, seven papers deal with this issue, and only one of them was published in an information science journal. Other topics reviewed in this section include journal quality, coverage and characterization.

### 9.1. The ‘impact factor’ debate

A report in *Nature* (Adam, 2002) explains the controversies related to citation counts and impact factors and interviews several informetricians. One of the issues discussed is the definition of “citable” papers (articles, notes and reviews) used for the calculation of the denominator of the impact factor. On the other hand, all citations even to non-citable items like correspondence are counted in the numerator of the impact factor. Moed (2002b) in response to (Anon., 2002) discusses issues related to citation analysis and impact factors. Interestingly, Moed’s response appears under correspondence, and as such it is not counted in the denominator for calculating the impact factor, but nonetheless, it has been cited more than 37 times. Walter, Bloch, Hunt, and Fisher (2003) state that “[t]he journal Impact Factor and citation counts are misconstrued and misused as measures of scientific quality” (p. 280), and suggest peer ranking of “best articles” instead. This opinion paper was included in the review based on its citation count, quite possible it would not have been reviewed had we followed their advice. Neuberger and Counsell (2002) discuss the limitations of the Impact Factor—again, all the points raised are well-known, but still the article has already been cited more than 25 times. Concerns regarding the usage of Impact Factors are raised by Kurmis (2003) as well. Bloch and Walter (2001) conclude: “[i]rremediable problems, both conceptual and technical, make the IF a flawed measure” (p. 563). Lundberg (2003) states that “[t]he IF is a poor measure of the worth of journals, journal articles and authors” (p. 253). Whitehouse (2001) critically reviews IF, and concludes that the system is “inherently flawed” (p. 3). Frank (2003) argues that the impact factor is not a good indicator of the quality of the author’s work—a point emphasized by the infometric community. A more balanced review is provided by Jones (2003)—the recommendation is not to judge a paper by the impact factor of the journal that publishes it. Colquhoun (2003) in a commentary to an article published by Lawrence (2003) about the politics of publications, emphasizes that the impact factor is an average: a paper may receive a high number of citations in a journal with a relatively low impact factor or a paper may be lowly cited even though it was published in a journal with a high impact factor. Rogers (2001), editor-in-chief of the *American Journal of Roentgenology*, complains on the “numbers game”, manipulations of the Ifs by journals editors, mainly because his journal ranked only 21st in the appropriate category of the JCR. Lankhorst and Franchignoni (2001) summarize their position as follows: “Although the use of the impact factor is recognized to be invalid, nonetheless it continues, in part because no better agreed measure has yet been developed” (p. 115). We found also one editor-in-chief, Trayhurn (2002) who does not negate citation counts and impact factors, probably because the impact factor of his journal has risen sharply over the years.

With such a negative attitude, should we give up on informetrics altogether? As we have seen in the section on indicators, informetricians are also concerned with the shortcomings of the IF and suggest different ways for improvements. Perhaps there is still some hope that scientific community will accept that the IF is a measure used to assess journals (and not individuals or individual articles) which should be complemented with other measures and evaluations. Yes it can be manipulated to some extent, but this is not to say that other means of evaluation, like rankings by peers cannot be manipulated as well.

### 9.2. Journal quality

Even though there is great controversy about the usage of journal impact factors, there seems to be a genuine need for journal rankings. This problem is especially acute for relatively young, multidisciplinary areas. One such area is business information which also includes information systems. DuBois and Reeb (2000) compared business journals according to the JIF, an adjusted JIF and the results of a survey of researchers in the area. Overall the results were similar using both approaches. Donohue and Fox (2000) also found positive and significant results between survey-based and citation based rankings of journals in decision and management sciences. Similarities between user rankings and rankings based on the IF were found in well-established fields as well. Saha, Saint, and Christakis (2003) asked medical practitioners and researchers to rank a list of nine medical journals. The correlation between the user rankings and the ranking induced by the impact factors were high and significant. Barrett, Olia, and Von Bailey (2000) created sub-discipline specific journal rankings in applied economics based on citations. Lewison (2002b) also compared user rankings with impact measures for biomedical journals. In this case the citation-based measures did not correlate well with the subjective views of the UK researchers.

A different approach to create rankings is for the institution to decide on the journals and their rankings (usually not an exact ranking, but a grouping, e.g. A+, A, B and C level journals). Such lists are usually used in the promotion process. van Fleet, McWilliams, and Siegel (2000) compared such lists from several management departments. Considerable

variations were found both in terms of the contents of the list (i.e. what are the major journals in the area) and in the perceptions of the quality of the journals.

Sometimes the ranking is based on a combination of several measures. [Katerattanakul, Han, and Hong \(2003\)](#) ranked journals in information systems based on several measures, including citation counts, uncitedness and self-citation, and on an average based on all the measures. The rankings, at least of the top-journals were similar to previous rankings based on surveys. [Zhou, Ma, and Turban \(2001\)](#) proposed a method that integrates rankings based on citation counts with rankings based on expert opinion employing a fuzzy-set approach. [Forgionne and Kohli \(2001\)](#) also used multiple methods and multiple criteria to create an overall measure that enabled them to rank journals in the area of decision support systems.

[Stegmann and Grohmann \(2001\)](#) compared short-term and long-term impact factors of dermatology journals—there were only minor differences in the rankings of the journals based on the two different impact factors.

[Nederhof, Luwel, and Moed \(2001\)](#) tried to assess the quality of journals in linguistics, a field where a large part of the publications is nationally oriented, published in the national language and not necessarily in journals. For assessing the quality of journals, scholars in the field were surveyed regarding the sub-disciplines they work in, the quality of scholarly journals and the quality of publishers of scholarly books. Based on the answers three indices were constructed: a quality weight, an international visibility weight and a combined index.

[Boldt, Haisch, and Maleck \(2000\)](#) studied the changes that occurred to the impact factors of anesthesia journals during a period of 10 years. They found a constant increase of the IF of most journals over the years, and conclude that “[a]lthough the IF is the only available operational measure of journals’ scientific quality, the IF should be interpreted with great caution” (p. 848).

A different method for assessing quality is considering the number of downloads of articles published in journals. [Coats \(2005\)](#) carried out such a study for the International Journal of Cardiology. [Mabe and Amin \(2002\)](#) found that there is a clear asymmetry in the author-reader relationship: researchers as authors want to publish more, but the same researchers as readers want to read less.

### 9.3. Journal coverage and structure

[Baumgartner and Pieters \(2003\)](#) used citation analysis to study the structural influence of marketing journals on the field. The index of structural influence does not have a high correlation with the impact factor of the journal. [Pieters and Baumgartner \(2002\)](#) again used citation analysis to establish the intra and inter-disciplinary data flows for economics journals. Economics emerged as the primary source of knowledge in a network of social science and business fields. [Vessey, Ramesh, and Glass \(2002\)](#) chose the classification approach of research articles in order to study diversity of information systems journals.

[Lascar and Mendelsohn \(2001\)](#) compiled a list of journals for structural biology based on the citation and publication patterns of a group of structural biologists. The aim of this study was to recommend to which journals the authors’ institution should subscribe. [McKibbon, Wilczynski, and Haynes \(2004\)](#) conducted a large-scale categorization of medical papers, based both on scientific merit and clinical relevance. One of their findings was that articles for each sub-discipline were concentrated in a small set of journals.

[Mabe and Amin \(2001\)](#) compared the growth in the number of journals and in the number of publications in two fields and conclude that journal growth closely follows growth characteristics of the field covered by the journals.

[Davis \(2002\)](#) carried out a study of publications of Cornell researchers in order to establish a list of core journals in the life sciences for Cornell. The title of the paper implies that this is a citation analysis, but actually records of publications of Cornell researchers in the Biosis previews database are analyzed. He calculated in which journals researchers from his institution publish most frequently, but did not take into account the journals they reference, however the author assumes that list would be similar for publications only and for publications and references. One of the aims of the study was to decide on the electronic subscriptions of Cornell. The results of the study show that relatively low cost society or association journals dominate the top-ranking journals.

### 9.4. Journal profiles

[Bonnievie \(2003\)](#) depicted a multi-faceted profile of the Journal of Information Science, discussing visibility, demographic data regarding the authors, self-citations, internationalization and scientific impact. [Sprott and Miyazaki \(2004\)](#)



created a profile of the Journal of Public Policy & Marketing, based on twenty years of publications, examined the scope and depth of research topics, authorship patterns and the impact of the journal on the field. Boyack (2004) characterized PNAS using mapping techniques for twenty year of publications in PNAS. Both funding information and citation data were used and a map of highest performing papers was created. Rosenberg, Holden, and Barker (2005) carried out a citation study of articles published in the journal Social Work in Health Care, in order to understand “what happens to our ideas?” (p. 103).

West and McIlwaine (2002) examined the papers published by the journal Addiction. They compared the citation counts that papers selected from this journal received, with the quality ratings of these papers by two experts. There was moderate agreement between the experts, but no correlation with the citation counts.

Lewis and Paraje (2004) classified biomedical journals as “basic”, “clinical” or “mixed” by counting the number of basic and clinical articles in each journal. The articles were identified as basic or clinical by searching for a set of about one hundred title words for each category. This method allowed for rapid classification. Woolf and Johnson (2000) manually classified the topics in one year of publications in JAMA and in the New England Journal of Medicine. One of the more interesting findings of this exercise was the low emphasis given by these journals to preventive medicine.

Koehler (2001) analyzed 50 years of JASIS in terms of authorship and funding. He found that there was a continuous increase in the number of funded and multi-authored articles and has become more international over the years. His conclusion was that information science is no longer a “little science”, but it is not a “big science” yet.

#### 9.5. Non-English journals and internationalization

Ren and Rousseau (2002) studied the visibility of Chinese journals and Chinese authors based on data from the Science Citation Index and the Science Citation Index Expanded. They found that even the journals that are included in the citation indexes suffer from low visibility. Ugolini and Casilli (2003) studied the visibility of Italian journals among the journals indexed by ISI, and found that the journal with the highest impact factor was the Journal of High Energy Physics.

Patel and Sumathipala (2001) studied the distribution of authorship by countries in six major psychiatric journals: they found that only 6% of the published literature is from regions that constitute 90% of the global population. He and Spink (2002) studied the distribution of foreign authors in JASIS and in the Journal of Documentation between 1950 and 1999. There has been a considerable increase in foreign authorship over the year. Danell (2000) studied the relations between American and European management journals and found that the percentage of references to European journals in the American ones is only about 4%, while the references to the American journals in the European journals constituted 69% of the references, showing a clear inequality between the positions of the two journal sets. A similar study of journals from different regions (this time from six different regions) in biomedicine was carried out by Soteriades, Rosmarakis, Paraschakis, and Falagas (2006). In this case too the US journals turned out to be most influential.

Publishing in languages other than English is customary in the humanities and the social sciences, but it is not without problems, because the visibility of these publications is low. Buela-Casal, Carretero-Dios, and de los Santos-Riog (2002) carried out a bibliometric study of the four Spanish language psychology journals indexed by the ISI. Gutiérrez and López-Nieva (2001) and Paasi (2005) both discuss human geography. Gutiérrez and López-Nieva (2001) show that journals in this area have not yet attained a high degree of internationalization, which they interpret as a sign of fragmentation in the field. Paasi (2005) discusses the “monopoly” of Anglo-American journals in the area of human geography, where geographers operating outside the context of these journals are considered marginal and non-relevant. Beller (2000) would like to see more German language journals in gynecology, that would be an appropriate venue for publishing German post-graduate studies, but instead he expects a further decline in German language medical publishing.

Schloegl and Stock (2004) compared the impact factors of English language LIS journals with the impact factor of German language LIS journals, where for the German language journal, instead of the regular impact factor; the so-called regional impact factor was computed. Even though the regional impact factors not only counts citations from the specific journal and other journals indexed by the ISI, but it also includes citations from other regional journals, still the German language LIS journals were ranked lower than the leading LIS journals. Regional impact factors can also be computed using regional citation databases (Jin, Zhang, Chen, & Zhu, 2002).

## 9.6. Journal editors

Marušić and Marušić (2001) view editors as educators, and call upon editors of influential journals to act as global educators, and editors of smaller journals to act as teachers to authors, in order to overcome the problem of weaker journals supporting weaker authors and vice versa.

Davis and Müllner (2002) surveyed editors of medical journals published by societies and other non-profit organizations in order to study the independence of editors. The large majority of the editors reported complete editorial freedom, but about 30% felt that there are some pressures exercised by the society/association that publishes the journal.

Nisonger (2002) studied the influence of international composition of editorial boards on the impact factors of the journals and found that a more international editorial board is not necessarily a marker of better journal quality.

Braun and Dióspatonyi (2005a, 2005b) studied the editors and editorial boards (the “gatekeepers”) of journals in a series of papers. They studied the international distribution of the gatekeepers of core journals (2005a) and of the journals published by major publishing houses. For core journals there is a strong dominance of US gatekeepers, but when considering journals by publishing houses, there were a few cases (e.g. IOP and Oxford University Press), where there were more European than US gatekeepers.

## 10. Open access and electronic publications

### 10.1. Open access

Lawrence (2001) showed that free, online access to scientific literature increases the citations these publications receive. This conclusion is based on comparing the citation rates of more than 10,000 online and offline computer science proceedings papers. The analysis is based on the assumption that papers published in the same venue are of similar quality. Harnad and Brody (2004) compared the citation counts of open-access vs. non-open access papers from in physics and showed that open-access articles have considerably higher impact. Antelman (2004) examined the citation impact of freely available online articles in philosophy, political science, electrical engineering and mathematics and found that the freely available online articles are more cited than the offline ones. On the other hand, Anderson, Sack, Krauss, and O’Keefe (2001) compared citation patterns of freely accessible online only articles with printed, access through subscription only articles in the journal *Pediatrics*, and found that the online only articles were less cited than their offline counterparts.

What are the reasons for this higher impact? Is it early access – the paper can be self-archived even before the refereeing process is over (early access effect)? Is it caused by a quality bias – authors self-archive only their “better” papers (selection bias)? Or is it a result of easier and wider access (OA effect)? Kurtz et al. (2005a) tried to separate the three possible effects based on data from the NASA Astrophysics Data System (ADS). Their results indicate that in astronomy there is no OA effect; there is a clear early access effect and also some selection bias. Kurtz et al. (2005b) also analyzed log data from the ADS system, and Kurtz et al. (2005c) measure obsolescence as a function of the actual downloads of a publication. Wren (2005) found that for a selection of fourteen biomedical and multidisciplinary journals, the percentage of freely available online papers was larger for journals with higher impact factors, possibly indicating a selection bias. Eysenbach (2005) studied open-access versus non-open access articles published in *PNAS*, and found that open access articles are more highly cited, and the differences in the number of citation even increase over time. This finding is contradictory to other findings that show that the relative advantage of early access diminishes over time. In a recent study, Brody, Harnad, and Carr (2006) found significant correlations ( $r$  around 0.46) between downloads from Arxiv and citations in physics and astrophysics, while the strength of the correlation was lower for mathematics and condensed matter ( $r$  around 0.34).

Brown (2001) studied the effects of the e-print archive Arxiv.org on the citation practices of physicists. She found that the number of citations and impact factors to the top physics journals has not changed since the inception of the archive, but the number of references to items from the archive is increasing. It is not known how many of the papers in these top-journals were also self-archived in Arxiv.org. In another study Brown (2003) considered eprint in chemistry, and found that only 6% of the editors who responded to her survey (a 17% response rate) allowed publication of papers that previously appeared in preprint—showing considerable differences between physics and chemistry. Harter and Park (2000) conducted a larger scale survey among editors in several disciplines and achieved a much better response

rate (57.4%). Most editors did not have a formal policy regarding prior electronic publication, but most were willing to consider “certain forms of such work for publication” (p. 940).

Harnad (2001) advocates institutional repositories for self-archiving and open access. Harnad and Carr (2000) and Hitchcock et al. (2000) describe the OpCit project that link between cited and citing items. Citebase (<http://www.citebase.org/>) is a Web based search service produced by the OpCit project (Hitchcock et al., 2002). Harnad et al. (2004) explain the terminology gold (open access) and green (subscription only, but allows self-archiving) access; they claim that 90% of the journals are already “green”, but only about 20% of the articles can be openly accessed. What is now needed in order to achieve true open access is mandating self-archiving by the academic institutions. Arzberger et al. (2004) lay down the principles for data sharing and access. These principles include: openness, transparency, interoperability, quality control, flexibility and complying with ethical and legal requirements.

Gadd, Oppenheim, and Proberts (2003a, 2003b, 2003c) investigated copyright and intellectual property rights of author self-archived research papers. The results are based on a survey of 542 academics and 80 publisher copyright transfer agreements.

## 10.2. *Electronic publications*

Borgman (2000) discusses the changes in scholarly communication due to the emergence of digital libraries, and points to possible research areas in the intersection of the two areas: scholarly communication and digital libraries. Odlyzko (2002) discusses the evolution of scholarly communication towards the usage of electronic formats and provides usage data of these formats. Oppenheim, Greenhalgh, and Rowl (2000) surveyed the publishers about the future of scholarly journal publishing—at the time of the survey; the publishers were not concerned with being replaced by an alternative system.

Curti, Pistotti, Gabutti, and Klersy (2001) studied the influence of electronic availability of biomedical papers on the impact factor of journals. They found that electronic availability, even if the paper is not freely available increases the impact factor of the journals, except for surgery and otorhinolaryngology.

Kaplan and Nelson (2000) attempted to assess the impact of a NASA Digital Library, the number of downloads was considerable, but there was almost no citation impact—one possible explanation is that the material of the NASA Digital Library is not covered by the ISI citation index.

Herring (2002) studied the percentage of electronic references in electronic journals. Her findings indicate that researchers are becoming more familiar with resources available on the Web.

## 11. Productivity and publications

In this section we review papers analyzing and characterizing publications, where the unit of analysis is the publication and not the journal and the main method is not citation analysis. Studies related to journals and/or applying only citation analysis were reviewed in earlier sections.

### 11.1. *Country level studies*

King (2004) carried out a large scale publication and citation study of the top thirty nations in terms of scientific productivity. He considered publications, citations, share of top 1% papers, GDP, normalized citation counts and financial inputs. The findings indicate that the share of US is declining and the EU countries are catching up.

Rahman and Fukui (2003a) analyzed the geographic distribution of biomedical literature based on data retrieved from Medline. Longitudinal trends were studied as well. Continents were compared in terms of publications per million population, with North America ranked first followed by Australia and Europe. All continents except Africa showed a positive trend over time. In another study (Rahman & Fukui, 2003b) they compared the number of publications with a number of other country level variables, e.g., the GDP and expenditure of health. Rahman and Fukui (2002) showed the decline of the share of the US in medical research. Their findings are based on articles from thirteen leading medical journals, both basic science and clinical ones. The US is still the highest producer, but comparing data from 1991 to data from 2000, the share of US paper is decreasing.

