Michael J. Nelson
Faculty of Information and Media Studies
University of Western Ontario
London, Ontario

An Alternate Method for Ranking Journals using Citations

Abstract: An alternate method for ranking journals based on the algorithm used in the Google search engine for pagerank is applied to the information science and library science set of journals from Journal Citation Reports. A method of calculating individual paper influence based on this algorithm is proposed.

Résumé : Une méthode alternative pour classer les périodiques et basée sur l'algorithme utilisé par la fonction *pagerank* du moteur de recherche Google est appliquée à l'ensemble des périodiques des sciences de l'information et de bibliothéconomie de *Journal Citation Reports.* Une méthode pour calculer l'influence individuelle des articles basés sur cet algorithme est proposée.

1. Introduction

A common measure of the importance of a journal in a field is the impact factor as calculated by the Institute of Scientific Information (ISI) and published in their *Journal Citation Reports* (JCR) annually. This is calculated by dividing the total number of citations to articles published in the previous two years from articles published in the JCR year by the total number of articles published in the previous two years. This is the average number of times that articles published in the journal in the previous two years have been cited. This includes self citations, cites from articles in the same journal. The only normalizing factor in this calculation is the number of articles that are published in the journal.

This measure has been criticized many times. For example by Seglen (1997), lists a number of limiting factors for the measure. First, the journal impact factor conceals the difference in article citation rates, so we should be looking at more article-based measures. Also, "Journal impact factors depend on the research field: high impact factors are likely in journals covering large areas of basic research with a rapidly expanding but short lived literature that use many references per article" (Seglen, 1997). Thus, it is useful to analyze subfields separately as we will here. Garfield (1996) himself suggested some modifications of impact factor that could be done and also emphasizes that individual articles are important when doing evaluations. He also discusses some of the differences between review articles, method articles, and others. Glänzel and Moed (2002) give a comprehensive review of previous work on journal measurements based on citations and summarize nine limitations of the impact factor measure discussed previously in the literature. The first two limitations listed are that there is no normalization for different referencing patterns in the difference disciplines and that the quality and merits of the citing journal are not incorporated. Vinkler (2002) also concentrates on the problems of journals in subfields and proposes an adjustment based on the citations within a set of journals. Mainly because of these reasons, we are interested in recalculating the ranking of the journals by bringing a new factor: the importance of the journal that references another journal. The journals analyzed will be limited to one subject field. We can do this by analogy with the pagerank factor that is used as weighting factor in some internet search engines, particularly in Google. The original white paper by Brin and Page (n.d.) is still available on the web. Before we discuss this measure for journals we will introduce some notation.

2. Notation

In order to discuss any of the measures based on journal citations including the measure in this paper, we start with an asymmetric journal cross-citation matrix. Using the same notation as Glänzel and Moed (2002) for the matrix, we have the data as displayed in table 1. There are journal titles in the rows (citing sources) and in the columns (cited journals). Each cell value C_{ij} is the number of references that the source journal J_{is} gives to a destination (cited) journal J_{dj} . The row sum R_i represents the total number of references given from J_i to all the other journals. The column sum C_j represents the total number of citations received by journal J_i from the source journals. In addition, one must specify the time period for both the source journals (t_r) and the destination journals (t_c).

	J_{d1}	…	J_{dj}	…	J_{dm}	sum
J_{1s}	C_{11}	…	C_{1j}	…	C_{1m}	R_1
…	…	…	…	…	…	…
J_{is}	C_{i1}	…	C_{ij}	…	C_{im}	R_i
…	…	…	…	…	…	…
J_{ns}	C_{n1}	…	C_{nj}	…	C_{nm}	R_n
Sum	C_1	…	C_j	…	C_m	

Table 1. Cross-citation matrix

We can then express the impact factor of a journal as:

$IF = C_i(t_c)/(|\ J_d(t_{c-2})\,| + |\ J_d(t_{c-1})\ |)$

Another measure that has the same objectives as our measure, is the *journal influence* measure of Pinski and Narin (1978), which can be expressed as:

$W_i = \sum_k W_k C_{ki} / R_i$

As we will show later, our measure based on page rank is closely related to their influence measure.