Glänzel, Schubert, and Braun (2002) studied the productivity and citation impact of the 32 most prolific countries based on ISI data for the years 1990–1998. They considered both overall and field specific productivity and computed

the relative citation rates for countries which is the quotient of the mean observed and the mean expected citation rates for journals in the different fields and created relation charts based on this indicator. Zhou and Leydesdorff (2006) concluded that China has become the fifth leading nation in terms of scientific productivity based on data from the Web of Science as of 2004, and more specifically China has become a major player in nanotechnology. However, China's citation rates are still low compared with citation rates of other nations, but its citation rates are increasing as well. This state of affairs has been termed China's "quantitative expansion phase" (Jin & Rousseau, 2005). Leydesdorff and Zhou (2005) also studied changes over time in publication and citation outputs of China and South Korea. Dore and Ojasoo (2001) studied the publication patterns of different countries in different disciplines and the changes that occurred to these patterns over time.

Soteriades and Falagas (2005) compared biomedical research in the US and the European Union. They investigated at individual European countries and groups of European countries (EU15, EU10, EU25 and EU Candidate Countries). Even though the productivity normalized for population of several European countries exceeds the normalized productivity of the US, even the EU15 as a whole has less normalized productivity than the US. Leydesdorff (2000) raised the question whether the EU is becoming a single publication system. The answer to this question at this point of time was inconclusive.

Man, Weinkauff, Tsang, and Sin (2004) tried to understand why some countries publish more than others in leading biomedical journals. They found a significant relationship between the national spending on research and TOEFL scores (indicates the level of English language proficiency) to publication output.

### 11.2. Field specific studies

Braun, Schubert and Kostoff (2000) analyzed the growth and trends in fullerene research based on journal publications and the citation impact of these publications. Tsay, Jou, and Ma (2000) investigated the growth of the semiconductor literature between 1978 and 1997, and identified the core journals in the field. Allen, Jacobs, and Levy (2006) studied nursing literature, based on the reference lists of publications in 53 selected journals for the years 1996–2000. The findings show that the most frequently referenced format was journal papers, and it was also possible to identify a list of core journals for nursing.

Moed (2000) studied the publication strategies of nine biotechnology and molecular biology research departments. He discusses the issue of publication counts of multi-authored paper from two or more different institutions, and compared rankings based on different counting schemes. Seglen and Aksnes (2000) analyzed the productivity of Norwegian microbiological research groups.

Tsay and Yang (2005) retrieved from Medline all the indexed items that were published between 1990 and 2000 and the article type was defined as "randomized controlled trial". They found exponential growth over the years, and 42 core journals were identified. Ray, Berkwits, and Davidoff (2000) studied the fate of 350 manuscripts that were rejected by the Annals of Internal Medicine, and found that most of these manuscripts were published by specialty journals within about a year and a half.

### 11.3. Media coverage of science

Blanchard, Erlich, Montgomery, and Bovbjerg (2002) showed that breast cancer is over represented in popular magazines relative to cardiovascular disease, and the disparity widened over the years, even though the women are far more likely to develop cardiovascular disease than breast cancer. Osteoporosis related articles in women's magazines and newspapers were studied by Wallace and Ballard (2003). They found that often the information in these articles was ambiguous and incomplete.

Tobacco is another health hazard: Durrant, Wakefield, McLeod, Clegg-Smith, and Chapman (2003) examined tobacco related Australian newspaper articles, and found that most of the news items were in favor of tobacco control.

Chronic disease coverage by Canadian aboriginal newspapers was studied by Hoffman-Goetz, Shannon, and Clarke (2003). They found far greater coverage of HIV/AIDS and diabetes than cancer and cardiovascular diseases, in spite of the fact that cancer and cardiovascular diseases are relatively frequent among the aboriginal population. Chan et al. (2002–2003) studied newspaper coverage of SARS in various countries.

The Web can be viewed as a communication medium as well. McCain (2000) studied whether researchers use the Web to disseminate their research related information, and Thelwall (2004c) examined whether the Web can provide information about the commercial uses of scientific research.

#### 11.4. Interdisciplinarity

Rinia, van Leeuwen, van Vuren, and van Raan (2001) carried out an empirical study examining whether interdisciplinary research in physics is less appreciated both in terms of citations and in terms of peer review. The results for the Netherlands showed that this was not the case. Morillo et al. (2001) studied interdisciplinarity in chemistry: several bibliometric indicators were used including multi-classification of journals into JCR categories, citations and references outside the journal category and multi-assignment of the document into Chemical Abstracts sections. Rinia, van Leeuwen, Bruins, van Vuren, and van Raan (2001) observed a citation delay in interdisciplinary knowledge transfer.

Pettigrew and McKechnie (2001) studied the use of theories in information science research. They found that almost half of the theories are from the social sciences, followed by the sciences and humanities. Steele and Styer (2000) studied interdisciplinarity in forestry research in terms of authorship, subject matter and cited literature and showed that interdisciplinary methods have made a measurable and positive impact on the forestry literature.

#### 11.5. Other studies

Bonaccorsi and Daraio (2003) carried out an extensive study on the influence of age on productivity, based on data on the researchers in the Italian National Research Council. They found that age has a negative effect on productivity. Liang, Kretschmer, Guo, and Beaver (2001) studied the age structures of Chinese collaborations in computer science.

Two papers dealt with the issue of duplication or near duplication of publications. von Elm, Poggia, Walder, and Tramer (2004) studied duplication in medical research based on review articles. Knothe (2006) provided concrete examples of different types of duplications and analyzed the citation patterns of the duplicates.

Lewison (2001) compared the productivity and citation counts of female and male researchers from Iceland. Such comparison was possible, because it is possible to differentiate between male and female surnames.

van Teijlingen and Hundley (2002) tell the story of how a methodological paper authored by got finally published, after being turned down by five journals. This paper has not only been cited ten times until now (according to Scopus), but also received two warm responses in a later issue of the publishing journal.

Lu and Katz (2005) studied the use of the term “Reiter syndrome” in medical literature. Reiter, a Nazi physician authorized experiments in concentration camps, and because of this some physicians have argued against further use of the Reiter eponym. The authors found a decreasing use of the terminology without qualification (i.e., without mentioning its disfavored use) between 1998 and 2003. In 2003 an international group of rheumatology journal editors decided against further use of the eponym in their journals.

Rothman, Kirk, and Knapp (2003) studied productivity vs. reputation of social work researchers. Reputation was based on nominations from journal editors, textbook editors and heads of relevant organizations. The researchers were partitioned into three groups according to their reputation. Productivity rates varied consistently among the groups, but there were considerable variations within the groups.

## 12. Collaborations

In this section we review collaborations between nations in all fields, collaborations in specific fields either on the national or on the group level, and collaborations on the author level. Collaborations from a network point of view are discussed in the section on network analysis (e.g., Barabási et al., 2002; Wagner & Leydesdorff, 2005; Newman, 2001c).

Beaver (2001) discusses the past, present and future of scientific collaborations, starting with collaborations of French chemists in the first half of the 1800s and finishing with e-mail, e-journals and the Internet. Beaver (2004), based on a longitudinal study of 33 researchers at Williams College, reached the conclusion that multi-authored papers have greater epistemic authority than single authored ones. Wagner-Döbler (2001) studied the long-term collaboration trends (since 1800) in mathematics, physics and logic.

Glänzel (2001) carried out an extensive study on collaborations based on publication and citation counts for 1995 and 1996. Relative specializations for eight major science fields were calculated and the citation impacts of domestic versus international papers were compared. Zitt, Bassecoulard, and Okubo (2000) concentrated on five large scientific countries: the US, the UK, France, Germany and Japan. Arunachalam and Doss (2000) considered Asia in general,

while Basu and Kumar (2000) studied international collaboration in India. About 30% of the Brazilian publications are a result of international collaboration (Leta & Chaimovich, 2002) and the percentage was more or less constant over the years. For India, the collaboration percentage was 13.2% in 1990 (Basu & Kumar, 2000) and the percentage grew to 17.6% (Arunachalam & Doss, 2000) in 1998.

Glänzel and Schubert (2001) investigated the effectiveness of collaboration in chemistry; Schummer (2004) analyzed nanoscience; Leydesdorff and Heimeriks (2001) and Heimeriks, Hörlesberger, and van den Besselaar (2003) considered biotechnology and Janssen, Schoon, Ke, and Börner (2006) concentrated on subfields of environmental studies. Glänzel and Schubert (2001) looked at the citation impact; Schummer (2004) carried out a co-author analysis and found that nanoscience is not so interdisciplinary; rather it consists of several mono-disciplinary subfields. Leydesdorff and Heimeriks (2001) applied co-word analysis; Heimeriks et al. (2003) based their findings on bibliographic coupling in journal papers, coauthorships and Web link analysis. Janssen et al. (2006) gathered data through multiple methods, using both a list of relevant keywords and references to central works in the sub-areas and provided visualizations of the collaboration networks.

Glänzel (2002) explored collaboration patterns in biomedical research, mathematics and chemistry. Although the number of multi-authored papers grew over time in all three fields, there was variability among the fields both in terms of coauthorship patterns and citation impact.

Melin (2000) studied motivations and reasons for collaboration based on the results of a questionnaire and interviews with collaborating authors. Bozeman and Corley (2004) looked at collaboration strategies using the survey method. Thorsteinsdóttir (2000) analyzed publication data and interviewed authors involved in international collaboration in and with Iceland and the Canadian province Newfoundland in order to understand the reasons for collaborating with researchers from small countries/regions. Chompalov, Genuth, and Shrum (2002) interviewed researchers in several subfields of physics and four different collaboration formats were identified: bureaucratic, leaderless, non-specialized and participatory.

Braun, Glänzel, and Schubert (2001) studied collaboration trends between 1991 and 1998 in neuroscience based on publication data derived from the ISI for each year, while King (2000) looked at the authorship patterns in two neuroscience journals between 1945 and 1995 in 10-year intervals.

Persson, Glänzel, and Danell (2004) analyzed the growth of the number of publications compared with the growth of the number of references in the publications to study the interactions between productivity, co-authorship and citation impact.

The relations between social ties, coauthorship and cocitation were examined by White et al. (2004) and Zuccala (2006b), and between coauthorship and Web links by Kretchmer and Aguillo (2004).

Bookstein, Moed, and Yitzahki (2006) proposed a model for measuring international collaboration, taking into account same country preference and bias for cooperating with researchers from specific countries. The notion of own-group preference was also discussed by Egghe and Rousseau (2002d).

### 13. Research policy

The whole area of science policy is of course not in the scope of the current review; here we only discuss the more scientometric aspects of research policy.

Etzkowitz and Leydesdorff (2000) discuss in detail their Triple Helix model of university-industry and government relations for explaining research systems. Shinn (2002) studied the visibility of the Triple Helix model both in the Citation Indexes and on the Internet. He found only a small number of citations in the Social Science Citation Indexes and on the Internet. His study was carried out in July 2002. By March 2007, the results are considerably different, the Etzkowitz and Leydesdorff (2000) paper was one of the most frequently cited paper among the papers mentioned in this review (see Table 1), and the phrase “Triple Helix” occurred on 717,000 pages on the Web as reported by Google in July 2007, but as can be expected the phrase is used in other contexts as well. In addition, Shinn (2002) also discussed the Triple Helix model at length. The model is further discussed and compared to other models by Leydesdorff and Meyer (2003).

Meyer, Siniläinen, and Utecht (2003) offer concrete ways to measure Triple Helix interactions. They show based on a study of Finnish academic patents, that patent analysis alone is not sufficient, much better results are acquired, when patent data are combined with an inventor survey. Leydesdorff (2003b) proposes mutual information (an entropy measure) as an indicator of university-industry-government relations. Danell and Persson (2003) analyze the Swedish regional innovation system in terms of Triple Helix relationships.

Technological innovation relies on basic research: McMillan, Narin, and Deeds (2000), based on patent analysis, show that the biotechnology industry relies heavily on public science, even more than other industries. Hicks, Breitzman, Hamilton, and Narin (2000) show that a highly cited US paper is nine times more likely to be cited by a US patent than a random paper, pointing to a relationship between research excellence and innovation. One of the major challenges of research policy makers is to emphasize the contribution of scientific and engineering research in technical inventions. Tijssen (2002) proposes to survey inventors to understand and quantify the role of basic science for innovation. Meyer (2002) provides several approaches for studying science and technology linkage: patent citation, industrial publication activity and university and academic patenting. He also discusses the limitations of informetric methods and suggests that a combination of formal and survey schemes can be more effective. Industrial publication activity was examined by Tijssen (2004) and he found that there is a decrease in this activity over time for the selected semiconductor and bio-pharmaceutical companies. Debackere and Veugelers (2005) assessed industry-science links through innovation indicators and more specific indicators that measure the use of university/public research institutional information sources in the innovation process. Not only research in science can have practical implications, Landry, Amara, and Lamari (2001) found that at least in Canada, nearly half of the research results in the social sciences are of some use to practitioners and/or policy makers.

It should be noted that university-industry cooperation and funded research may have a negative effect on academic freedom. However, Behrens and Gray (2001) conducted a survey of graduate students in US engineering departments and found that industry sponsorship does not negatively affect student experiences or outcomes. Must (2006) studied innovation strategies using bibliometric data for countries in Central and Eastern Europe.

Coccia (2004) constructed a macro-index based on a large number of indicators to produce a research performance score. Research institutes within the Italian National Research Council were classified as high or low performance institutes.

Godin (2003, 2006b) and Godin and Gingras (2000b) studied in several papers longitudinal trends related research policy: R&D in general (Godin, 2006b); the place of universities in knowledge production (Godin & Gingras, 2000b) and the emergence of S&T indicators (Godin, 2003).

### 13.1. Research evaluation

One of the major, periodical research evaluations is the UK Research Assessment Exercises (RAE; <http://www.rae.ac.uk>). The RAE is based on assessment of publications submitted to a panel of experts. An ongoing debate is whether peer assessment should be supplemented/replaced by bibliometric indicators. Holmes and Oppenheim (2001) tried to predict the outcome of the 2001 RAE exercise based on citation counts. When comparing the prediction with the actual results (<http://www.hero.ac.uk/rae/Results/>), we see that the rankings and the research ratings are not identical, there is one case where the actual rating is 4 vs. the predicted 2 (a rather low rating, 5\* is highest). It is interesting to point out that the actual ratings are never lower than the predicted ones. Even if there are considerable differences between the citation based and the peer review based rankings, it is not clear which (if any) of the rankings is “correct”. In a later study, Norris and Oppenheim (2003) found high and significant correlations between RAE rankings and rankings based on citation counts for archeology.

Warner (2000) is critical of the bibliometric approach, although he advocates the use of a combination of methods. Weingart (2005) discusses the problems of using indicators for research evaluation. One of the issues he raises is that the results can be manipulated. An interesting example of this is Australian science, where funding is mainly based on the number of publications, as a result of which Australia’s share of publication increased by 25% between 1988 and 1997, while its citation impact decreased (Butler, 2002, 2003). Researchers can divide their articles to “least publishable units” (Weingart, 2005, p. 125).

Dodgson and Hinze (2000) review indicators for the evaluation of the innovation process. They conclude that “a much more complete reflection of the highly complex innovation process can be gained by combining a number of these different indicators for analytic purposes” (p. 112).

As can be seen from the reviewed articles, there does not exist one “correct” or “perfect” method or indicator for evaluation. van Raan (2000c) calls for “the application of objective, high-quality analytic tools based on advanced bibliometric methods” (p. 86). Roessner (2000) calls for methods and measures to be driven by theory, and not to make a definite choice between qualitative and quantitative measures, but to use multiple methods.

Cronin and Shaw (2002b) compare citation counts, Web hits and media mentions of prominent American information scientists. Rankings based on each of the measures are considerably different.

Inonu (2003) computes country rankings both according to articles per capita and GDP per capita and creates two categories according to whether the ranking based on articles per capita is lower or higher than the ranking based on GDP per capita. The results are interpreted by the influence of cultural factors on scientific production.

In a series of publications, Bornmann and Daniel (2005b, 2005c, 2006) assessed the peer review process for selecting PhD and post-doc students at the Boehringer Ingelheim Fonds (BIF). Citation analysis of the accepted and rejected applicants show a considerably higher impact for works published by the accepted applicants (Bornmann & Daniel, 2006) and the citation rates of their papers is higher than the average citation rate of the journals that published the works (Bornmann & Daniel, 2005b). The Bornmann and Daniel studies, like the results of Norris and Oppenheim (2003) show that there is high agreement between rankings by experts and by bibliometric indicators. However in the BIF case, bibliometric indicators cannot replace expert assessment, because these indicators are not available at the time the candidate applies for the fellowship.

### 13.2. University rankings

Recently, there is increased interest in rankings of universities. Examples of such rankings are the Shanghai Ranking (Liu & Cheng, 2005a) or the Times of London Higher Education Survey. van Raan (2005b) discusses the conceptual and methodological problems when ranking universities by bibliometric methods. One of the major problems is that policy makers prefer “quick and dirty” methods. van Raan’s doubts about the methods used by the Shanghai group were replied by Liu and Cheng (2005b), which in turn received a reply by van Raan (2005c). This debate is far from being over, and will obviously be covered in future reviews as well. As an example of the debated issues, Webster (2001) analyzed an earlier ranking published by the US News and World Report, and showed that the actual weights assigned to the different criteria are not as published, because of the interdependence of the ranking parameters.

## 14. Patent analysis

Oppenheim (2000) reviews patent citation analysis, and raises the question whether patent citations count. His answer is: “[d]espite the volume of work that has been carried out on the topic, the case is as yet unproved” (p. 424). In this section, we review recent developments in the field, and it is for the reader to judge whether she agrees with Oppenheim’s reservations.

### 14.1. Patent citations and science and technology links

Meyer (2000a) discussed the differences between references that appear in patents (“citations”) and references that appear in scholarly works. Tijssen and colleagues (Tijssen, 2001; Tijssen, Buter, & van Leeuwen, 2000) analyzed the front page references of patents patented in the US to study the utilization of Dutch industry relevant science. The findings indicate that there is genuine linkage between science and technology. van Looy et al. (2003) studied the science intensity of different technological domains operationalized by the number of references to scientific articles in the patents. Michel and Bettels (2001) compared the number of citations that appear on patents filed at different PTOs and also provided the percentage of non-patent references by different technical fields. Meyer (2000c) also emphasized the differences between the citation practices at different PTOs.

Meyer (2000b) used the case study method to investigate the reasons for citing scientific literature in patents. Meyer (2000b, 2000c) concluded that patent citation analysis alone was insufficient to convey a complete picture of science and technology linkage. Murray (2002) looked at patent-paper pairs to study the co-evolution of scientific and technological networks.

Verbeek et al. (2002) developed a method to study science and technology linkage through patent citation data. They observed that the distribution of non-patent references (NPRs) is highly skewed, with the majority of patents having no NPRs, and that science and technology links are concentrated in a few fields.

The abovementioned papers looked for references to scientific publications in patents, Glänzel and Meyer (2003) took the opposite approach and studied patent references in scientific publications. The findings showed that chemistry-



related subfields tended to cite patents more than other scientific areas. At this point, it is not clear how to interpret references to patents in scientific publications.

#### 14.2. Other patent indicators

Mowery and colleagues (Mowery, Nelson, Sampat, & Ziedonis, 2001; Mowery & Ziedonis, 2002) studied the changes that occurred to patenting by US universities after the introduction of the Bayh-Dole act, that allows for the transfer of exclusive control over many government funded inventions to universities. Hicks, Breitzman, Olivastro, and Hamilton (2001) studied trends in US innovation based on citation analysis. They observed a shift towards large corporate basic research laboratories and growth in university patenting. Leydesdorff (2004c) also studied patents assigned to or invented by universities—his study was not limited to US universities. He created visualizations based on the title words of the patents. Yoon and Park (2004) use text-based methods to create patent networks as well.

Acs, Anselin, and Varga (2002) studied the relationship between patent and innovation counts and concluded that patents provide a fairly reliable measure of innovation activity. Maurseth and Verspagen (2002) looked at knowledge spillovers between European regions based on patent citations. The findings indicate that there are barriers to knowledge flows within Europe. Zitt, Ramanana-Rahary and Bassecoulard (2003) also considered knowledge spillover based both on publications and patents. European regions were classified into five different levels; the top level regions were responsible for almost two thirds of the European science output and for more than half of the European patents.

Harhoff, Scherer, and Vopel (2003) found a positive relationship between the value of a patent and the number of citation it receives. Hall, Jaffe, and Trajtenberg (2005) also found that the number of citations a patent receives is a useful indicator of the patent's "importance" as indicated by the firm's value on the stock market.

Huang et al. (2003) carried out an extensive analysis of nanotechnology patents and created content maps based on the titles and abstracts of more than 70,000 patents in their data set. Citation networks on the country institution levels were also created. The study was updated to include data from 2003 (Huang, Chen, Chen, & Roco, 2004). Hullmann and Meyer (2003) and Meyer (2000c, 2001) also studied nanotechnology.

## 15. Databases

In the last section of the review, we discuss research related to bibliographic databases: comparison between databases, coverage, retrieval techniques and database specifications.

Probably the most important data source for informetric research is the ISI (Thomson) Citation Databases. Until the fall of 2004, these databases (i.e., the Science Citation Index Expanded, the Social Science Citation Index and the Arts and Humanities Index) when searched together (e.g. using the Web of Science) were the only comprehensive sources of citation data. In the fall of 2004 two new citation databases appeared – Elsevier's Scopus (Scopus, 2007) and Google Scholar (Payne, 2004). Scopus provides full citation data from 1996 and onwards. The coverage both in terms of time and journals of Google Scholar is much fuzzier – this information is not disclosed by Google. Google Scholar employs automatic techniques to extract information from electronic files that look like scholarly publications from publisher sites, conference sites and other sites (even homepages of individuals). It extracts information (title, authors, references, etc.) from the collected items automatically, thus it contains errors, but on the other hand it is a free tool.

Both Scopus and Google Scholar were reviewed and compared (mostly with WOS) in a number of papers. For example Jacsó (2005a) and Wleklinski (2005) described Google Scholar's features; Noruzi (2005) compared the results of WOS and Google Scholar for articles in the field of webometrics. Jacsó (2005b) and Bauer and Bakkalbasi (2005) compared the three citation databases. Jacsó considered the coverage in general and citation counts for the most cited items in the journal *Current Science*, while Bauer and Bakkalbasi (2005) compared the citations counts for papers published in *JASIST* in 2000. Glänzel and Persson (2005) calculated the *h*-index of Price medal winners based on WOS and Bar-Ilan (2006b) based on Google Scholar. Jacsó (2006) discussed the Bauer and Bakkalbasi results on *JASIST* citation counts. Bakkalbasi et al. (2006) compared the three databases (WOS, Scopus and Scholar) using eleven journal titles in oncology and condensed matter physics and reached the conclusion that the "study did not identify any one of these three resources as the answer to all citation needs" (p. 7).

The new citation databases were compared with other databases as well, for example Henderson (2005) compared the coverage of Google Scholar on medical topics with MEDLINE; Gardner and Eng (2005) compared Google Scholar

with several social science databases and Neuhaus, Neuhaus, Asher, and Wrede (2006) compared the content of Google Scholar with that of 47 other databases. All these comparisons discuss current weaknesses of Google Scholar, while emphasizing its potential. Bar-Ilan (2006a) used three citation databases, WOS, Citeseer and Google Scholar to gather data on publications and citations for an ego-centered citation analysis.

Citeseer (also called ResearchIndex, <http://citeseer.ist.psu.edu/>) is an autonomous citation database of computer science literature (Bollacker, Lawrence, & Giles, 2000)—items to be indexed are gathered and parsed and references are extracted automatically. In addition to displaying citing items, the citing context is displayed as well. Because of the heuristics involved, the system is not error proof. It is highly appreciated by computer science researchers, mainly because the proceedings literature not covered by the Science Citation Index is of utmost importance in this field. Goodrum, McCain, Lawrence, and Giles (2001) compared Citeseer with ISI's SCISEARCH. Zhao and Logan (2002) compared the ResearchIndex with the Science Citation Index in a specific area: XML research. One of the problems with the ResearchIndex data was duplicates, which had to be removed manually. ResearchIndex is not selective and it indexes proceedings papers as well. These are probably the major reasons for the different results obtained from the two result sets. The authors conclude that in spite of these shortcomings, using ResearchIndex is a valid method for evaluating scholarly contributions. In a more recent paper, Zhao (2005) compared Citeseer and the Science Citation Index again in the area of XML research—Citeseer included more than four times more items on XML research than ISI.

Moed (2002a) studied China's research performance based on the ISI Science Citation Index and on the Chinese Science Citation Database. He reached the conclusion that the ISI Citation Index reflects China's international scientific position, while the Chinese database is best for assessing Chinese research from the national perspective. Liang, Wu and Li (2001) investigated the differences between the Science Citation Index, the Chinese Scientific and Technical Papers and Citations and the Chinese Science Citation Database for studying the structures of Chinese medical universities.

Glänzel, Schlemmer, Schubert, and Thijs (2006) discuss the importance of proceedings literature for studying scholarly communication and the use of the ISI Proceedings database as an additional source for bibliometric studies.

Techniques and issues related to retrieving data from online databases are explored in (Hood & Wilson, 2003a; Marx, Schier, & Wanitschek, 2001). Butler and Visser (2006) showed how to extend citation analysis to non-source items using the ISI databases. The coverage of the databases and the need to use more than one database are discussed in several papers (Hood & Wilson, 2001b, 2003b; Meho & Spurgin, 2005). Braun, Glänzel et al. (2000) and Braun, Schubert et al. (2000) assessed the representativeness of the journal coverage of ISI's Science Citation Index at the level of countries, disciplines and publishers.

Ruiz-Pérez, López-Cózar, and Jiménez-Contreras (2002) explored how different bibliographic databases handle Spanish names, where the most frequent Spanish name structure consist of first name, middle name patronymic surname and matronymic surname. For more than 50% of the examined cases several variations occurred even in the same database. Buchanan (2006) studied data entry errors during indexing in SciFinder and in the Science Citation Index and found that error rates ranged between 1.2 and 6.9%. Archambault, Vignola-Gagné, Côte, Larivière, and Gingras (2006) analyzed the country and language distribution of journals covered by the ISI databases and compared them with the peer reviewed journals indexed by Ulrich.

## 16. Summary

In this paper I attempted to review the huge informetric literature published between 2000 and 2006. Some of the more recent studies may be less visible than those from the beginning of the review period, thus I recommend that the next review of informetrics should cover the years 2005 and onwards.

At the beginning of the 21st century, besides the "traditional" major informetric topics, citation analysis, theory and productivity measures that continue to thrive, we witness the emergence of new topics, like webometrics, the *h*-index and open and electronic access and the strengthening of previously existing topics, like mapping and visualization, text and data mining. A major development that influences informetrics is that besides the ISI Citation Databases there is an additional comprehensive citation database, Elsevier's Scopus, and even though it has limitations, Google Scholar provides freely available citation data. In addition the number of regional (citation) databases is growing, for example in China, Japan, Taiwan, and SCiELO in Latin America.

An additional major new development is, of course, the new Journal of Informetrics (see Mayr & Umstätter, 2007 for an explanation of why this journal appeared at exactly the "right time"). The first issue of the Journal of Informetrics

appeared in January 2007, thus the articles published in it are beyond the scope of this review, but for the next review, I expect that the Journal of Informetrics will appear among the major journals (see Table 2 for the current list) in the field.

The two major conference series, the ISSI conferences and the S&T indicators conferences continued on a bi-annual basis and several COLLNET Meetings (<http://www.collnet.de/>) were also held during this period. The ISSI Society has a well-developed Website (<http://www.issi-society.info/>) and during 2005 and 2006 eight issues of the Society's Newsletter were published.

Electronic access and its influence on scholarly communication is one of the topics reviewed in this paper. Without electronic access writing the paper would have been considerably more difficult. With a very few exceptions, all reviewed articles were accessible electronically, mostly through institutional subscription.

## Acknowledgements

I thank Leo Egghe, Editor-in-Chief of the Journal of Informetrics for inviting me to write this review article. My special thanks to Elsevier for providing access to the Scopus database. Access to Thomson's Web of Science was provided through MALMAD, the Israeli consortium providing information services to all the Israeli universities. It would have been impossible to write a reasonable review article without access to both these citation databases. I also thank the reviewers for their useful comments and suggestions.

## References

- Acs, Z. J., Anselin, L., & Varga, A. (2002). Patents and innovation counts as measures of regional production of new knowledge. *Research Policy*, 31(9), 1069–1085.
- Adam, D. (2002). The counting house. *Nature*, 415, 725–729.
- Adams, J. (2005). Early citation counts correlate with accumulated impact. *Scientometrics*, 63(3), 567–581.
- Aguillo, I. F., Granadino, B., Ortega, J. L., & Prieto, J. A. (2006). Scientific research activity and communication measured with cybermetrics indicators. *Journal of the American Society for Information Science and Technology*, 57(10), 1296–1302.
- 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(6), 550–560.
- Aksnes, D. W. (2006). Citation rates and perceptions of scientific contribution. *Journal of the American Society for Information Science and Technology*, 57(2), 169–185.
- Aksnes, D. W. (2003a). Characteristics of highly cited papers. *Research Evaluation*, 12(3), 159–170.
- Aksnes, D. W. (2003b). A macro study of self-citation. *Scientometrics*, 56(2), 235–246.
- Aksnes, D. W. (2004). The effect of highly cited papers on national citation indicators. *Scientometrics*, 59(2), 213–224.
- Aksnes, D. W., & Taxt, R. E. (2004). Peer reviews and bibliometric indicators—A comparative study at a Norwegian university. *Research Evaluation*, 13(1), 33–41.
- Albert, R., & Barabási, A. L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1), 47–97.
- Allen, M. P., Jacobs, S. K., & Levy, J. R. (2006). Mapping the literature of nursing—1996–2000. *Journal of the Medical Library Association*, 94(2), 206–220.
- Anderson, K., Sack, J., Krauss, L., & O'Keefe, L. (2001). Publishing online-only peer-reviewed biomedical literature: Three years of citation, author perception, and usage experience. *Journal of Electronic Publishing*, 6(3). Retrieved July 22, 2007, from <http://www.press.umich.edu/jep/06-03/anderson.html>
- Anon. (2002). Errors in citation statistics. *Nature*, 415(6868), 101.
- Antelman, K. (2004). Do open-access articles have a greater research impact? *College & Research Libraries*, 65(9), 372–382.
- Archambault, É., Vignola-Gagné, É., Côte, G., Larivière, V., & Gingras, Y. (2006). Benchmarking scientific output in the social sciences and humanities—The limits of existing databases. *Scientometrics*, 68(3), 329–342.
- Arunachalam, S., & Doss, M. J. (2000). Mapping international collaboration in science in Asia through coauthorship analysis. *Current Science*, 79(5), 621–628.
- Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., et al. (2004). *An international framework to promote access to data Science*, 303(5665), 1777–1778.
- Baeza-Yates, R., Castillo, C., & López, V. (2005). Characteristics of the Web of Spain. *Cybermetrics*, 9(1), paper 3. Retrieved July 12, 2007, from <http://www.cindoc.csic.es/cybermetrics/articles/v9i1p3.html>.
- Bailón-Moreno, R., Jurado-Alameda, E., Ruiz-Baños, R. R., & Courtial, J. P. (2005). Bibliometric laws—Empirical flaws of fit. *Scientometrics*, 63(2), 209–229.
- Bakkalbasi, N., Bauer, K., Glover, J., & Wang, L. (2006). Three options for citation tracking: Google Scholar, Scopus and Web of Science. *Biomedical Digital Libraries*, 3, 7–14. Retrieved July 31, 2007, from <http://www.bio-diglib.com/content/3/1/7>.
- Balconi, M., Breschi, S., & Lissoni, F. (2004). Networks of inventors and the role of academia: An exploration of Italian patent data. *Research Policy*, 33(1), 127–145.

- Banks, M. G. (2006). An extension of the Hirsch index—Indexing scientific topics and compounds. *Scientometrics*, 69(1), 161–168.
- Barabási, A. L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A*, 311(3–4), 590–614.
- Bar-Ilan, J. (1997). The 'mad cow disease', Usenet newsgroups and bibliometric laws. *Scientometrics*, 39(1), 29–35.
- Bar-Ilan, J. (2000a). Evaluating the stability of the search tools Hotbot and Snap: A case study. *Online Information Review*, 24(6), 439–450.
- Bar-Ilan, J. (2000b). The Web as an information source on informetrics? A content analysis. *Journal of the American Society for Information Science and Technology*, 51(5), 432–443.
- Bar-Ilan, J. (2000c). Results of an extensive search for "S&T indicators" on the Web—A content analysis. *Scientometrics*, 49(2), 257–277.
- Bar-Ilan, J. (2001). Data collection methods on the Web for informetric purposes—A review and analysis. *Scientometrics*, 50(1), 7–32.
- Bar-Ilan, J. (2002a). How much information do search engines disclose on the links to a web page? A longitudinal case study of the 'cybermetrics' home page. *Journal of Information Science*, 28(6), 455–466.
- Bar-Ilan, J. (2002b). Methods for measuring search engine performance over time. *Journal of the American Society for Information Science and Technology*, 53(4), 308–319.
- Bar-Ilan, J. (2004a). The use of Web search engines in information science research. In B. Cronin (Ed.), *Annual review of information science and technology* (pp. 231–288). Medford, NJ: Information Today.
- Bar-Ilan, J. (2004b). A microscopic link analysis of academic institutions within a country—The case of Israel. *Scientometrics*, 59(3), 391–403.
- Bar-Ilan, J. (2004c). Self-linking and self-linked rates of academic institutions on the Web. *Scientometrics*, 59(1), 29–41.
- Bar-Ilan, J. (2005a). Expectations versus reality—Search engine features needed for Web research at mid 2005. *Cybermetrics*, 9(1), paper 2. Retrieved July 14, 2007, from <http://www.cindoc.csic.es/cybermetrics/articles/v9i1p2.html>.
- Bar-Ilan, J. (2005). What do we know about links and linking? A framework for studying links in academic environments. *Information Processing and Management*, 41(4), 966–973.
- Bar-Ilan, J. (2006a). An ego-centric citation analysis of the works of Michael O. Rabin based on multiple citation indexes. *Information Processing and Management*, 42(6), 1553–1566.
- Bar-Ilan, J. (2006b). H-Index for Price medalists revisited. *ISSI Newsletter*, 2(1), 3–5.
- Bar-Ilan, J., & Peritz, B. C. (2002). Informetric theories and methods for exploring the Internet—An analytical survey of recent research literature. *Library Trends*, 50(3), 371–392.
- Bar-Ilan, J., & Peritz, B. C. (2004). Evolution, continuity, and disappearance of documents on a specific topic on the Web—A longitudinal study of "informetrics". *Journal of the American Society for Information Science and Technology*, 55(11), 980–990.
- Barjak, F. (2006a). The role of the Internet in informal scholarly communication. *Journal of the American Society for Information Science and Technology*, 57(10), 1350–1367.
- Barjak, F. (2006b). Research productivity in the Internet era. *Scientometrics*, 68(3), 343–360.
- Barrat, A., Barthélémy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(11), 3747–3752.
- Bauer, K., & Bakkalbasi, N. (2005). An examination of citation counts in a new scholarly communication environment. *D-Lib Magazine*, 11(9). Retrieved July 31, 2007, from <http://www.dlib.org/dlib/september05/bauer/09bauer.html>
- Basu, A., & Kumar, B. S. V. (2000). International collaboration in Indian scientific papers. *Scientometrics*, 48(3), 381–402.
- Barrett, C. B., Oliá, A., & Von Bailey, D. (2000). Subdiscipline-specific journal rankings—Whither applied economics? *Applied Economics*, 32(2), 239–252.
- Batista, P. D., Campiteli, M. G., Kinouchi, O., & Martinez, A. S. (2006). Is it possible to compare researchers with different scientific interests? *Scientometrics*, 68(1), 179–189.
- Baumgartner, H., & Pieters, R. (2003). The structural influence of marketing journals—A citation analysis of the discipline and its subareas over time. *Journal of marketing*, 67(2), 123–139.
- Beaver, D. D. (2001). Reflections on scientific collaboration, (and its study)—Past, present, and future. *Scientometrics*, 52(3), 365–377.
- Beaver, D. D. (2004). Does collaborative research have greater epistemic authority? *Scientometrics*, 60(3), 399–408.
- Behrens, T. R., & Gray, D. O. (2001). Unintended consequences of cooperative research—Impact of industry sponsorship on climate for academic freedom and other graduate student outcome. *Research Policy*, 30(2), 179–199.
- Beller, F. K. (2000). The future of the German language in science. *Gynakologisch-Geburtshilfliche Rundschau*, 40(1), 50–54.
- Bensman, S. J. (2001). Urquhart's and Garfield's laws—The British controversy over their validity. *Journal of the American Society for Information Science and Technology*, 52(9), 714–724.
- Bensman, S. J. (2004). Pearson's r and author cocitation analysis—A commentary on the controversy. *Journal of the American Society for Information Science and Technology*, 55(10), 935.
- Bensman, S. J. (2005). Urquhart and probability—The transition from librarianship to library and information science. *Journal of the American Society for Information Science and Technology*, 56(2), 189–214.
- Bhattacharya, S., Kretschmer, H., & Meyer, M. (2003). Characterizing intellectual spaces between science and technology. *Scientometrics*, 58(2), 369–390.
- Björneborn, L. (2001). Small-world linkage and co-linkage. In *Proceedings of the 12th ACM conference on hypertext and hypermedia* (pp. 133–134). ACM.
- Björneborn, L. (2004). *Small-world link structures across an academic web space: A library and information science approach*. PhD dissertation, Copenhagen: Department of Information Studies, Royal School of Library and Information Science. xxxvi, 399 pp. ISBN: 87-7415-276-9. Retrieved July 14, 2007, from <http://vip.db.dk/lb/phd/phd-thesis.pdf>.
- Björneborn, L., & Ingwersen, P. (2001). Perspectives of webometrics. *Scientometrics*, 50(1), 65–82.