3. Calculating a "Journal Rank" and Influence.

We will start with a definition of pagerank for web pages:

> "PageRank is defined as follows:
>
> We assume page A has pages T1...Tn which point to it (i.e., are citations). The parameter d is a damping factor which can be set between 0 and 1. We usually set d to 0.85. There are more details about d in the next section. Also C(A) is defined as the number of links going out of page A. The PageRank of a page A is given as follows:
>
> PR(A) = (1-d) + d (PR(T1)/C(T1) + ... + PR(Tn)/C(Tn))

Note that the PageRanks form a probability distribution over web pages, so the sum of all web pages' PageRanks will be one."
(Brin and Page, Section 2.1.1)

The damping factor is not justified in the paper but seems to included to allow an adjustment for how much the other pages influence the pagerank of page A. Also, the correct wording is that the *average* pagerank is one, not the sum. Another way of explaining the formula is that the page rank of a web page is divided up evenly amongst the pages that it links to.

We can calculate the "pagerank" of a journal by using the same formula but substituting journals for web pages. We must first interpret $C(T_i)$ for journals. If we stay strictly to the analogy then $C(J_i)$ is the number of other different journals that journal T_i references. It does not count the number of references and does not count self citations. The formula for *journal rank* F of journal A then becomes:

$$F(A) = (1-d) + d\,(F(J_1)/C(J_1) + \ldots + F(J_n)/\,C(J_n))$$

where journal A has journals J_1 to J_n which reference journal A. We will call it *journal rank* to continue the terminology used for web pages. Thus the journal rank of a journal is divided up among the journals it references. The parameter d is a damping factor which can be set between 0 and 1. Brin and Page usually set d to 0.85 to start. Since all weights are distributed over the journals the average of all journal ranks will be one. So F is not an actual rank number but a journal ordering can be calculated from it. In order to calculate the actual values an iterative procedure must be employed, as we don't know any of the F values to insert in the formula above. We can start with simple initial values such as one and in most cases the formula will converge to stable values. Since this is n equations in n unkowns we can use standard matrix theory to explain the solution of the equations (see also Pinski and Narin,1976).

To test the feasibility of this procedure a data subset from 54 journals in information science and library science as designated in JCR 2004 for the social sciences has been downloaded. The journal rank values have been calculated and the results compared with the ranking by impact factor and Pinski and Narin's influence measure (See Appendix 1). Before we make comparisons between the different measures we will develop one more measure.

There are a number of things in the formula that could be easily changed. First, if we truly want to reflect the effect of other journals we could remove the damping factor. It is not necessary for the calculation of the values. In addition we could also include the journal itself as an influence (i.e. self citations), which would have the effect of slightly raising the influence of journals at the top of the range and lowering the journals near the bottom. The third possible improvement would be to use the actual number of references that a journal makes in place of just the number journals. In this case to keep with the original idea of proportioning the total journal rank over all journals referenced we must take the proportion C_{ki} / R_i of each journal rank. If we make all these changes then we have a measure that we will call the modified journal rank FN:

$$F'(J_i) = \sum_k F'(J_k)\,C_{ki} / R_i$$

Now compare this to the influence measure of Pinski and Narin (1976) which is:

$$W_i = \sum_k W_k C_{ki} / R_i$$

we see this exactly the same as our modified journal rank. So coming from an information retrieval analogy we have arrived at the same formula. For a matrix and eigenvalue derivation of the solutions see Pinski and Narin (1976). Geller (1977) also derives sufficient conditions for the solution to exist using Markov chain theory.

To see how the three measures differ, tables 2, 3, and 4 show the ten top ranked journals as determined by each of the three measures, impact factor, journal rank and influence respectively. First, one should remember that impact factor is calculated over all journals in the social sciences whereas we only make calculations for the subset of information science and library science. We can see that the Annual Review of Information Science moves down in the rankings for the new measures (it was twelfth by influence). Also MIS Quarterly has a high impact factor in general but is not receive many citations from the top information science journals so moves down the rankings. The same thing happens for a few other journals. On the other hand, journals such as the Journal of Information Science and Information Research move up in the rankings because they are more central to the field. For the subject area of Information and library science both journal rank and influence seem to give a better measure of the important journals in the subject area. The journal rank measure seems to be between impact factor and influence in placement of journals in the ranking.