- Björneborn, L., & Ingwersen, P. (2004). Toward a basic framework for webometrics. *Journal of the American Society for Information Science and Technology*, 55(14), 1216–1227.
- Blanchard, D., Erlich, J., Montgomery, G. H., & Bovbjerg, D. H. (2002). Read all about it—The over-representation of breast cancer in popular magazines. *Preventive Medicine*, 35(4), 343–348.
- Bloch, S., & Walter, G. (2001). The impact factor—Time for change. *Australian and New Zealand Journal of Psychiatry*, 35(5), 563–568.
- Boldt, J., Haisch, G., & Maleck, W. H. (2000). Changes in the impact factor of anesthesia/critical care journals within the past 10 years. *Acta Anaesthesiologica Scandinavica*, 44(7), 842–849.
- Bollacker, K. D., Lawrence, S., & Giles, C. L. (2000). Discovering relevant scientific literature on the Web. *IEEE Intelligent Systems and their Applications*, 15(2), 42–47.
- Bollen, J., Rodriguez, M. A., & van de Sompel, H. (2006). Journal status. *Scientometrics*, 69(3), 669–687.
- Bonaccorsi, A., & Daraio, C. (2003). Age effects in scientific productivity—The case of the Italian National Research Council (CNR). *Scientometrics*, 58(1), 49–90.
- Bonitz, M., & Scharnhorst, A. (2001). Competition in science and the Matthew core journals. *Scientometrics*, 51(1), 37–54.
- Bonnevie, E. (2003). A multifaceted portrait of a library and information science journal—The case of the Journal of Information Science. *Journal of Information Science*, 29(1), 11–23.
- Bookstein, A., Moed, H., & Yitzahki, M. (2006). Measures of international collaboration in scientific literature. Part II. *Information Processing and Management*, 42(9), 1422–1427.
- Bordons, M., Fernández, M. T., & Gómez, I. (2002). Advantages and limitations in the use of impact factor measures for the assessment of research performance in a peripheral country. *Scientometrics*, 53(2), 195–206.
- Borgman, C. L. (2000). Digital libraries and the continuum of scholarly communication. *Journal of Documentation*, 56(4), 412–430.
- Borgman, C. L., & Furner, J. (2002). Scholarly communication and bibliometrics. In B. Cronin (Ed.), *Annual review of information science and technology* (pp. 3–72). Medford, NJ: Information Today.
- Börner, K., Chen, C. M., & Boyack, K. W. (2003). *Visualizing knowledge domains*. Annual review of information science and technology Medford, NJ: Information Today, pp. 179–255
- Börner, K., Maru, J. T., & Goldstone, R. L. (2004). The simultaneous evolution of author and paper networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl. 1), 5266–5273.
- Bornmann, L., & Daniel, H. D. (2005a). Does the *h*-index for ranking of scientists really work? *Scientometrics*, 65(3), 391–392.
- Bornmann, L., & Daniel, H. D. (2005b). Selection of research fellowship recipients by committee peer review. Reliability, fairness and predictive validity of Board of Trustees' decisions. *Scientometrics*, 63(2), 297–320.
- Bornmann, L., & Daniel, H. D. (2005c). Committee peer review at an international research foundation—Predictive validity and fairness of selection decisions on post-graduate fellowship applications. *Research Evaluation*, 14(1), 15–20.
- Bornmann, L., & Daniel, H. D. (2006). Selecting scientific excellence through committee peer review—A citation analysis of publications previously published to approval or rejection of post-doctoral research fellowship applicants. *Scientometrics*, 68(3), 427–440.
- Boyack, K. W. (2004). Mapping knowledge domains—Characterizing PNAS. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl. 1), 5192–5199.
- Boyack, K. W., & Börner, K. (2003). Indicator-assisted evaluation and funding of research - Visualizing the influence of grants on the number and citation counts of research papers. *Journal of the American Society for Information and Technology*, 54(5), 447–461.
- Boyack, K. W., Klavans, R., & Börner, K. (2005). Mapping the backbone of science. *Scientometrics*, 64(3), 351–374.
- Boyack, K. W., Wylie, B. N., & Davidson, G. S. (2002). Domain visualization using VxInsight (R) for science and technology management. *Journal of the American Society for Information Science and Technology*, 53(9), 764–774.
- Bozeman, B., & Corley, E. (2004). Scientists' collaboration strategies—Implications for scientific and technical human capital. *Research Policy*, 33(4), 599–616.
- Braun, T. (Ed.). (2006). *Evaluations of individual scientists and research institutions*. *Scientometrics guidebook series 1*. Budapest: Akadémia Kiadó.
- Braun, T., & Dióspatonyi, I. (2005a). The journal gatekeepers of major publishing houses of core science journals. *Scientometrics*, 62(3), 297–319.
- Braun, T., & Dióspatonyi, I. (2005b). World flash on basic research—The counting of core journal gatekeepers as science indicators really counts. The scientific scope of action and strength of nations. *Scientometrics*, 64(2), 113–120.
- Braun, T., Glänzel, W., & Schubert, A. (2000). How balanced is the Science Citation Index's journal coverage? A preliminary overview of macrolevel statistical data. In B. Cronin & H. Atkins (Eds.), *The Web of knowledge—A Festschrift in honor of Eugene Garfield*, *ASIS Monograph Series* (pp. 251–277). Medford, NJ: Information Today.
- Braun, T., Glänzel, W., & Schubert, A. (2001). Publication and cooperation patterns of the authors of neuroscience journals. *Scientometrics*, 51(3), 499–510.
- Braun, T., Glänzel, W., & Schubert, A. (2005). A Hirsch-type index for journals. *The Scientist*, 19(22), 8.
- Braun, T., Glänzel, W., & Schubert, A. (2006). A Hirsch-type index for journals. *Scientometrics*, 69(1), 169–173.
- Braun, T., Schubert, A. P., & Kostoff, R. N. (2000). Growth and trends of fullerene research as reflected in its journal literature. *Chemical Reviews*, 100(1), 23–37.
- Braun, T., Szabadi-Peresztegi, Z., & Kovacs-Németh, É. (2003). No-bells for ambiguous lists of ranked Nobelists as science indicators of national merit in physics, chemistry and medicine, 1901–2001. *Scientometrics*, 56(1), 3–28.
- Brin, S., & Page, L. (1998, 1998). The anatomy of a large-scale hypertextual Web search engine. Proceedings of the 7th international World Wide Web conference. *Computer Networks and ISDN Systems*, 30, 107–117.
- Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., et al. (2000, May). Graph structure in the Web. Proceedings of the 9th international World Wide Web conference. *Computer Networks and ISDN Systems*, 33, 309–320. Retrieved July 31, 2007, from <http://www9.org/w9cdrom/160/160.html>

- Brody, T., Harnad, S., & Carr, L. (2006). Earlier web usage statistics as predictors of later citation impact. *Journal of the American Society for Information Science and Technology*, 57(8), 1060–1072.
- Brown, C. (2003). The role of electronic preprints in chemical communication—Analysis of citation, usage, and acceptance in the journal literature. *Journal of the American Society for Information Science and Technology*, 54(5), 362–371.
- Brown, C. (2001). The e-evolution of preprints in the scholarly communication of physicists and astronomers. *Journal of the American Society for Information Science and Technology*, 52(3), 187–200.
- Buela-Casal, G. (2002). Evaluating quality of articles and scientific journals. Proposal of weighted impact factor and a quality index? *Psichothema*, 15(1), 23–35.
- Buela-Casal, G., Carretero-Dios, H., & de los Santos-Riog, R. (2002). Comparative study of the Psychology journals with impact factor written in Spanish. *Psichotema*, 1494, 837–852.
- Buchanan, R. A. (2006). Accuracy of cited references—The role of citation databases. *College & Research Libraries*, 67(4), 292–303.
- Burrell, Q. L. (2001). Stochastic modelling of the first-citation distribution. *Scientometrics*, 52(1), 3–12.
- Burrell, Q. L. (2002). The *n*th-citation distribution and obsolescence. *Scientometrics*, 53(3), 309–323.
- Burrell, Q. L. (2005a). Symmetry and other transformation features of Lorenz/Leimkuhler representations of informetric data. *Information Processing and Management*, 41(6), 1317–1329.
- Burrell, Q. L. (2005b). Measuring informetric similarity of concentration between different distributions—Two new approaches. *Journal of the American Society for Information Science and Technology*, 56(4), 704–714.
- Burrell, Q. L. (2005c). Are “sleeping beauties” to be expected? *Scientometrics*, 65(3), 381–389.
- Butler, L. (2002). A list of published papers is no measure of value—The present system rewards quantity, not quality—But hasty changes could be as bad. *Nature*, 419, 877.
- Butler, L. (2003). Explaining Australia’s increased share of ISI publications—The effects of a funding formula based on publication counts. *Research Policy*, 32(1), 143–155.
- Butler, L., & Visser, M. S. (2006). Extending citation analysis to non-source items. *Scientometrics*, 66(2), 327–343.
- Cahlik, T. (2000). Comparison of the maps of science. *Scientometrics*, 49(3), 373–387.
- Callahan, M., Wears, R. L., & Weber, E. (2002). Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. *Journal of the American Medical Association*, 287(21), 2847–2850.
- Chan, L. C. Y., Jin, B., Rousseau, R., Vaughan, L., & Yu, Y. (2002–2003). Newspaper coverage of SARS—A comparison among Canada, Hong Kong, Mainland China and Western Europe. *Cybermetrics*, 6/7(1), paper 1. Retrieved July 23, 2007, from <http://www.cindoc.csic.es/cybermetrics/articles/v6i1p1.html>.
- Chen, C. M. (2003). *Mapping scientific frontiers: The quest for knowledge visualization*. London: Springer-Verlag.
- Chen, C. M. (2004a). *Information visualization: Beyond the horizon* (2nd ed.). London: Springer-Verlag.
- Chen, C. M. (2004b). Searching for intellectual turning points—Progressive knowledge domain visualization. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 5303–5310.
- Chen, C. M. (2006). CiteSpace II—Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, 57(3), 359–377.
- Chen, C. M., Cribbin, T., Macredie, R., & Morar, S. (2002). Visualizing and tracking the growth of competing paradigms—Two case studies. *Journal of the American Society for Information Science and Technology*, 53(8), 678–689.
- Chen, C. M., & Kuljis, J. (2003). The rising landscape—A visual exploration of superstring revolutions in physics. *Journal of the American Society for Information Science and Technology*, 56(4), 704–714.
- Chen, C. M., & Paul, R. J. (2001). Visualizing a knowledge domain’s intellectual structure. *IEEE Computer*, 34(3), 65–71.
- Chen, C. M., Paul, R. J., & O’Keefe, B. (2001). Fitting the jigsaw of citation—Information visualization in domain analysis. *Journal of the American Society for Information Science and Technology*, 52(4), 315–330.
- Chompalov, I., Genuth, J., & Shrum, W. (2002). The organization of scientific collaborations. *Research Policy*, 31(5), 749–767.
- Chu, H. (2005). Taxonomy of inlinked Web entities—What does it imply for webometric research? *Library & Information Science Research*, 27(1), 8–27.
- Coats, A. J. S. (2005). Top of the Charts—Download versus citations in the International Journal of Cardiology. *International Journal of Cardiology*, 105(2), 123–125.
- Coccia, M. (2004). New models for measuring the R&D performance and identifying the productivity of public research institutes. *R&D Management*, 34(3), 267–280.
- Coelho, P. M. Z., Antunes, C. M. F., Costa, H. M. A., Kroon, E. G., Sousa Lima, M. C., & Linardi, P. M. (2003). The use and misuse of the “impact factor” as a parameter for evaluation of scientific publication quality—A proposal to rationalize its application. *Brazilian Journal of Medical and Biological Research*, 36(12), 1605–1612.
- Colquhoun, D. (2003). Challenging the tyranny of impact factors. *Nature*, 423, 479.
- Cothey, V. (2004). Web-crawling reliability. *Journal of the American Society for Information Science and Technology*, 55(14), 1228–1238.
- Cothey, V., Aguillo, I., & Arroyo, N. (2006). Operationalising “Websites”—Lexically, semantically or topologically? *Cybermetrics*, 10(1), paper 3. Retrieved July 12, 2007, from <http://www.cindoc.csic.es/cybermetrics/articles/v10i1p3.html>.
- Cronin, B. (2000). Semiotics and evaluative bibliometrics. *Journal of Documentation*, 56(4), 440–453.
- Cronin, B. (2001a). 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(7), 558–569.
- Cronin, B. (2001b). Bibliometrics and beyond—Some thoughts on web-based citation analysis. *Journal of Information Science*, 27(1), 1–7.
- Cronin, B., & Atkins, H. (Eds.). (2000). *The Web of knowledge—A Festschrift in honor of Eugene Garfield*. ASIS Monograph Series. Medford, NJ: Information Today.

- Cronin, B., & Meho, L. (2006). Using the *h*-index to rank influential information scientists. *Journal of the American Society for Information Science and Technology*, 57(9), 1275–1278.
- Cronin, B., & Shaw, D. (2002a). Identity-creators and image-makers—Using citation analysis and thick description to put authors in their place. *Scientometrics*, 54(1), 34–49.
- Cronin, B., & Shaw, D. (2002b). Banking (on) different forms of symbolic capital. *Journal of the American Society for Information Science and Technology*, 53(14), 1267–1270.
- Cronin, B., Shaw, D., & Le Barre, K. (2003). A cast of thousands—Coauthorship and subauthorship collaboration in the 20th century as manifested in the scholarly journal literature of psychology and philosophy. *Journal of the American Society for Information Science and Technology*, 54(9), 855–871.
- Cronin, B., Shaw, D., & Le Barre, K. (2004). Visible, less visible, and invisible work—Patterns of collaboration in 20th century chemistry. *Journal of the American Society for Information Science and Technology*, 55(2), 160–168.
- Curti, M., Pistotti, V., Gabutti, G., & Klerys, C. (2001). Impact factor and electronic versions of biomedical scientific journals. *Haematologica*, 86(10), 1015–1020.
- Danell, R. (2000). Stratification among journals in management research—A bibliometric study of interaction between European and American journals. *Scientometrics*, 49(1), 23–28.
- Danell, R., & Persson, O. (2003). Regional R&D activities and interactions in the Swedish Triple Helix. *Scientometrics*, 58(2), 205–218.
- Davenport, E., & Cronin, B. (2000). The citation network as a prototype for representing trust in virtual environments. In B. Cronin & H. Atkins (Eds.), *The Web of knowledge—A Festschrift in honor of Eugene Garfield, ASIS Monograph Series* (pp. 517–534). Medford, NJ: Information Today.
- Davis, P. (2002). Where to spend our E-journal money—Defining a university library's core collection through citation analysis. *PORTAL—Libraries and the Academy*, 2(1), 155–166.
- Davis, P. M., & Cohen, S. A. (2001). The effect of the Web on undergraduate citation behavior 1996–1999. *Journal of the American Society for Information Science and Technology*, 52(4), 319–324.
- Davis, R. M., & Müllner, M. (2002). Editorial independence at medical journals owned by professional associations—A survey of editors. *Science and Engineering Ethics*, 8(4), 513–528.
- Debackere, K., & Veugelers, R. (2005). The role of academic technology transfer organizations in improving industry science links. *Research Policy*, 34(3), 321–342.
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society of Information Science*, 41(6), 391–407.
- de Nooy, W., Mrvar, A., & Batagelj, V. (2005). *Exploratory social network analysis with Pajek*. Cambridge: Cambridge University Press.
- Dietz, J. S., Chompalov, I., Bozeman, B., Lane, E. O., & Park, J. (2000). Using the curriculum vita to study the career paths of scientists and engineers—An exploratory assessment. *Scientometrics*, 49(3), 412–442.
- Ding, Y., Chowdhury, G. G., & Foo, S. (2000). Journal as markers of intellectual space—Journal co-citation analysis of information Retrieval area, 1987–1997. *Scientometrics*, 47(1), 55–73.
- Ding, Y., Chowdhury, G. G., & Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing and Management*, 37(6), 817–842.
- Ding, Y., Chowdhury, G. G., Foo, S., & Qian, W. (2000). Bibliometric Information Retrieval System (BIRS)—A web search interface utilizing bibliometric research results. *Journal of the American Society for Information Science and Technology*, 51(13), 1190–1204.
- Dodgson, M., & Hinze, S. (2000). Indicators used to measure the innovation process—Defects and possible remedies. *Research Evaluation*, 8(2), 101–114.
- Donohue, J. M., & Fox, J. B. (2000). A multi-method evaluation of journals in the decision and management sciences by US academics. *Omega*, 28(1), 17–36.
- Dore, J. C., & Ojasoo, T. (2001). How to analyze publication time trends by correspondence factor analysis—Analysis of publications by 48 countries in 18 disciplines over 12 years. *Journal of the American Society for Information Science and Technology*, 52(9), 763–769.
- DuBois, F. L., & Reeb, D. (2000). Ranking the international business journals. *Journal of International Business Studies*, 31(4), 689–704.
- Durrant, R., Wakefield, M., McLeod, K., Clegg-Smith, K., & Chapman, S. (2003). Tobacco in the news—An analysis of newspaper coverage of tobacco issues in Australia, 2001. *Tobacco Control*, 12(Supplement 2), 75–81.
- Duy, J., & Vaughan, L. (2006). Can electronic journal usage data replace citation data as a measure of journal use? An empirical examination. *Journal of Academic Librarianship*, 32(5), 512–517.
- Egghe, L. (2000a). A heuristic study of the first-citation distribution. *Scientometrics*, 48(3), 345–359.
- Egghe, L. (2000b). The distribution of *N*-grams. *Scientometrics*, 47(2), 237–252.
- Egghe, L. (2001). New informetric aspects of the Internet—Some reflections—Many problems. *Journal of Information Science*, 26(5), 329–335.
- Egghe, L. (2002). Construction of concentration measures for General Lorenz curves using Riemann–Stieltjes integrals. *Mathematical and Computer Modelling*, 35(9/10), 1149–1163.
- Egghe, L. (2003). Type/Token-Taken informetrics. *Journal of the American Society for Information Science and Technology*, 54(7), 603–610.
- Egghe, L. (2004a). Positive reinforcement and 3-dimensional informetrics. *Scientometrics*, 60(3), 497–509.
- Egghe, L. (2004b). The source-item coverage of the Lotka function. *Scientometrics*, 61(1), 103–115.
- Egghe, L. (2005a). Expansion of the field of informetrics: Origins and consequences. *Information Processing and Management*, 41(6), 1311–1316.
- Egghe, L. (2005b). *Power laws in the information production process: Lotkaian informetrics*. Amsterdam: Academic Press.
- Egghe, L. (2005c). Zipfian and Lotkaian continuous concentration theory. *Journal of the American Society for Information Science and Technology*, 56(9), 935–945.