Abbreviated Journal Title	Total Cites	Impact Factor	Articles	Journal Rank	Influence
ANNU REV INFORM SCI	296	4.292	12	2.290	1.763
INFORM SYST RES	615	3.512	20	0.451	0.067
MIS QUART	1869	2.884	24	0.573	0.212
J AM MED INFORM ASSN	1468	2.612	61	0.361	0.523
J AM SOC INF SCI TEC	2254	2.086	97	2.899	5.818
INFORM MANAGE-AMSTER	1013	1.815	72	1.041	1.356
J DOC	654	1.542	26	1.538	2.003
INFORM PROCESS MANAG	992	1.295	54	1.786	2.104
J MANAGE INFORM SYST	944	1.271	35	0.466	0.255
INT J GEOGR INF SCI	899	1.234	37	0.341	3.194

Table 2. Top 10 journals by Impact Factor.

Abbreviated Journal Title	Total Cites	Impact Factor	Articles	Journal Rank	Influence
J AM SOC INF SCI TEC	2254	2.086	97	2.899	5.818
INFORM RES	104	0.841	46	2.372	2.490
J INF SCI	366	0.899	44	2.351	2.810
ANNU REV INFORM SCI	296	4.292	12	2.290	1.763
LIBR INFORM SCI RES	214	0.842	25	1.832	1.811
ELECTRON LIBR	55	0.233	47	1.794	1.272
INFORM PROCESS MANAG	992	1.295	54	1.786	2.104
ONLINE INFORM REV	93	0.581	41	1.730	1.793
COLL RES LIBR	384	1.159	26	1.716	1.263
ASLIB PROC	112	0.456	35	1.684	1.408

Table 3. Top 10 journals by Journal Rank.

Abbreviated Journal Title	Total Cites	Impact Factor	Articles	Journal Rank	Influence
SCIENTOMETRICS	860	1.12	89	1.401	6.640
J AM SOC INF SCI TEC	2254	2.086	97	2.899	5.818
KNOWL ORGAN	73	0.444	8	1.163	3.277
INT J GEOGR INF SCI	899	1.234	37	0.341	3.194
J INF SCI	366	0.899	44	2.351	2.810
INFORM RES	104	0.841	46	2.372	2.490
LIBR RESOUR TECH SER	109	0.3	23	1.186	2.328
INFORM PROCESS MANAG	992	1.295	54	1.786	2.104
J DOC	654	1.542	26	1.538	2.003
LIBR INFORM SCI RES	214	0.842	25	1.832	1.811

Table 4. Top 10 journals by Influence.

4. Conclusion

A better analogy might be to substitute a single article for a web page. Then we have article matrix and we can calculate a "paper influence" for individual papers and $C(T_i)$ is just the total number of references a paper makes. We could then calculate the influence of a journal as the average paper influence for the papers published in one year (or other time period). The data for this analysis is more difficult to collect and analyze, particularly for a set of journals in one subject area, because it is not usually tabulated in this way by ISI.

We have shown how journal rank calculated as in Google is a simplified version of the journal influence measure. The information and library science field may not have been the best example as it is very inter-disciplinary and many of the references and citations are to journals outside the group of 54 journals. Even so, the journal rank and influence measures offer two better alternatives for evaluating journals within a field.

References:

Garfield, Eugene. 1996. How can impact factors be improved? *British Medical Journal* 313:411-3.

Brin, Sergey and Page, Lawrence. n.d. The Anatomy of a Large-Scale Hypertextual Web Search Engine. [Accessed Jan. 15, 2006. http://www.db.stanford.edu/~backrub/google.html]

Bordonds, Maria; Fernandez, M.T. and Gomez, Isabel. 2002. Advantages and limitations in the use of impact factor measures for the assesment of research performance in a peripheral country. *Scientometrics* 53(2): 195-206.

Geller, N.L. 1978. On the citation influence methodology of Pinski and Narin. *Information Processing & Management* 14: 93.

Glanzel, W. and Moed, H.F. 2002. Journal impact measures in bibliometric research, *Scientometrics* 53(2): 171-193

Institute of Scientific Information. 2004. *Journal Citation Reports.*

Kleinberg, J.M. 1999. Authoritative sources in a hyperlinked environment, *Journal of the ACM* 46 (5): 604-632.

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

Seglen, Per O. 1997. Why the impact factor of journals should not be used for evaluating research. *British Medical Journal* 314: 497.

Vinkler, Peter. 2002. Subfield problems in applying the Garfield (Impact) factors in practice. *Scientometrics* 53(2): 267-279.

Yanovsky, V.I. 1981. Citation analysis significance of scientific journals, *Scientometrics* 3(3): 223-233

Appendix 1. Impact Factor, Journal rank and Influence measures for all journals.

	Abbreviated Journal Title	Total Cites	Impact Factor	Articles	Journal Rank	Influence
1	ANNU REV INFORM SCI	296	4.292	12	2.290	1.763
2	ASLIB PROC	112	0.456	35	1.684	1.408
3	CAN J INFORM LIB SCI	52	0.308	6	1.300	0.438
4	COLL RES LIBR	384	1.159	26	1.716	1.263
5	ECONTENT	32	0.068	57	0.150	0.000
6	ELECTRON LIBR	55	0.233	47	1.794	1.272
7	GOV INFORM Q	115	1.125	25	0.768	0.556
8	INFORM MANAGE-AMSTER	1013	1.815	72	1.041	1.356
9	INFORM PROCESS MANAG	992	1.295	54	1.786	2.104
10	INFORM RES	104	0.841	46	2.372	2.490
11	INFORM SOC	219	0.667	26	1.108	0.360
12	INFORM SYST J	108	0.727	17	0.436	0.089
13	INFORM SYST RES	615	3.512	20	0.451	0.067
14	INFORM TECHNOL LIBR	82	0.24	18	0.635	0.253
15	INT J GEOGR INF SCI	899	1.234	37	0.341	3.194
16	INT J INFORM MANAGE	234	0.371	35	1.009	0.389
17	INTERLEND DOC SUPPLY	81	0.537	34	1.128	1.389
18	J ACAD LIBR	282	1.034	56	1.601	1.442
19	J AM MED INFORM ASSN	1468	2.612	61	0.361	0.523
20	J AM SOC INF SCI TEC	2254	2.086	97	2.899	5.818
21	J DOC	654	1.542	26	1.538	2.003
22	J GOV INFORM	58	0.227	30	0.598	1.346
23	J HEALTH COMMUN	201	0.671	46	0.150	1.000
24	J INF SCI	366	0.899	44	2.351	2.810
25	J INF TECHNOL	217	0.85	27	0.295	0.103
26	J INFORM ETHICS	19	0.217	8	0.185	0.002
27	J LIBR INF SCI	64	0.514	9	0.474	0.136
28	J MANAGE INFORM SYST	944	1.271	35	0.466	0.255
29	J MED LIBR ASSOC	96	0.92	51	1.061	1.747
30	J SCHOLARLY PUBL	19	0.172	19	0.375	0.128
31	KNOWL ORGAN	73	0.444	8	1.163	3.277

32	LAW LIBR J	147	0.414	32	0.407	0.598
33	LIBR COLLECT ACQUIS	58	0.386	34	1.128	0.479
34	LIBR INFORM SCI	54	0.167		0.728	0.217
35	LIBR INFORM SCI RES	214	0.842	25	1.832	1.811
36	LIBR J	362	0.265	146	0.230	0.040
37	LIBR QUART	207	0.933	12	0.970	0.488
38	LIBR RESOUR TECH SER	109	0.3	23	1.186	2.328
39	LIBR TRENDS	284	0.517	50	1.474	1.454
40	LIBRI	86	0.327	25	1.673	1.085
41	MIS QUART	1869	2.884	24	0.573	0.212
42	ONLINE	104	0.212	28	0.232	0.010
43	ONLINE INFORM REV	93	0.581	41	1.730	1.793
44	PORTAL-LIBR ACAD	66	0.541	28	1.154	0.760
45	PROGRAM-ELECTRON LIB	112	0.171	20	0.560	0.311
46	REF USER SERV Q	89	0.446	12	0.579	0.259
47	RES EVALUAT	53	0.371	13	0.229	0.662
48	RESTAURATOR	121	0.575	23	0.150	0.444
49	SCIENTIST	396	0.247	213	0.150	0.013
50	SCIENTOMETRICS	860	1.12	89	1.401	6.640
51	SOC SCI COMPUT REV	191	0.687	40	0.670	0.274
52	SOC SCI INFORM	218	0.174	23	0.150	0.001
53	TELECOMMUN POLICY	210	0.329	43	0.204	0.079
54	Z BIBL BIBL	14	0.119	21	0.284	0.066