- Egghe, L. (2005d). The power of power laws and an interpretation of Lotkaian informetric systems as self-similar fractals. *Journal of the American Society for Information Science and Technology*, 56(7), 669–675.
- Egghe, L. (2006). Theory and practise of the *g*-index. *Scientometrics*, 69(1), 131–152.
- Egghe, L., & Rao, I. K. R. (2001). Theory of first-citation distributions and applications. *Mathematical and Computer Modelling*, 34(1–2), 84–90.
- Egghe, L., & Rao, I. K. R. (2002). Duality revisited: Construction of fractional frequency distributions based on two dual Lotka laws. *Journal of the American Society for Information Science and Technology*, 53(10), 789–801.
- Egghe, L., & Rousseau, R. (2000a). The influence of publication delays on the observed aging distribution of scientific literature. *Journal of the American Society for Information Science*, 51(2), 158–165.
- Egghe, L., & Rousseau, R. (2000b). Aging, obsolescence, impact, growth, and utilization—Definitions and relations. *Journal of the American Society for Information Science*, 51(11), 1004–1017.
- Egghe, L., & Rousseau, R. (2001). Symmetric and asymmetric theory of relative concentration and applications. *Scientometrics*, 52(2), 261–290.
- Egghe, L., & Rousseau, R. (2002a). Co-citation, bibliographic coupling and a characterization of lattice citation networks. *Scientometrics*, 55(3), 349–361.
- Egghe, L., & Rousseau, R. (2002b). A proposal to define a core of a scientific subject—A definition using concentration and fuzzy sets. *Scientometrics*, 54(1), 51–62.
- Egghe, L., & Rousseau, R. (2002c). A general framework for relative impact indicators. *Canadian Journal of Information and Library Science*, 27(1), 29–48.
- Egghe, L., & Rousseau, R. (2002d). How to measure own-group preference? A novel approach to a sociometric problem. *Scientometrics*, 59(2), 233–252.
- Egghe, L., & Rousseau, R. (2006a). Classical retrieval and overlap measures satisfy the requirements for rankings based on a Lorenz curve. *Information Processing and Management*, 42(1), 106–120.
- Egghe, L., & Rousseau, R. (2006b). An informetric model for the Hirsch-index. *Scientometrics*, 69(1), 1221–1219.
- Egghe, L., Rousseau, R., & Van Hooydonk, G. (2000). Methods for accrediting publications to authors or countries: Consequences for evaluation studies. *Journal of the American Society for Information Science*, 51(2), 145–157.
- Eysenbach, G. (2005). Citation advantage of open access articles. *PLoS Biology*, 4(5), 692–698.
- Essential Science Indicators. (2005). *Online help*. Retrieved July 10, 2007, from <http://esi.isiknowledge.com/help/h.datres.htm>.
- Etzkowitz, H., & Leydesdorff, L. (2000). The dynamics of innovation—From National Systems and “Mode 2” to a Triple Helix of university–industry–government relations. *Research Policy*, 29(2), 109–123.
- Fang, Y., & Rousseau, R. (2001). Lattices in citation networks—An investigation into the structure of citation graph. *Scientometrics*, 50(2), 273–287.
- Farkas, I., Derenyi, I., Jeong, H., Néda, Z., Oltvai, Z. N., Ravasz, E., et al. (2002). Networks in life—Scaling properties and eigenvalue spectra. *Physica A*, 314(1–4), 25–34.
- Fassoulaki, A., Paraskeva, A., Papilas, K., & Karabinis, G. (2000). Self-citations in six anaesthesia journals and their significance in determining the impact factor. *British Journal of Anaesthesia*, 84(2), 266–269.
- Fassoulaki, A., Papilas, K., Paraskeva, A., & Patris, N. (2002). Impact factor bias and proposed adjustments for its determination. *Acta Anaesthesiologica Scandinavica*, 46(7), 902–905.
- Fenton, J. E., Brazier, H., De Souza, A., Hughes, J. P., & McShane, D. P. (2000). The accuracy of citation and quotation in otolaryngology/head and neck surgery journals. *Clinical Otolaryngology and Allied Sciences*, 25(1), 40–44.
- Forgionne, G. A., & Kohli, R. (2001). A multiple criteria assessment of decision technology system journal quality. *Information & Management*, 38(7), 421–435.
- 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.
- Frandsen, T. F., Rousseau, R., & Rowlands, I. (2006). Diffusion factors. *Journal of Documentation*, 62(1), 58–72.
- Frank, M. (2003). Impact factors—Arbiter of excellence? *Journal of the Medical Library Association*, 91(1), 4–6.
- Gadd, E., Oppenheim, C., & Proberts, S. (2003a). RoMEO studies. 1. The impact of copyright ownership on academic author self-archiving. *Journal of Documentation*, 59(3), 243–277.
- Gadd, E., Oppenheim, C., & Proberts, S. (2003b). RoMEO studies 2—How academics want to protect their open-access research papers? *Journal of Information Science*, 29(5), 333–356.
- Gadd, E., Oppenheim, C., & Proberts, S. (2003c). RoMEO studies. 3. How academics expect to use open access research papers. *Journal of Librarianship and Information Science*, 35(3), 171–187.
- Gardner, S., & Eng, S. (2005). Gaga over Google? Scholar in the Social Sciences. *Library Hi Tech News*, 22(8), 42–45.
- Garfield, E. (no date). *Free full-text access to Paul Wouters’ “Citation Culture”*. Retrieved June 10, 2007, from <http://www.asis.org/Chapters/europe/announcements/citation.htm>.
- Garfield, E. (2000). Use of Journal Citation Reports and Journal Performance Indicators in measuring short and long term journal impact. *Croatian Medical Journal*, 41(4), 368–374.
- Garfield, E. (2001). Impact factors, and why they won’t go away. *Nature*, 411, 522.
- Garfield, E. (2004). Historiographic mapping of knowledge domains literature. *Journal of Information Science*, 30(2), 119–145.
- Garfield, E. (2006). The history and meaning of the journal impact factor. *Journal of the American Medical Association*, 295(1), 90–93.
- Garfield, E., Pudovkin, A. I., & Istomin, V. S. (2003). Why do we need algorithmic historiography? *Journal of the American Society for Information Science and Technology*, 54(5), 400–412.
- Gaughan, M., & Bozeman, B. (2002). Using curriculum vitae to compare some impacts of NSF research grants with research center funding. *Research Evaluation*, 11(1), 17–26.



- Giles, C. L., & Councill, I. G. (2004). Who gets acknowledged: Measuring scientific contributions through automatic acknowledgment indexing. *Proceedings of the National Academy of Sciences of the United States of America*, 101(51), 17599–17604.
- Girvan, & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12), 7821–7826.
- Glänzel, W. (2001). National characteristics in international scientific co-authorship relations. *Scientometrics*, 51(1), 69–115.
- Glänzel, W. (2002). Coauthorship patterns and trends in the sciences (1980–1998)—A bibliometric study with implications for database indexing and search strategies. *Library Trends*, 50(3), 461–473.
- Glänzel, W. (2006). On the *h*-index—A mathematical approach to a new measure of publication activity and citation impact. *Scientometrics*, 67(2), 315–321.
- Glänzel, W., & Meyer, M. (2003). Patents cited in the scientific literature—An exploratory study of 'reverse' citation relations. *Scientometrics*, 58(2), 415–428.
- Glänzel, W., & Moed, H. F. (2002). Journal impact measures in bibliometric research. *Scientometrics*, 53(20), 171–193.
- Glänzel, W., & Persson, O. (2005). *H*-Index for prize medalists. *ISSI Newsletter*, 1(4), 15–18.
- Glänzel, W., Schlemmer, B., Schubert, A., & Thijs, B. (2006). Proceedings literature as additional data source for bibliometric analysis. *Scientometrics*, 68(3), 457–473.
- Glänzel, W., Schlemmer, B., & Thijs, B. (2003). Better late than never? On the chance to become highly cited only beyond the standard bibliometric time horizon. *Scientometrics*, 58(3), 571–586.
- Glänzel, W., & Schubert, A. (2001). Double effort=double impact? A critical view at international co-authorship in chemistry. *Scientometrics*, 50(2), 199–214.
- Glänzel, W., & Schubert, A. (2003). A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. *Scientometrics*, 56(3), 357–367.
- Glänzel, W., Schubert, A., & Braun, T. (2002). A relational charting approach to the world of basic research in twelve science fields at the end of the second millennium. *Scientometrics*, 55(3), 335–348.
- Glänzel, W., & Thijs, B. (2004). The influence of author self-citations on bibliometric macro indicators. *Scientometrics*, 59(3), 281–310.
- Glänzel, W., Thijs, B., & Schlemmer, B. (2004). A bibliometric approach to the role of author self-citations in scientific communication. *Scientometrics*, 59(1), 63–77.
- Glenisson, P., Glänzel, W., Janssens, F., & De Moor, B. (2005). Combining full text and bibliometric information in mapping scientific disciplines. *Information Processing and Management*, 41(6), 1548–1572.
- Glenisson, P., Glänzel, W., & Persson, O. (2005). Combining full-text analysis and bibliometric indicators. A pilot study. *Scientometrics*, 63(1), 163–180.
- Gmur, M. (2003). Co-citation analysis and the search for invisible colleges—A methodological evaluation. *Scientometrics*, 57(1), 27–57.
- Godin, B. (2003). The emergence of S&T indicators—Why did governments supplement statistics with indicators? *Research Policy*, 32(4), 679–691.
- Godin, B. (2005). *Measurement and statistics on science and technology: 1920 to the present*. London: Routledge.
- Godin, B. (2006a). On the origins of bibliometrics. *Scientometrics*, 68(1), 109–133.
- Godin, B. (2006b). Research and development—How the 'D' got into R&D. *Science and Public Policy*, 33(1), 59–78.
- Godin, B., & Gingras, Y. (2000a). What is scientific and technological culture and how is it measured? A multidimensional model. *Public Understanding of Science*, 9(1), 43–58.
- Godin, B., & Gingras, Y. (2000b). The place of universities in the system of knowledge production. *Research Policy*, 29(2), 273–278.
- Goldstein, M. L., Morris, S. A., & Yen, G. G. (2004). Problems with fitting to the power-law distribution. *European Physical Journal B*, 41(2), 255–258.
- Goldstein, M. L., Morris, S. A., & Yen, G. G. (2005). Group-based Yule model for bipartite author-paper networks. *Physical Review E*, 71(2), 6108-1–6108-6.
- Goodrum, A. A., McCain, K. W., Lawrence, S., & Giles, C. L. (2001). Scholarly publishing in the Internet age—A citation analysis of computer science literature. *Information Processing and Management*, 37(5), 661–675.
- Griggs, R. A., & Proctor, D. L. (2002). A citation analysis of who's who in introductory textbooks. *Teaching of Psychology*, 29(3), 203–206.
- Gutiérrez, J., & López-Nieva, P. (2001). Are international journals of human geography really international? *Progress in Human Geography*, 25(1), 53–69.
- Hall, B. H., Jaffe, A., & Trajtenberg, M. (2005). Market value and patent citations. *RAND Journal of Economics*, 36(1), 16–38.
- Harhoff, D., Scherer, F. M., & Vopel, K. (2003). Citations, family size, opposition and the value of patent rights. *Research Policy*, 32(8), 1343–1363.
- Harnad, S. (2001). The self-archiving initiative. *Nature*, 410, 1024–1025.
- Harnad, S., & Brody, T. (2004). Comparing the impact of Open Access (OA) vs. non-OA articles in the same journals. *D-Lib Magazine*, 10(6). Retrieved July 22, 2007, from <http://www.dlib.org/dlib/june04/harnad/06harnad.html>
- Harnad, S., Brody, T., Vallières, F., Carr, L., Hitchcock, S., Gingras, Y., et al. (2004). The access/impact problem and the green and gold roads to open access. *Serials Review*, 30(4), 310–314.
- Harnad, S., & Carr, L. (2000). Integrating, navigating, and analysing open Eprint archives through open citation linking (the OpCit project). *Current Science*, 79(5), 629–638.
- Harries, G., Wilkinson, D., Price, L., Fairclough, R., & Thelwall, M. (2004). Hyperlinks as a data source for science mapping. *Journal of Information Science*, 30(5), 436–477.
- Harter, S. P., & Park, T. K. (2000). Impact of prior electronic publication on manuscript consideration policies of scholarly journals. *Journal of the American Society for Information Science and Technology*, 51(10), 940–948.
- He, Y. L., & Hui, S. C. (2002). Mining a Web Citation Database for author co-citation analysis. *Information Processing and Management*, 38(4), 491–508.

- He, S., & Spink, A. (2002). A comparison of foreign authorship distribution in JASIST and the Journal of Documentation. *Journal of the American Society for Information Science and Technology*, 53(11), 953–959.
- Heimeriks, G., Hörlesberger, M., & van den Besselaar, P. (2003). Mapping communication and collaboration in heterogeneous research networks. *Scientometrics*, 58(2), 391–413.
- Heimeriks, G., & van den Besselaar, P. (2006). Analyzing hyperlinks networks—The meaning of hyperlink based indicators of knowledge production. *Cybermetrics*, 10(1), paper 1. Retrieved July 12, 2007, from <http://www.cindoc.csic.es/cybermetrics/articles/v10i1p1.html>.
- Henderson, J. (2005). Google scholar—A source for clinicians? *Canadian Medical Association Journal*, 172(12), 1549–1550.
- Herring, S. D. (2002). Use of electronic resources in scholarly electronic journals—A citation analysis. *College & Research Libraries*, 63(4), 334–340.
- Hicks, D., Breitzman, A., Sr., Hamilton, K., & Narin, F. (2000). Research excellence and patented innovation. *Science and Public Policy*, 27(5), 310–320.
- Hicks, D., Breitzman, T., Olivastro, D., & Hamilton, K. (2001). The changing composition of innovative activity in the US—A portrait based on patent analysis. *Research Policy*, 30(4), 681–703.
- Hirsch, J. E. (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(46), 16569–16572.
- Hitchcock, S., Bergmark, D., Brody, T., Gutteridge, C., Carr, L., Hall, W., et al. (2002). Open citation linking. The way forward. *D-Lib Magazine*, 8(10). Retrieved July 23, 2007, from <http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/october02/hitchcock/10hitchcock.html>
- Hitchcock, S., Carr, L., Jiao, Z., Bergmark, D., Hall, W., Lagoze, C., et al. (2000). Developing services for open eprint archives—Globalization, integration and the impact of links. In *Proceedings of the ACM international conference on digital libraries* (pp. 143–151).
- Ho, Y.-S. (2004). Citation review of Lagergren kinetic rate equation on adsorption reactions. *Scientometrics*, 59(1), 171–177.
- Holden, G., Rosenberg, G., & Barker, K. (2005). Tracing thought through time and space—A selective review of bibliometrics in social work. *Social Work in Health Care*, 41(3/4), 1–34.
- Holmes, A., & Oppenheim, C. (2001). Use of citation analysis to predict the outcome of the 2001 Research Assessment Exercise for Unit of Assessment (UoA) 61—Library and information management. *Information Research*, 6(2). Retrieved July 24, 2007, from <http://informationr.net/ir/6-2/paper103.html>
- Hoffman-Goetz, L., Shannon, C., & Clarke, J. N. (2003). Chronic disease coverage in Canadian aboriginal newspapers. *Journal of Health Communication*, 8(5), 475–488.
- Hood, W. W., & Wilson, C. S. (2001a). The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics*, 52(2), 291–314.
- Hood, W. W., & Wilson, C. S. (2001b). The scatter of documents over databases in different subject domains—How many databases are needed? *Journal of the American Society for Information Science and Technology*, 52(14), 1242–1254.
- Hood, W. W., & Wilson, C. S. (2003a). Informetric studies using databases—Opportunities and challenges. *Scientometrics*, 58(3), 587–608.
- Hood, W. W., & Wilson, C. S. (2003b). Overlap in bibliographic databases. *Journal of the American Society for Information Science and Technology*, 54(12), 1091–1103.
- Huang, Z., Chen, H., Chen, Z.-K., & Roco, M. C. (2004). International nanotechnology development in 2003—Country, institution, and technology field analysis based on USPTO patent database. *Journal of Nanoparticle Research*, 6(4), 325–354.
- Huang, Z., Chen, H., Yip, A., Ng, G., Guo, F., Chen, Z.-K., et al. (2003). Longitudinal patent analysis for nanoscale science and engineering—Country, institution and technology field. *Journal of Nanoparticle Research*, 5(3–4), 333–363.
- Huber, J. C. (2002). A new model that generates Lotka's Law. *Journal of the American Society for Information Science and Technology*, 53(3), 209–219.
- Huber, J. C., & Wagner-Döbler, W. (2001). Scientific production—A statistical analysis of authors in mathematical logic. *Scientometrics*, 50(2), 323–337.
- Hullmann, A., & Meyer, M. (2003). Publications and patents in nanotechnology—An overview of previous studies and the state of the art. *Scientometrics*, 58(3), 507–527.
- Hyland, K. (2003). Self-citation and self-reference—Credibility and promotion in academic publication. *Journal of the American Society for Information Science and Technology*, 54(3), 251–259.
- Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles. *English for Special Purposes*, 20(3), 207–226.
- Ingwersen, P. (1998). The calculation of Web Impact Factors. *Journal of Documentation*, 54(2), 236–243.
- Ingwersen, P., Larsen, B., Rousseau, R., & Russell, J. (2001). The publication-citation matrix and its derived quantities. *Chinese Science Bulletin*, 46(6), 524–528.
- Inonu, E. (2003). The influence of cultural factors on scientific production. *Scientometrics*, 56(1), 137–146.
- Ioannidis, P. A. (2005). Contradicted and initially stronger effects in highly cited clinical research. *JAMA*, 294(2), 218–228.
- Jacsó, P. (2001). A deficiency in the algorithm for calculating the impact factor of scholarly journals—The journal impact factor. *Cortex*, 37(4), 590–594.
- Jacsó, P. (2005a). Google scholar—The pros and the cons. *Online Information Review*, 29(2), 208–214.
- Jacsó, P. (2005b). As we may search—comparison of major features of web of science, Scopus and Google Scholar citation-based and citation-enhanced databases. *Current Science*, 89(9), 1537–1547.
- Jacsó, P. (2006). Deflated, inflated and phantom citation counts. *Online Information Review*, 30(3), 297–309.
- Janssen, M. A., Schoon, M. L., Ke, W. M., & Börner, K. (2006). Scholarly networks on resilience, vulnerability and adaptation within the human dimensions of global environmental change. *Global Environmental Change*, 16(3), 240–252.
- Janssens, F., Leta, J., Glänzel, W., & De Moor, B. (2006). Towards mapping library and information science. *Information Processing and Management*, 42(6), 1614–1642.

- Jarvelin, K., Ingwersen, P., & Niemi, T. (2000). A user-oriented interface for generalised informetric analysis based on applying advanced data modelling techniques. *Journal of Documentation*, 56(3), 250–278.
- Jin, B., & Rousseau, R. (2005). China's quantitative expansion phase: Exponential growth but low impact. In P. Ingwersen & B. Larsen (Eds.), *Proceedings of the ISSI 2005* (pp. 362–370). Stockholm, Sweden: Karolinska University Press.
- Jin, B., Zhang, J. G., Chen, D. O., & Zhu, X. Y. (2002). Development of the Chinese Scientometric Indicators (CSI). *Scientometrics*, 54(1), 145–154.
- Jones. (2003). Impact factors of forensic science and toxicology journals—What do the numbers really mean? *Forensic Science International*, 133(102), 1–8.
- Kaplan, N. R., & Nelson, M. L. (2000). Determining the publication impact of a digital library. *Journal of the American Society for Information Science and Technology*, 51(4), 324–329.
- Katerattanakul, P., Han, B., & Hong, S. (2003). Objective quality ranking of computing journal. *CACM*, 46(10), 111–114.
- Kelly, C. D., & Jennions, M. D. (2006). The *h*-index and career assessment by numbers. *Trends in Ecology and Evolution*, 21(4), 167–170.
- King, D. A. (2004). The scientific impact of nations. *Nature*, 430(6997), 311–316.
- King, J. T. (2000). How many neurosurgeons does it take to write a research article? Authorship proliferation in neurosurgical research. *Neurosurgery*, 47(2), 435–440.
- Kjaergard, L. L., & Gluud, C. (2002). Citation bias of hepato-biliary randomized clinical trials. *Journal of Clinical Epidemiology*, 55, 407–410.
- Kleijnen, J. P. C., & Van Groenendaal, W. (2000). Measuring the quality of publications—New methodology and case study. *Information Processing and Management*, 36(4), 551–570.
- Knothe, G. (2006). Comparative citation analysis of duplicate or highly related publications. *Journal of the American Society for Information Science and Technology*, 57(13), 1830–1839.
- Koehler, W. (2001). Information science as “Little Science”—The implications of a bibliometric analysis of the Journal of the American Society for Information Science. *Scientometrics*, 51(10), 117–132.
- Koehler, W. (2002). Web page change and persistence—A four-year longitudinal study. *Journal of the American Society for Information Science and Technology*, 53(2), 162–171.
- Kostoff, R. N. (2002). Citation analysis of research performer quality. *Scientometrics*, 53(1), 49–71.
- Kostoff, R. N., del Rió, J. A., Humenik, J. A., García, E. O., & Ramírez, A. M. (2001). Citation mining—Integrating text mining and bibliometrics for research user profiling. *Journal of the American Society for Information Science and Technology*, 52(13), 1148–1156.
- Kostoff, R. N., Braun, T., Schubert, A., Toothman, D. R., & Humenik, J. A. (2000). Fullerene data mining using bibliometrics and database tomography. *Journal of Chemical Information and Computer Sciences*, 40(1), 19–39.
- Kostoff, R. N., Green, K. A., Toothman, D. R., & Humenik, J. A. (2000). Database tomography applied to an aircraft science and technology investment strategy. *Journal of Aircraft*, 37(4), 727–730.
- Kostoff, R. N., Miles, D. L., & Eberhardt, H. J. (1995). System and method for database tomography. US Patent 5,44,0481.
- Kousha, K., & Thelwall, M. (2006). Motivations for URL citations to open access library and information science articles. *Scientometrics*, 68(3), 501–517.
- Kretschmer, H., & Aguillo, I. (2004). Visibility of collaboration on the Web. *Scientometrics*, 61(3), 405–426.
- Kurmis, A. P. (2003). Current concepts review—Understanding the limitations of the journal impact factor. *Journal of Bone & Joint Surgery—American Volume*, 85(12), 2449–2454.
- Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., Henneken, E., et al. (2005). The effect of use and access on citations. *Information Processing and Management*, 41(6), 1395–1402.
- Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., & Murray, S. S. (2005). Worldwide use and impact of the NASA astrophysics data system digital library. *Journal of the American Society for Information Science and Technology*, 56(1), 36–45.
- Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., Murray, S. S., et al. (2005). The bibliometric properties of article readership information. *Journal of the American Society for Information Science and Technology*, 56(2), 111–128.
- Landry, R., Amara, N., & Lamari, M. (2001). Utilization of social science research knowledge in Canada. *Research Policy*, 30(2), 333–349.
- Lafouge, T., & Michel, C. (2001). Links between information construction and information gain. Entropy and bibliometric distributions. *Journal of Information Science*, 27(1), 39–49.
- Lankhorst, G. J., & Franchignoni, F. (2001). The ‘impact factor’—An explanation and its application to rehabilitation journals. *Clinical Rehabilitation*, 15(2), 115–118.
- Landauer, T. K., Laham, D., & Derr, M. (2004). From paragraph to graph—Latent semantic analysis for information visualization. *Proceedings of the National Academy of Sciences of the United States of America*, 101(suppl. 1), 5214–5219.
- Lange, L. L. (2001). Citation counts of multi-authored papers—First-named authors and further authors. *Scientometrics*, 52(3), 457–470.
- Larson, R. (1996). Bibliometrics of the World Wide Web: An exploratory analysis of the intellectual structure of Cyberspace. In *Proceedings of the ASIS96* (pp. 71–78). Retrieved July 11, 2007, from <https://sherlock.sims.berkeley.edu/asis96/asis96.html>
- Lascar, C., & Mendelsohn, L. D. (2001). An analysis of journal use by structural biologists with applications for journal collection development decisions. *College & Research Libraries*, 62(5), 422–433.
- Lawrence, P. A. (2003). The politics of publication—Authors, reviewers and editors must act to protect the quality of research. *Nature*, 422, 259–261.
- Lawrence, S. (2001). Free online availability substantially increases a paper's impact. *Nature*, 411, 521.
- Lawrence, S., Pennock, D. M., Flake, G. W., Krovetz, R., Coetzee, F. M., Glover, E., et al. (2000). Persistence of web references in scientific research. *IEEE Computer*, 34(2), 26–31.
- Lee, K. P., Schotland, M., Bacchetti, P., & Bero, L. A. (2002). Association of journal quality indicators with methodological quality of clinical research articles. *Journal of the American Medical Association*, 287(21), 2805–2808.
- Leta, J., & Chaimovich, H. (2002). Recognition and international collaboration—The Brazilian case. *Scientometrics*, 53(2), 325–335.
- Lewison, G. (2001). The quantity and quality of female researchers—A bibliometric study of Iceland. *Scientometrics*, 52(1), 29–43.

- Lewison, G. (2002a). From biomedical research to health improvement. *Scientometrics*, 54(2), 179–192.
- Lewison, G. (2002b). Researchers' and users' perceptions of the relative standing of biomedical papers in different journals. *Scientometrics*, 53(2), 229–240.
- Lewison, G., & Paraje, G. (2004). The classification of biomedical journals by research level. *Scientometrics*, 60(2), 145–157.
- Lewison, G., Rippon, I., & Wooding, S. (2005). Tracking knowledge diffusion through citations. *Research Evaluation*, 14(1), 5–14.
- Leydesdorff, L. (2000). Is the European Union becoming a single publication system? *Scientometrics*, 47(2), 265–280.
- Leydesdorff, L. (2001). *The challenge of scientometrics: The development, measurement, and self-organization of scientific communications* (2nd ed.). Universal Publishers.
- Leydesdorff, L. (2001b). Indicators of innovation in a knowledge-based economy. *Cybermetrics*, 5, paper 2. Retrieved June 20, 2007, from <http://www.cindoc.csic.es/cybermetrics/articles/v5i1p2.html>.
- Leydesdorff, L. (2002a). Dynamic and evolutionary updates of classificatory schemes in scientific journal structures. *Journal of the American Society for Information Science and Technology*, 53(12), 987–994.
- Leydesdorff, L. (2002b). Indicators of structural change in the dynamics of science—Entropy statistics of the SCI Journal Citation Reports. *Scientometrics*, 53(1), 131–159.
- Leydesdorff, L. (2003a). Can networks of journal-journal citations be used as indicators of change in the social sciences? *Journal of Documentation*, 59(1), 84–104.
- Leydesdorff, L. (2003b). The mutual information of university–industry–government relations—An indicator of the Triple Helix dynamics. *Scientometrics*, 58(2), 445–467.
- Leydesdorff, L. (2004a). Clusters and maps of science journals based on bi-connected graphs in Journal Citation Reports. *Journal of Documentation*, 60(4), 371–424.
- Leydesdorff, L. (2004b). Top-down decomposition of the Journal Citation Report of the Social Science Citation Index—Graph- and factor-analytical approaches. *Scientometrics*, 60(2), 159–180.
- Leydesdorff, L. (2004c). The university–industry knowledge relationship—Analyzing patents and the science base of technologies. *Journal of the American Society for Information Science and Technology*, 55(11), 999–1001.
- Leydesdorff, L. (2005). Similarity measures, author cocitation analysis, and information theory. *Journal of the American Society for Information Science and Technology*, 56(7), 769–772.
- Leydesdorff, L. (2006). 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 and Technology*, 57(5), 601–613.
- Leydesdorff, L., & Curran, M. (2000). Mapping university–industry–government relations on the Internet—The construction of indicators for a knowledge-based economy. *Cybermetrics*, 4, paper 2. Retrieved June 20, 2007, from <http://www.cindoc.csic.es/cybermetrics/articles/v4i1p2.html>.
- Leydesdorff, L., & Bensman, S. (2006). Classification and powerlaws—The logarithmic transformation. *Journal of the American Society for Information Science and Technology*, 57(1), 1170–1186.
- Leydesdorff, L., & Heimeriks, G. (2001). The self-organization of the European information society—The case of “biotechnology”. *Journal of the American Society for Information Science and Technology*, 52(14), 1262–1274.
- Leydesdorff, L., & Hellsten, I. (2005). Metaphors and diaphors in science communication. *Science Communication*, 27(1), 64–99.
- Leydesdorff, L., & Meyer, M. (2003). The Triple Helix of university–industry–government relations. *Scientometrics*, 58(2), 191–203.
- Leydesdorff, L., & Vaughan, L. (2006). Co-occurrence matrices and their applications in information science—Extending ACA to the Web environment. *Journal of the American Society for Information Science and Technology*, 57(12), 1616–1628.
- Leydesdorff, L., & Zhou, P. (2005). Are the contributions of China and Korea upsetting the world system of science? *Scientometrics*, 63(3), 617–630.
- Li, X. M. (2003). A review of the development and application of the Web impact factor. *Online Information Review*, 27(6), 407–417.
- Li, X. M., Thelwall, M., Musgrove, P., & Wilkinson, D. (2003). The relationship between the WIFs or inlinks of Computer Science Departments in UK and their RAE ratings or research productivities in 2001. *Scientometrics*, 57(2), 239–255.
- Li, X. M., Thelwall, M., Wilkinson, D., & Musgrove, P. (2005). National and international university departmental Web site interlinking. Part 1. Validation of departmental link analysis. *Scientometrics*, 64(2), 151–185.
- Liang, L. (2005). R-sequences—Relative indicators for the rhythm of science. *Journal of the American Society for Information Science and Technology*, 56(10), 1045–1049.
- Liang, L. (2006). *H-Index sequence and h-index matrix: Constructions and applications*. *Scientometrics*, 69(1), 163–169.
- Liang, L. M., Kretschmer, H., Guo, Y. Z., & Beaver, D. D. (2001). Age structures of scientific collaboration in Chinese computer science. *Scientometrics*, 52(3), 471–486.
- Liang, L. M., Wu, Y. S., & Li, J. (2001). Selection of databases, indicators and models for evaluating research performance of Chinese universities. *Research Evaluation*, 10(2), 105–113.
- Lin, X., White, H. D., & Buzydlowski, J. (2003). Real-time author co-citation mapping for online searching. *Information Processing and Management*, 39(5), 689–706.
- Liu, N. C., & Cheng, Y. (2005a). The Academic Ranking of world universities—Methodologies and problems. *Higher Education in Europe*, 30(2), 127–136.
- Liu, N. C., & Cheng, Y. (2005b). Academic ranking of world universities using scientometrics—A comment to the “Fatal Attraction”. *Scientometrics*, 64(1), 101–109.
- Lok, C. K. W., Chan, M. T. V., & Martinson, I. M. (2001). Risk factors for citation errors in peer-reviewed nursing journals. *Journal of Advanced Nursing*, 34(2), 223–229.
- Losiewicz, P., Oard, D. W., & Kostoff, R. N. (2000). Textual data mining to support science and technology management. *Journal of Intelligent Information Systems*, 15(2), 99–119.

- Lu, D. W., & Katz, K. A. (2005). Declining use of the eponym “Reiter’s syndrome” in the medical literature, 1998–2003. *Journal of the American Academy of Dermatology*, 53(4), 720–723.
- Lundberg, J. D. (2003). The “omnipotent” Science Citation Index impact factor. *Medical Journal of Australia*, 178(6), 253–254.
- Mabe, M. A., & Amin, M. (2001). Growth dynamics of scholarly and scientific journals. *Scientometrics*, 51(1), 147–162.
- Mabe, M. A., & Amin, M. (2002). Dr Jekyll and Dr Hyde—Author-reader asymmetries in scholarly publishing. *Aslib Proceedings*, 54(3), 149–157.
- Mahlck, P., & Persson, O. (2000). Socio-bibliometric mapping of intra-departmental networks. *Scientometrics*, 49(1), 81–91.
- Man, J. P., Weinkauff, J. G., Tsang, M., & Sin, D. D. (2004). Why do some countries publish more than others? An international comparison of research funding. English proficiency and publication output in highly ranked general medical journals. *European Journal of Epidemiology*, 19(8), 811–817.
- Marušić, M., & Marušić, A. (2001). Good editorial practice—Editors as educators. *Croatian Medical Journal*, 42(2), 113–120.
- Marx, W., Schier, H., & Wanitschek, M. (2001). Citation analysis using online databases—Feasibilities and shortcomings. *Scientometrics*, 52(1), 59–82.
- Maurseth, P. B., & Verspagen, B. (2002). Knowledge spillovers in Europe—A patent citations analysis. *Scandinavian Journal of Economics*, 104(4), 531–545.
- Mayr, P., & Tosques, F. (2005). Webometrische Analysen mit Hilfe der Google Web APIs. *Information Wissenschaft und Praxis*, 56(1), 41–48.
- Mayr, P., & Umstätter, W. (2007). Why is a new Journal of Informetrics needed? *Cybermetrics*, 11(1), paper 1.
- McCain, K. W. (2000). Sharing digitized research-related information on the World Wide Web. *Journal of the American Society for Information Science and Technology*, 51(14), 1321–1327.
- McKibbin, K. A., Wilczynski, N. L., & Haynes, R. B. (2004). What do evidence-based secondary journals tell us about the publication of clinically important articles in primary healthcare journals? *BMC Medicine*, 2, 33–46.
- McMillan, G. S., & Hamilton, R. D., III. (2000). Using bibliometrics to measure firm knowledge—An analysis of the US pharmaceutical industry. *Technology Analysis and Strategic Management*, 12(4), 467–475.
- McMillan, G. S., Narin, F., & Deeds, D. L. (2000). An analysis of the critical role of public science in innovation—The case of biotechnology. *Research Policy*, 29(1), 1–8.
- Meadows, J. (2005). A practical Line in bibliometrics. *Interlending & Document Supply*, 33(2), 90–94.
- Meho, L. I., & Sonnenwald, D. H. (2000). Citation ranking versus peer evaluation of senior faculty research performance—A case study of Kurdish scholarship. *Journal of the American Society for Information Science and Technology*, 51(2), 123–138.
- Meho, L. I., & Spurgin, K. M. (2005). Ranking the research productivity of library and information science faculty and schools—An evaluation of data sources and research methods. *Journal of the American Society for Information Science and Technology*, 56(12), 1314–1331.
- Melin, G. (2000). Pragmatism and self-organization—Research collaboration on the individual level. *Research Policy*, 29(1), 31–40.
- Mettrop, W., & Nieuwenhuysen, P. (2001). Internet search engines—Fluctuations in document accessibility. *Journal of Documentation*, 57(5), 623–651.
- Meyer, M. (2000a). What is special about patent citations? Differences between scientific and patent citations. *Scientometrics*, 49(1), 93–123.
- Meyer, M. (2000b). Does science push technology? Patents citing scientific literature. *Research Policy*, 29(3), 409–434.
- Meyer, M. (2000c). Patent citations in a novel field of technology—What can they tell about interactions between emerging communities of science and technology? *Scientometrics*, 48(2), 151–178.
- Meyer, M. S. (2001). Patent citation analysis in a novel field of technology—An exploration of nano-science and nano-technology. *Scientometrics*, 51(1), 163–183.
- Meyer, M. (2002). Tracing knowledge flows in innovation systems. *Scientometrics*, 54(2), 193–212.
- Meyer, M., Siniläinen, T., & Utecht, J. T. (2003). Towards hybrid triple helix indicators—A study of university-related patents and a survey of academic inventors. *Scientometrics*, 58(2), 321–350.
- Michel, J., & Bettels, B. (2001). Patent citation analysis—A closer look at the basic input data from patent search reports. *Scientometrics*, 51(1), 185–201.
- Moed, H. F. (2000). Bibliometric indicators reflect publication and management strategies. *Scientometrics*, 47(2), 323–346.
- Moed, H. F. (2002a). Measuring China’s research performance using the Science Citation Index. *Scientometrics*, 53(3), 281–296.
- Moed, H. F. (2002b). The impact-factors debate—The ISI’s uses and limits. *Nature*, 415, 731–732.
- Moed, H. F. (2005). *Citation analysis in research evaluation*. Dordrecht, The Netherlands: Springer.
- Moed, H. F., Burger, W. J. M., Frankfort, J. G., & van Raan, A. F. J. (1985). The use of bibliometric data for the measurement of university research performance. *Research Policy*, 14(3), 131–149.
- Moed, H. F., Glänzel, W., & Schmoch, U. (Eds.). (2004). *Handbook of quantitative science and technology research: The use of publication and patent statistics in studies of S&T systems*. Dordrecht, London: Kluwer Academic.
- Moed, H. F., Luwel, M., & Nederhof, A. J. (2002). Towards research performance in the humanities. *Library Trends*, 50(3), 498–520.
- Moed, H. F., & van Leeuwen, T. N. (1995). Improving the accuracy of institute for scientific information’s journal impact factors. *Journal of the American Society for Information Science and Technology*, 46(6), 461–467.
- Moed, H. F., & van Leeuwen, T. N. (1996). Impact factors can mislead. *Nature*, 381, 166.
- Montemurro, M. A. (2001). Beyond the Zipf-Mandelbrot law in quantitative linguistics. *Physica A*, 300(39175), 567–578.
- Moody, J. (2004). The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociological Review*, 69(2), 213–238.
- Morillo, F., Bordons, M., & Gómez, I. (2001). An approach to interdisciplinarity bibliometric indicators. *Scientometrics*, 51(1), 203–222.
- Morillo, F., Bordons, M., & Gómez, I. (2003). Interdisciplinarity in science—A tentative typology of disciplines and research areas. *Journal of the American Society for Information Science and Technology*, 54(13), 1237–1249.

- Morris, S. A., Yen, G., Wu, Z., & Asnake, B. (2003). Time line visualization of research fronts. *Journal of the American Society for Information Science and Technology*, 54(5), 413–422.
- Mowery, D. C., Nelson, R. R., Sampat, B. N., & Ziedonis, A. A. (2001). The growth of patenting and licensing by US universities—An assessment of the effects of the Bayh-Dole act of 1980. *Research Policy*, 30(1), 99–119.
- Mowery, D. C., & Ziedonis, A. A. (2002). Academic patent quality and quantity before and after the Bayh-Dole act in the United States. *Research Policy*, 31(3), 399–418.
- Murray, F. (2002). Innovation as co-evolution of scientific and technological networks—Exploring tissue engineering. *Research Policy*, 31(8–9), 1389–1403.
- Must, Ü. (2006). New countries in Europe—Research, development and innovation strategies vs bibliometric data. *Scientometrics*, 66(2), 241–248.
- Nederhof, A. J., Luwel, M., & Moed, H. F. (2001). Assessing the quality of scholarly journals in Linguistics—An alternative to citation-based journal impact factors. *Scientometrics*, 51(1), 241–265.
- Nederhof, A. J., & van Raan, A. F. J. (1987). Peer review and bibliometric indicators of scientific performance: A comparison of cum laude doctorates with ordinary doctorates in physics. *Scientometrics*, 11(5–6), 333–350.
- Neuberger, J., & Counsell, C. (2002). Impact factors—Uses and abuses. *European Journal of Gastroenterology & Hepatology*, 14(3), 209–211.
- Neuhaus, C., Neuhaus, E., Asher, A., & Wrede, C. (2006). The depth and breadth of Google Scholar—An empirical study. *portal: Libraries and the Academy*, 6(2), 127–141.
- Newman, M. E. J. (2001a). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 64(1). Art. No. 016132 Part 2
- Newman, M. E. J. (2001b). Clustering and preferential attachment in growing networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 64(2). Art. No. 025102 Part 2.
- Newman, M. E. J. (2001c). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America*, 98(2), 404–409.
- Newman, M. E. J. (2003a). The structure and function of complex networks. *SIAM Review*, 45, 167–256.
- Newman, M. E. J. (2003b). Ego-centered networks and the ripple effect. *Social Networks*, 25(1), 83–95.
- Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 69(2) [Art. No. 026113, Part 2].
- Nicholls, P. T. (1989). Bibliometric modeling processes and the empirical validity of Lotka's law. *Journal of the American Society for Information Science*, 40(6), 379–385.
- Nisonger, T. E. (2000). Use of the Journal Citation Reports for serials management in research libraries—An investigation of the effect of self-citation on journal rankings in library and information science and genetics. *College & Research Libraries*, 61(3), 263–275.
- Nisonger, T. E. (2002). The relationship between international editorial board composition and citation measures in political science, business, and genetics journals. *Scientometrics*, 54(2), 257–268.
- Nisonger, T. E. (2005). Citation autobiography—An investigation of ISI database coverage in determining author citedness. *College & Research Libraries*, 65(2), 152–163.
- Norris, M., & Oppenheim, C. (2003). Citation counts and the research assessment exercise V—Archaeology and the 2001 RAE. *Journal of Documentation*, 59(6), 709–730.
- Noruzi, A. (2005). Google Scholar: The new generation of citation indexes. *LIBRI*, 55(4), 170–180.
- Noyons, E. (2001). Bibliometric mapping of science in a science policy context. *Scientometrics*, 50(1), 83–98.
- Odlyzko, A. (2002). The rapid evolution of scholarly communication. *Learned Publishing*, 15, 7–19.
- Onyancha, O. B., & Ocholla, D. N. (2005). An informetric investigation of the relatedness of opportunistic infections to HIV/AIDS. *Information Processing and Management*, 41(6), 1573–1588.
- Ortega, J. L., Aguillo, I., & Prieto, J. A. (2006). Longitudinal study of content and elements in the scientific web environment. *Journal of Information Science*, 32(4), 344–351.
- Oppenheim, C. (2000). Do patent citations count? In B. Cronin & H. Atkins (Eds.), *The Web of knowledge—A Festschrift in honor of Eugene Garfield*, ASIS Monograph Series (pp. 405–432). Medford, NJ: Information Today.
- Oppenheim, C., Greenhalgh, C., & Rowland, F. (2000). The future of scholarly journal publishing. *Journal of Documentation*, 56(4), 361–398.
- Otte, E., & Rousseau, R. (2002). Social network analysis—A powerful strategy, also for the information sciences. *Journal of Information Science*, 28(6), 441–453.
- Patel, V., & Sumathipala, A. (2001). International representation in psychiatric literature—Survey of six leading journals. *British Journal of Psychiatry*, 178(5), 406–409.
- Paasi, A. (2005). Globalisation, academic capitalism, and the uneven geographies of international journal publishing spaces. *Environment and Planning A*, 37(5), 769–789.
- Paladugu, R., Schein, M., Gardezi, S., & Wise, L. (2002). One hundred citation classics in general surgical journals. *World Journal of Surgery*, 26(9), 1099–1105.
- Park, H. W., & Thelwall, M. (2003). Hyperlink analyses of the World Wide Web—A review. *Journal of Computer Mediated Communication*, 8(4.). Retrieved July 11, 2007, from <http://jcmc.indiana.edu/vol8/issue4/park.html>
- Patsopoulos, N. A., Analatos, A. A., & Ioannidis, J. P. A. (2005). Relative citation impact of various study designs in the health sciences. *JAMA*, 293(19), 2362–2366.
- Payne, D. (2004). Google Scholar welcomed. *The Scientist*, 5(1), 20041123-01. Retrieved July 31, 2007, from <http://www.the-scientist.com/news/20041123/01/>
- Peritz, B. C., & Bar-Ilan, J. (2002). The sources used by bibliometrics—scientometrics as reflected in references. *Scientometrics*, 54(2), 269–284.
- Perkel, J. M. (2005). The future of citation analysis. *The Scientist*, 19(20), 24–30.

- Persson, O. (2001). All author citations versus first author citations. *Scientometrics*, 50(2), 339–344.
- Persson, O., Glänzel, W., & Danell, R. (2004). Inflationary bibliometric values—The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics*, 421–432.
- Pettigrew, K. E., & McKechnie, L. (2001). The use of theory in information science research. *Journal of the American Society for Information Science and Technology*, 52(1), 62–73.
- Pichappan, P., & Sarasvady, S. (2002). The influence of author self-citations on bibliometric macro indicators. *Scientometrics*, 54(2), 285–290.
- Pieters, R., & Baumgartner, H. (2002). Who talks to whom? Intra- and interdisciplinary communication of economics journals. *Journal of Economic Literature*, 40(2), 483–509.
- Pinski, G., & Narin, F. (1976). Citation influence for journal aggregates of scientific publications: Theory with an application to the literature of physics. *Information Processing and Management*, 12, 297–312.
- Porter, A. L., Kongthon, A., & Lui, J. C. (2002). Research profiling—Improving the literature review. *Scientometrics*, 53(3), 351–370.
- Price, D. J. de S. (1986). *Little science, big science and beyond*. New York: Columbia University Press.
- Prime, C., Bassecoulard, E., & Zitt, M. (2002). Co-citations and co-sitations—A cautionary view on an analogy. *Scientometrics*, 54(2), 291–308.
- Pudovkin, A. I., & Garfield, E. (2002). Algorithmic procedure for finding semantically related journals. *Journal of the American Society for Information Science and Technology*, 53(13), 1113–1119.
- Rahman, M., & Fukui, T. (2002). A decline in the U.S. share of research articles. *New England Journal of Medicine*, 347(15), 1211–1212.
- Rahman, M., & Fukui, T. (2003a). Biomedical publication—Global profile and trend. *Public Health*, 117, 274–280.
- Rahman, M., & Fukui, T. (2003b). Biomedical research productivity—Factors across the countries. *International Journal of Technology Assessment in Health Care*, 19(1), 249–260.
- Ramasco, J. J., Dorogovtsev, S. N., & Pastor-Satorras, R. (2004). Self-organization of collaboration networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 70(3) [Art. No. 036106, Part 2].
- Ramírez, A. M., García, E. O., & Del Río, J. (2000). Renormalized impact factor. *Scientometrics*, 47(1), 3–9.
- Ramos-Rodríguez, A. R., & Ruiz-Navarro, J. (2004). Changes in the intellectual structure of strategic management research—A bibliometric study of the Strategic Management Journal, 1980–2000. *Strategic Management Journal*, 25(10), 981–1004.
- Ray, J., Berkswits, M., & Davidoff, F. (2000). The fate of manuscripts rejected by a general medical journal. *American Journal of Medicine*, 109(2), 131–135.
- Redman, J., Willett, P., Allen, F. H., & Taylor, R. (2001). A citation analysis of the Cambridge Crystallographic Data Centre. *Journal of Applied Crystallography*, 34, 375–380.
- Ren, S. L., & Rousseau, R. (2002). International visibility of Chinese scientific journals. *Scientometrics*, 53(3), 389–405.
- Rinia, E. J., van Leeuwen, T. N., Bruins, E. E. W., van Vuren, H. G., & van Raan, A. F. J. (2001). Citation delay in interdisciplinary knowledge exchange. *Scientometrics*, 51(1), 293–309.
- Rinia, E. J., van Leeuwen, T. N., Bruins, E. E. W., van Vuren, H. G., & van Raan, A. F. J. (2002). Measuring knowledge transfer between fields of science. *Scientometrics*, 54(3), 347–362.
- Rinia, E. J., van Leeuwen, T. N., van Vuren, H. G., & van Raan, A. F. J. (2001). Influence of interdisciplinarity on peer-review and bibliometric evaluations in physics research. *Research Policy*, 30(3), 357–361.
- Rogers, L. F. (2001). Impact factor—The numbers game. *American Journal of Roentgenology*, 178(3), 541–542.
- Roessner, D. (2000). Quantitative and qualitative methods and measures in the evaluation of research. *Research Evaluation*, 8(2), 125–132.
- Rosenberg, G., Holden, G., & Barker, K. (2005). What happens to our ideas? A bibliometric analysis of articles in Social Work in Health Care in the 1990s. *Social Work in Health Care*, 41(3/4), 103–108.
- Rothman, J., Kirk, S. A., & Knapp, H. (2003). Reputation and publication productivity among social work researchers. *Social Work Research*, 27(2), 105–115.
- Rousseau, B., & Rousseau, R. (2000). LOTKA—A program to fit a power law distribution to observed frequency data. *Cybermetrics*, 4(4). Retrieved May 31, 2007, from <http://www.cindoc.csic.es/cybermetrics/articles/v4i1p4.html>
- Rousseau, R. (1997). Sitations: An exploratory study. *Cybermetrics*, 1(1). Retrieved July 30, 2002, from <http://www.cindoc.csic.es/cybermetrics/articles/v1i1p1.pdf>
- Rousseau, R. (2002a). Lack of standardisation in informetric research. Comments on “Power laws of research output. Evidence for journals of economics” by Matthias Sutter and Martin G. Kocher. *Scientometrics*, 55(2), 317–327.
- Rousseau, R. (2002b). Journal evaluation: Technical and practical issues. *Library Trends*, 50(3), 418–439.
- Rousseau, R. (2005a). Robert Fairthorne and the empirical power laws. *Journal of Documentation*, 61(2), 194–202.
- Rousseau, R. (2005b). Median and percentile impact factors—A set of new indicator. *Scientometrics*, 63(3), 431–441.
- Rousseau, R., Jin, B. H., Yang, N. H., & Liu, X. S. (2001). Observations concerning the two- and three-year synchronous impact factor, based on the Chinese Science Citation Database. *Journal of Documentation*, 57(3), 349–357.
- Rousseau, R., & Zuccala, A. (2004). A classification of author co-citations—Definitions and search strategies. *Journal of the American Society for Information Science and Technology*, 55(6), 513–529.
- Rowlands, I. (2002). Journal diffusion factors—A new approach to measuring research influence. *ASLIB Proceedings*, 54(2), 77–84.
- Ruiz-Pérez, R., López-Cózar, E. D., & Jiménez-Contreras, E. (2002). Spanish personal name variations in national and international biomedical databases—Implications for information retrieval and bibliometric studies. *Journal of the Medical Library Association*, 90(4), 411–430.
- 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(1), 117–120.
- Saha, S., Saint, S., & Christakis, D. A. (2003). Impact factor—A valid measure of journal quality? *Journal of the Medical Library Association*, 91(1), 41–46.
- Sandstrom, P. E. (2001). Scholarly communication as a socioecological system. *Scientometrics*, 51(3), 573–605.

- Schamhorst, A., & Wouters, P. (2006). Web indicators—A new generation of S&T indicators? *Cybermetrics*, 10(1), paper 6. Retrieved July 12, 2007, from <http://www.cindoc.csic.es/cybermetrics/articles/v10i1p6.html>.
- Schloegl, C., & Stock, W. G. (2004). Impact and relevance of LIS journals—A scientometric analysis of international and German-language LIS journals—Citation analysis versus reader survey. *Journal of the American Society for Information Science and Technology*, 56(1), 1155–1168.
- Schoepflin, U., & Glänzel, W. (2001). Two decades of “Scientometrics”—An interdisciplinary field represented by its leading journal. *Scientometrics*, 50(2), 301–312.
- Schubert, A. (2002). The Web of Scientometrics—A statistical overview of the first 50 volumes of the journal. *Scientometrics*, 53(1), 3–20.
- Schummer, J. (2004). Multidisciplinarity, interdisciplinarity, and patterns of research collaboration in nanoscience and nanotechnology. *Scientometrics*, 59(3), 425–465.
- Scopus. (2007). *Elsevier is proud to announce today's launch of the world's largest scientific abstract database*. Retrieved July 31, 2007, from [http://www.info.scopus.com/news/press/pr\\_041103.asp](http://www.info.scopus.com/news/press/pr_041103.asp).
- Seglen, P. O., & Aksnes, D. W. (2000). Scientific productivity and group size—A bibliometric analysis of Norwegian microbiological research. *Scientometrics*, 49(1), 125–143.
- Shinn, T. (2002). The triple helix and new production of knowledge—Prepackaged thinking on science and technology. *Social Studies of Science*, 32(4), 599–614.
- Sinai, M. (2007). *Brandon Hill selected lists—Important announcement*. Retrieved June 18, 2007, from <http://www.mssm.edu/library/brandon-hill/>.
- Small, H. (2003). Paradigms, citations, and maps of science—A personal history. *Journal of the American Society for Information Science and Technology*, 54(5), 394–399.
- Small, H. (2004). On the shoulders of Robert Merton—Towards a normative theory of citation. *Scientometrics*, 60(1), 71–79.
- Small, H. (2006). Tracking and predicting growth areas in science. *Scientometrics*, 68(3), 595–610.
- Smith, A. G., & Thelwall, M. (2002a). Web impact factors for Australasian universities. *Scientometrics*, 54(3), 363–380.
- Solari, A., & Magri, M. H. (2000). A new approach to the SCI Journal Citation Reports, a system for evaluating scientific journals. *Scientometrics*, 47(3), 605–625.
- Sombatsompop, N., & Markpin, T. (2005). Making an equality of ISI impact factors for different subject fields. *Journal of the American Society for Information Science and Technology*, 56(7), 676–683.
- Sombatsompop, N., Markpin, T., & Premkamolnetr, N. (2004). A modified method for calculating the Impact Factors of journals in ISI Journal Citation Reports—Polymer Science Category in 1997–2001. *Scientometrics*, 60(2), 235–271.
- Soteriades, E. S., & Falagas, M. E. (2005). Comparison of amount of biomedical research originating from the European Union and the United States. *British Medical Journal*, 331(7510), 192–194.
- Soteriades, E. S., Rosmarakis, E. S., Paraschakis, K., & Falagas, M. E. (2006). Research contribution of different world regions in the top 50 biomedical journals (1995–2002). *The FASEB Journal*, 20(1), 19–34.
- Sprott, D. E., & Miyazaki, A. D. (2004). Two decades of contributions to marketing and public policy—An analysis of research published in Journal of Public Policy & Marketing. *Journal of Public Policy & Marketing*, 21(1), 105–125.
- Steele, T. W., & Styer, J. C. (2000). The impact of interdisciplinary research in the environmental sciences—A forestry case study. *Journal of the American Society for Information Science and Technology*, 51(5), 476–484.
- Stegmann, J., & Grohmann, G. (2001). Citation rates, knowledge export and international visibility of dermatology journals listed and not listed in the Journal Citation Reports. *Scientometrics*, 50(3), 483–502.
- Stegmann, J., & Grohmann, G. (2003). Hypothesis generation guided by co-word clustering. *Scientometrics*, 56(1), 111–135.
- Swanson, D. R., Smalheiser, N. R., & Bookstein, A. (2001). Information discovery from complementary literatures—Categorizing viruses as potential weapons. *Journal of the American Society for Information Science and Technology*, 52(10), 797–812.
- Tague-Sutcliffe, J. (1992). An introduction to informetrics. *Information Processing and Management*, 28(1), 1–3.
- Tang, R., & Thelwall, M. (2002a). Exploring the pattern of links between Chinese university Websites. *Proceedings of the American Society for Information Science and Technology*, 39, 417–424.
- Tang, R., & Thelwall, M. (2003). U.S. academic departmental Web-site interlinking in the United States disciplinary differences. *Library & Information Science Research*, 25(4), 437–458.
- Thelwall, M. (2000). Web impact factors and search engine coverage. *Journal of Documentation*, 56(2), 185–189.
- Thelwall, M. (2001a). Extracting macroscopic information from Web links. *Journal of the American Society for Information Science and Technology*, 52(13), 1157–1168.
- Thelwall, M. (2001b). A web crawler design for data mining. *Journal of Information Science*, 27(5), 319–325.
- Thelwall, M. (2001c). Results from a web impact factor crawler. *Journal of Documentation*, 57(2), 177–191.
- Thelwall, M. (2001d). The responsiveness of search engine indexes. *Cybermetrics*, 5(1), paper 1, Retrieved July 14, 2007, from <http://www.cindoc.csic.es/cybermetrics/articles/v5i1p1.html>.
- Thelwall, M. (2001e). Exploring the link structure of the Web with network diagrams. *Journal of Information Science*, 27(6), 393–401.
- Thelwall, M. (2002a). A comparison of sources of links for academic Web impact factor calculations. *Journal of Documentation*, 58(1), 66–78.
- Thelwall, M. (2002b). Conceptualizing documentation on the Web—An evaluation of different heuristic-based models for counting links between university Web sites. *Journal of the American Society for Information Science and Technology*, 53(12), 995–1005.
- Thelwall, M. (2002c). Methodologies for crawler based Web surveys. *Internet Research—Electronic Networking Applications and Policy*, 12(2), 124–138.
- Thelwall, M. (2002d). Evidence for the existence of geographic trends in university Web site interlinking. *Journal of Documentation*, 58(5), 563–574.
- Thelwall, M. (2002e). An initial exploration of the link relationship between UK university Web sites. *ASLIB Proceedings*, 54(2), 118–126.
- Thelwall, M. (2002f). The top 100 linked-to pages on UK university web sites—High inlink counts are not usually associated with quality scholarly content. *Journal of Information Science*, 28(6), 483–491.



- Thelwall, M. (2002g). Research dissemination and invocation on the Web. *Online Information Review*, 26(6), 413–420.
- Thelwall, M. (2003a). Web use and peer interconnectivity metrics for academic web sites. *Journal of Information Science*, 29(1), 1–10.
- Thelwall, M. (2003b). Can Google's PageRank be used to find the most important academic Web pages? *Journal of Documentation*, 59(2), 205–217.
- Thelwall, M. (2003c). What is this link doing here? Beginning a fine-grained process of identifying reasons for academic hyperlink creation. *Information Research*, 8(3), paper 151. Retrieved July 14, 2007, from <http://informationr.net/ir/8-3/paper151.html>.
- Thelwall, M. (2004a). *Link analysis: An information science approach*. Amsterdam: Elsevier.
- Thelwall, M. (2004b). Methods for reporting on the targets of links from national systems of university Web sites. *Information Processing and Management*, 40(1), 125–144.
- Thelwall, M. (2004c). Can the Web give useful information about commercial uses of scientific research? *Online Information Review*, 28(2), 120–130.
- Thelwall, M. (2006). Interpreting social science link analysis research—A theoretical framework. *Journal of the American Society for Information Science and Technology*, 57(1), 60–68.
- Thelwall, M., Barjak, F., & Kretschmer, H. (2006). Web links and gender in science—An exploratory analysis. *Scientometrics*, 67(3), 373–383.
- Thelwall, M., Binns, R., Harries, G., Page-Kennedy, T., Price, L., & Wilkinson, D. (2002). European Union associated university websites. *Scientometrics*, 53(1), 95–111.
- Thelwall, M., & Harries, G. (2003). The connection between the research of a university and counts of links to its web pages—An investigation based upon a classification of the relationships of pages to the research of the host university. *Journal of the American Society for Information Science and Technology*, 54(7), 594–602.
- Thelwall, M., & Harries, G. (2004a). Can personal web pages that link to universities yield information about the wider dissemination of research? *Journal of Information Science*, 30(3), 240–253.
- Thelwall, M., & Harries, G. (2004b). Do the Web sites of higher rated scholars have significantly more online impact? *Journal of the American Society for Information Science and Technology*, 55(2), 149–159.
- Thelwall, M., & Smith, A. (2002). Interlinking between Asia-Pacific University Web sites. *Scientometrics*, 55(3), 363–376.
- Thelwall, M., & Tang, R. (2003). Disciplinary and linguistic considerations for academic Web linking—An exploratory hyperlink mediated study with Mainland China and Taiwan. *Scientometrics*, 58(1), 158–181.
- Thelwall, M., Tang, R., & Price, L. (2003). Linguistic patterns of academic Web use in Western Europe. *Scientometrics*, 56(3), 417–432.
- Thelwall, M., Vann, K., & Fairclough, R. (2006). Web issue analysis—An integrated water resource management case study. *Journal of the American Society for Information Science and Technology*, 57(10), 1303–1314.
- Thelwall, M., Vaughan, L., & Björneborn, L. (2005). Webometrics. In B. Cronin (Ed.), *Annual review of information science and technology* (pp. 81–135). Medford, NJ: Information Today.
- Thelwall, M., Vaughan, L., Cothey, V., Li, X. M., & Smith, A. G. (2003). Which academic subjects have most online impact? A pilot study and a new classification process. *Online Information Review*, 27(5), 333–343.
- Thelwall, M., & Wilkinson, D. (2003a). Three target document range metrics for university web sites. *Journal of the American Society for Information Science and Technology*, 54(6), 490–497.
- Thelwall, M., & Wilkinson, D. (2003b). Graph structure in three national academic webs—Power laws with anomalies. *Journal of the American Society for Information Science and Technology*, 54(8), 706–712.
- Thomas, O., & Willet, P. (2000). Webometric analysis of departments of librarianship and information science. *Journal of Information Science*, 26(6), 421–428.
- Thompson, J. W. (2002). The death of the scholarly monograph in the humanities? Citation patterns in literary scholarship. *LIBRI*, 52(3), 121–136.
- Thorsteinsdóttir, O. H. (2000). External research collaboration in two small science systems. *Scientometrics*, 49(1), 145–160.
- Tijssen, R. J. W. (2001). Global and domestic utilization of industrial relevant science—Patent citation analysis of science-technology interactions and knowledge flows. *Research Policy*, 30(1), 35–54.
- Tijssen, R. J. W. (2002). Science dependence of technologies—Evidence from inventions and their inventors. *Research Policy*, 31(4), 509–526.
- Tijssen, R. J. W. (2004). Is the commercialisation of scientific research affecting the production of public knowledge? Global trends in the output of corporate research articles. *Research Policy*, 33(5), 709–733.
- Tijssen, R. J. W., Buter, R. K., & van Leeuwen, T. N. (2000). Technological relevance of science—An assessment of citation linkages between patents and research papers. *Scientometrics*, 47(2), 389–412.
- Tijssen, R. J. W., Visser, M. S., & van Leeuwen, T. N. (2002). Benchmarking international scientific excellence—Are highly cited research papers an appropriate frame of reference? *Scientometrics*, 54(3), 381–397.
- Trayhurn, G. (2002). Citations and 'impact factor'—The Holy Grail. *British Journal of Nutrition*, 88(1), 1–2.
- Tsallis, C., & de Albuquerque, M. P. (2000). Are citations of scientific papers a case of nonextensivity? *European Physical Journal B*, 13(4), 777–780.
- Tsay, M. Y., Jou, S. J., & Ma, S. S. (2000). A bibliometric study of semiconductor literature, 1978–1997. *Scientometrics*, 49(3), 491–509.
- Tsay, M. Y., & Yang, Y. H. (2005). Bibliometric analysis of the literature of randomized controlled trials. *Journal of the Medical Library Association*, 93(4), 450–458.
- Ugolini, D., & Casilli, C. (2003). The visibility of Italian journals. *Scientometrics*, 56(3), 345–355.
- van den Besselaar, P. (2000). Communication between science and technology studies journals—A case study in differentiation and integration in scientific fields. *Scientometrics*, 47(2), 169–193.
- van den Besselaar, P. (2001). The cognitive and the social structure of STS. *Scientometrics*, 51(2), 441–460.
- van Dalen, H. P., & Henkens, K. (2001). What makes a scientific article influential? The case of demographers. *Scientometrics*, 50(3), 455–482.
- van Fleet, D. D., McWilliams, A., & Siegel, D. S. (2000). A theoretical and empirical analysis of journal rankings—The case of formal lists. *Journal of Management*, 26(5), 839–861.

- van Leeuwen, T. N., & Moed, H. (2002). Development and application of journal impact measures in the Dutch science system. *Scientometrics*, 53(2), 249–266.
- van Leeuwen, T. N., & Moed, H. (2005). Characteristics of Journal Impact Factors—The effects of uncitedness and citation distribution on the understanding of journal impact factors. *Scientometrics*, 63(2), 357–371.
- van Leeuwen, T. N., Moed, H. F., Tijssen, R. J. W., Visser, M. S., & van Raan, A. F. J. (2000a). First evidence of serious language-bias in the use of citation analysis for the evaluation of national science systems. *Research Evaluation*, 9(2), 155–156.
- van Leeuwen, T. N., Moed, H. F., Tijssen, R. J. W., Visser, M. S., & van Raan, A. F. J. (2001). Language biases in the coverage of the Science Citation Index and its consequences for international comparisons of national research performance. *Scientometrics*, 51(1), 335–346.
- van Leeuwen, T. N., Visser, M. S., Moed, H. F., Nederhof, T. J., & Van Raan, A. F. J. (2003). The holy grail of science policy—Exploring and combining bibliometric tools in search of scientific excellence. *Scientometrics*, 57(2), 257–280.
- van Looy, B., Zimmermann, E., Veugelers, R., Verbeek, A., Mello, J., & Debackere, K. (2003). Do science-technology interactions pay off when developing technology? An exploratory investigation of 10 science-intensive technology domains. *Scientometrics*, 57(3), 355–367.
- van Raan, A. F. J. (1996). Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. *Scientometrics*, 36(3), 397–420.
- van Raan, A. F. J. (2000a). On growth, ageing, and fractal differentiation of science. *Scientometrics*, 47(2), 347–362.
- van Raan, A. F. J. (2000b). The Pandora's box of citation analysis—Measuring scientific excellence—The last evil? In B. Cronin & H. Atkins (Eds.), *The Web of knowledge—A Festschrift in honor of Eugene Garfield, ASIS Monograph Series* (pp. 301–319). Medford, NJ: Information Today.
- van Raan, A. F. J. (2000c). R&D evaluation at the beginning of the new century. *Research Evaluation*, 8(2), 81–86.
- van Raan, A. F. J. (2001a). Bibliometrics and Internet: Some observations and expectations. *Scientometrics*, 50(1), 59–63.
- van Raan, A. F. J. (2001b). Competition amongst scientists for publication status—Toward a model of scientific publication and citation distributions. *Scientometrics*, 50(1), 347–357.
- van Raan, A. F. J. (2004). Sleeping beauties in science. *Scientometrics*, 59(3), 467–472.
- van Raan, A. F. J. (2005a). Reference-based publication networks with episodic memories. *Scientometrics*, 63(3), 549–566.
- van Raan, A. F. J. (2005b). Fatal attraction—Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics*, 62(1), 133–143.
- van Raan, A. F. J. (2005c). Reply to the comments of Liu et al. *Scientometrics*, 64(1), 111–112.
- van Raan, A. F. J. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*, 67(3), 491–502.
- van Raan, A. F. J., & van Leeuwen, T. N. (2002). Assessment of the scientific basis of interdisciplinary, applied research—Application of bibliometric methods in Nutrition and Food Research. *Research Policy*, 31(4), 611–632.
- van Teijlingen, E., & Hundley, V. (2002). Getting your paper to the right journal—A case study of an academic paper. *Journal of Advanced Nursing*, 37(6), 506–511.
- Vaughan, L. (2004a). New measurements for search engine evaluation proposed and tested. *Information Processing and Management*, 40(4), 677–691.
- Vaughan, L. (2004b). Web hyperlinks reflect business performance—A study of US and Chinese IT companies. *Canadian Journal of Information and Library Science*, 28(1), 17–31.
- Vaughan, L. (2004c). Exploring website features for business information. *Scientometrics*, 61(3), 467–477.
- Vaughan, L. (2006). Visualizing linguistic and cultural differences using web co-link data. *Journal of the American Society for Information Science and Technology*, 57(9), 1178–1193.
- Vaughan, L., & Hysen, K. (2002). Relationship between links to journal Web sites and impact factors. *ASLIB Proceedings*, 54(6), 356–361.
- Vaughan, L., Gao, Y., & Kipp, M. (2006). Why are hyperlinks to business Websites created? A content analysis. *Scientometrics*, 67(2), 291–300.
- Vaughan, L., & Shaw, D. (2003). Bibliographic and web citations—What is the difference? *Journal of the American Society for Information Science and Technology*, 54(14), 1313–1322.
- Vaughan, L., & Shaw, D. (2005). Web citation data for impact assessment—A comparison of four science disciplines. *Journal of the American Society for Information Science and Technology*, 56(10), 1075–1087.
- Vaughan, L., & Thelwall, M. (2003). Scholarly use of the Web—What are the key inducers of links to journal Web sites? *Journal of the American Society for Information Science and Technology*, 54(1), 29–38.
- Vaughan, L., & Thelwall, M. (2004). Search engine coverage bias—Evidence and possible causes. *Information Processing and Management*, 40(4), 693–707.
- Vaughan, L., & Thelwall, M. (2005). A modeling approach to uncover hyperlink patterns—The case of Canadian universities. *Information Processing and Management*, 41(2), 347–359.
- Vaughan, L., & Wu, G. Z. (2004). Links to commercial websites as a source of business information. *Scientometrics*, 60(3), 487–496.
- Vaughan, L., & You, J. (2006). Comparing business competition positions based on Web co-link data—The global market vs. the Chinese market. *Scientometrics*, 68(3), 611–628.
- Verbeek, A., Debackere, K., Luwel, M., Andries, P., Zimmermann, E., & Deleus, F. (2002). Linking science to technology—Using bibliographic references in patents to build linkage schemes. *Scientometrics*, 54(3), 399–420.
- Vessey, I., Ramesh, V., & Glass, R. L. (2002). Research in information systems—An empirical study of diversity in the discipline and its journals. *Journal of Management Information Systems*, 19(2), 129–174.
- Vinkler, P. (2001). An attempt for defining some basic categories of scientometrics and classifying the indicators of evaluative scientometrics. *Scientometrics*, 50(3), 539–544.
- Vinkler, P. (2002). Subfield problems in applying the Garfield (Impact) Factors in practice. *Scientometrics*, 53(2), 267–279.

- Vinkler, P. (2004). Characterization of the impact of sets of scientific papers—The Garfield (Impact) Factor. *Journal of the American Society for Information Science and Technology*, 55(5), 431–435.
- von Elm, E., Poglia, G., Walder, B., & Tramer, M. R. (2004). Different patterns of duplicate publication—An analysis of articles used in systematic reviews. *JAMA*, 291(8), 974–980.
- Vreeland, R. C. (2000). Law libraries in hyperspace—A citation analysis of World Wide Web sites. *Law Library Journal*, 92(1), 9–25.
- Wagner, C. S., & Leydesdorff, L. (2005). Network structure, self-organization, and the growth of international collaboration in science. *Research Policy*, 34(10), 1608–1618.
- Wagner-Döbler, R. (2001). Continuity and discontinuity of collaboration behaviour since 1800—From a bibliometric point of view. *Scientometrics*, 52(3), 503–517.
- Wallace, L. S., & Ballard, J. E. (2003). Osteoporosis coverage in selected women's magazines and newspapers, 1998–2001. *American Journal of Health Behavior*, 27(1), 75–83.
- Walter, G., Bloch, S., Hunt, G., & Fisher, K. (2003). Counting on citations—A flawed way to measure quality. *Medical Journal of Australia*, 178(6), 280–281.
- Warner, J. (2000). Critical review of the application of citation studies to the Research Assessment Exercises. *Journal of Information Science*, 26(6), 453–460.
- Webster, T. J. (2001). A principal component analysis of the US News & World Report tier rankings of colleges and universities. *Economics of Education Review*, 20(3), 235–244.
- Weale, A. R., Bailey, M., & Lear, P. A. (2004). The level of non-citation of articles within a journal as a measure of quality—A comparison to the impact factor. *BMC Medical Research Methodology*, 4(14). Retrieved June 18, 2007, from <http://www.biomedcentral.com/1471-2288/4/14>
- Weingart, P. (2005). Impact of bibliometrics upon the science system—Inadvertent consequences? *Scientometrics*, 62(1), 117–131.
- West, R., & McIlwaine, A. (2002). What do citation counts count for in the field of addiction? An empirical evaluation of citation counts and their link with peer ratings of quality. *Addiction*, 97(5), 501–504.
- Wetterer, J. K. (2006). Quotation error, citation copying, and ant extinctions in Madeira. *Scientometrics*, 67(3), 351–372.
- Wilkinson, D., Harries, G., Thelwall, M., & Price, L. (2003). Motivations for academic web site interlinking—Evidence for the Web as a novel source of information on informal scholarly communication. *Journal of Information Science*, 29(1), 49–56.
- Wilson, C. S. (1999). Informetrics. In M. E. Williams (Ed.), *Annual review of information science and technology* (pp. 107–247). Medford, NJ: Information Today.
- Winkmann, G., Schlutius, S., & Schweim, H. (2002). Publication languages of Impact Factor Journals and of medical bibliographic databanks. *Deutsche Medizinische Wochenschrift*, 127(4), 131–137.
- White, H. D. (2000). *Toward ego-centered citation analysis. The Web of knowledge—A Festschrift in honor of Eugene Garfield. ASIS Monograph Series*. Medford, NJ: Information Today., pp. 475–496.
- White, H. D. (2001a). Authors as citers over time. *Journal of the American Society for Information Science and Technology*, 52(2), 87–108.
- White, H. D. (2001b). Author-centered bibliometrics through CAMEOs—Characterizations automatically made and edited online. *Scientometrics*, 51(3), 607–637.
- White, H. D. (2003a). Pathfinder networks and author cocitation analysis—A remapping of paradigmatic information scientists. *Journal of the American Society for Information Science and Technology*, 54(5), 423–434.
- White, H. D. (2003b). Author cocitation analysis and Pearson's *r*. *Journal of the American Society for Information Science and Technology*, 54(13), 1250–1259.
- White, H. D. (2004). Citation analysis and discourse analysis revisited. *Applied Linguistics*, 25(1), 89–116.
- White, H. D., Wellman, B., & Nazer, N. (2004). Does citation reflect social structure? Longitudinal evidence from the “Globenet” interdisciplinary research group. *Journal of the American Society for Information Science and Technology*, 55(2), 111–126.
- Whitehouse, G. H. (2001). Citation rates and impact factors—Should they matter? *British Journal of Radiology*, 74(877), 1–3.
- Wleklinski, J. M. (2005). Studying google scholar—Wall to wall coverage? *Online*, 29(3), 22–26.
- Wolfram, D. (2006). Applications of SQL for informetric frequency distribution processing. *Scientometrics*, 67(2), 301–313.
- Wormell, I. (2000a). Critical aspects of the Danish Welfare State—As revealed by issue tracking. *Scientometrics*, 48(2), 237–250.
- Wormell, I. (2000b). Bibliometric analysis of the Welfare topic. *Scientometrics*, 48(2), 203–236.
- Wolf, S. H., & Johnson, R. E. (2000). A one-year audit of topics and domains in the Journal of the American Medical Association and the New England Journal of Medicine. *American Journal of Preventive Medicine*, 19(2), 79–86.
- Wouters, P. (1999). *The citation culture*. Retrieved June 10, 2007, from <http://garfield.library.upenn.edu/wouters/wouters.pdf>
- Wren, J. D. (2005). Open access and openly accessible—A study of scientific publications shared via the Internet. *British Medical Journal*, 330(7500), 1128–1131.
- Yoon, B., & Park, Y. (2004). A text-mining-based patent network—Analytical tool for high-technology trend. *Journal of High Technology Management Research*, 15(1), 37–50.
- Yoshikane, F., Kageura, K., & Tsuji, K. (2003). A method for the comparative analysis of concentration of author productivity, giving consideration to the effect of sample size dependency of statistical measure. *Journal of the American Society for Information Science and Technology*, 54(6), 521–528.
- Yu, G., Wang, X. H., & Yu, D. R. (2005). The influence of publication delays on impact factors. *Scientometrics*, 64(20), 235–246.
- Zhao, D. Z. (2005). Challenges of scholarly publications on the Web to the evaluation of science—A comparison of author visibility on the Web and in print journals. *Information Processing and Management*, 41(6), 1403–1418.
- Zhao, D. Z. (2006). Towards all-author co-citation analysis. *Information Processing and Management*, 42(6), 1578–1591.
- Zhao, D., & Logan, E. (2002). Citation analysis using scientific publications on the Web as data source—A case study in the XML research area. *Scientometrics*, 54(3), 449–472.

- Zhou, D. N., Ma, H., & Turban, E. (2001). Journal quality assessment—An integrated subjective and objective approach. *IEEE Transactions of Engineering Management*, 48(4), 479–490.
- Zhou, P., & Leydesdorff, L. (2006). The emergence of China as a leading nation in science. *Research Policy*, 35(1), 83–104.
- Zitt, M., & Bassecouard, E. (2006). Delineating complex scientific fields by an hybrid lexical-citation method—An application to nanosciences. *Information Processing and Management*, 42(6), 1513–1531.
- Zitt, M., Bassecouard, E., & Okubo, Y. (2000). Shadows of the past in international cooperation—Collaboration profiles of the top five producers of science. *Scientometrics*, 47(3), 627–657.
- Zitt, M., Ramanana-Rahary, S., & Bassecouard, E. (2003). Correcting glasses help fair comparisons in international science landscape—Country indicators as a function of ISI database delineation. *Scientometrics*, 56(2), 259–282.
- Zitt, M., Ramanana-Rahary, S., & Bassecouard, E. (2005). Relativity of citation performance and excellence measures—From cross-field to cross-scale effects of field-normalisation. *Scientometrics*, 63(2), 373–401.
- Zitt, M., Ramanana-Rahary, S., Bassecouard, E., & Laville, F. (2003). Potential science-technology spillovers in regions—An insight on geographic co-location of knowledge activities in the EU. *Scientometrics*, 57(2), 295–320.
- Zuccala, A. (2006a). Author cocitation analysis is to intellectual structure as web colink analysis is to ...? *Journal of the American Society for Information Science and Technology*, 57(11), 1487–1502.
- Zuccala, A. (2006b). Modeling the invisible college. *Journal of the American Society for Information Science and Technology*, 57(2), 152–168.