Reference analysis: a view in the mirror of citation analysis

Published in: Geomatics and Information Science of Wuhan University (2010), Vol. 35, spec. November 2010, p. 6-9

Liming Liang
Institute for Science, Technology and Society
Henan Normal University
Xinxiang, P.R. China
e-mail: liangliming1949@sina.com

Ronald Rousseau
Industrial Sciences and Technology
KHBO
Oostende, Belgium
e-mail: ronald.rousseau@khbo.be

***Abstract*—An indicator framework based on references instead of citations is proposed. The reference factor and the reference-based *h*-index are discussed as examples of new indicators derived from this framework.**

Keywords-references; citations

I. INTRODUCTION: CITATION ANALYSIS

Citation analysis is defined as that subfield of informetrics where patterns and frequencies of citations, given as well as received are analyzed. Such an analysis is performed on the level of authors, journals, scientific disciplines and any other useful unit or level. Citation analysis further studies relations between cited and citing units (documents, authors, countries etc.). From an application point of view citation analysis may be considered as a collaborative peer effort to analyze and promote the quality of scholarly publication and research [1], [2]. Although this definition clearly states that citation analysis studies citations given (this is: references) as well as received, it is a fact that the majority of articles written in this field deals with citations received. The different types of journal impact factors [3], the h-index [4] and its applications in different fields (such as the journal [5] or country h-index), as well as its generalizations and variants (such as the g-index [6], the A-index [7] and the R-index [8]) are all based on received citations.

II. RECENT DEVELOPMENTS

Two articles are bibliographically coupled if they have references in common [9], [10]. This notion, based on the notion of 'references', is one of the basic concepts in the field and as such dates from more than fifty years ago. Yet, its dual notion, namely cocitation [11], [12] has been used and studied in much more articles. A simple search in the Web of Science revealed that the ratio is 1 to 10. Recently, however, more attention has been given to bibliographic coupling [13] and to references in general. Indeed, new proposals for journal impact factors make explicit use of the numbers of references [14], [15], [16], [17]. In [18] we already showed how references can be used in citation analysis, in particular in the framework of rhythm studies [19]. In this contribution we go one step further and show that basically everything that has been studied for received citations can also be studied for references.

III. A BASIC FRAMEWORK FOR REFERENCES

We consider a journal B and form its publication-reference matrix (p-r matrix). Recall that the publication-reference matrix can be considered as the mirror image of the better known publication-citation (p-c) matrix. As presented in Table 1 it consists of rows of publication years shown from top to bottom in anti-chronological order. The first column gives the actual years, while the second column is a row counter. The third column gives the number of publications published in the year shown in the first column, possibly restricted to certain types of publications, e.g. normal articles, reviews, contributions in conference proceedings, and so on. The next column gives numbers of references (=cited articles). Cited years are numbered in the same way as publication years. R_{ij} denotes the number of references (in journal B) published themselves in year j, and referred to (= cited) in journal B in year i. Hence increasing row and column numbers refer to older publications. Usually (for logical reasons) $R_{ij} = 0$ if $i > j$. Instead of ranks as indices we will occasionally also use actual years as indices. If year Y corresponds to row rank i then R_{ij} corresponds to $R_{Y,Y+i-j}$.

Similar to the relative citation impact factor we introduced a w-year relative synchronous reference factor for the year Y and journal B, denoted as $REF_w(Y,B)$ as:

$$REF_w(Y,B) = \frac{\sum_{k=1}^{w} R_{Y,Y-k+1-s}}{P_B(Y)}$$

where R-indices are years; $P_B(Y)$ denotes the number of articles published in year Y in journal B and w denotes the window. Usually we do not include the letter B, when the journal is clear from the context or when this is of no importance. The w-year relative reference factor is the average number of references per article, included in journal

B in the year Y, that are published between the years Y-w+1-s (the oldest year in the window) and the year Y-s (the most recent year in the window), where s is a natural number indicating the ending of the window.

TABLE I. p-r matrix of journal B

publication year *i* and number of publications *P_i* of journal B			cited year *j* (reference's publication year) and R_{ij}								
			2010	2009	2008	2007	2006	2005	2004	2003	2002
			1	2	3	4	5	6	7	8	9
2010	1	P_1	R_{11}	R_{12}	R_{13}	R_{14}	R_{15}	R_{16}	R_{17}	R_{18}	R_{19}
2009	2	P_2		R_{22}	R_{23}	R_{24}	R_{25}	R_{26}	R_{27}	R_{28}	R_{29}
2008	3	P_3			R_{33}	R_{34}	R_{35}	R_{36}	R_{37}	R_{38}	R_{39}
2007	4	P_4				R_{44}	R_{45}	R_{46}	R_{47}	R_{48}	R_{49}
2006	5	P_5					R_{55}	R_{56}	R_{57}	R_{58}	R_{59}
2005	6	P_6						R_{66}	R_{67}	R_{68}	R_{69}
2004	7	P_7							R_{77}	R_{78}	R_{79}
2003	8	P_8								R_{88}	R_{89}
2002	9	P_9									R_{99}

A generalized impact factor calculated over the same window can be defined as (using the notation of [3]):

$$IF_B(1,w,Y,Y-w+1-s) = \frac{\sum_{k=1}^{w} Cit_{Y,Y-k+1-s}}{P_B(Y)}$$

where Cit $_{Y, Y-k+1-s}$ denotes the number of citations received by journal B in year Y to articles published during the publication window of length w. What could be the meaning of the reference factor, $REF_w(Y,B)$? When is it high and when is it low? What is its relation to the impact factor $IF_B(1,w,Y,Y-w+1-s)$?

Clearly $\frac{IP_B(1,w,Y,Y-w+1-s)}{REF_w(Y,B)} = \frac{\sum_{k=1}^{w} Cit_{Y,Y-k+1}}{\sum_{k=1}^{w} R_{Y,Y-k+1}}$. As $REF_w(Y,B)$ is a ratio it is determined by two elements: the numerator and the denominator. From the denominator we see that the larger the number of published articles the lower *REF* and vice versa, the lower the number of published articles the higher *REF* (keeping the numerator fixed). Such a reference index only differs from the impact factor $IF_B(1,w,Y,Y-w+1-s)$ through its numerator: received citations are replaced by references. Clearly one expects review journals to have a high *REF*-value. Similarly, based on the numerator, letter journals, and generally journals in fields that develop fast are expected to have a high *REF*-value, especially if the window is short and referencing occurs to relatively recent publications. Note that R in Table I refers to a number of publications (articles and other) published in many different sources. They are all cited in journal B (in year Y). The symbol C refers to citations of journal B, given in many different journals. In [18] we showed that a p-r matrix has a constant rR-sequence (with window w) if and only if its w-year relative reference factor *REF* is constant during the period under study. For the definition and further elaboration of rR-sequences we refer to [19].

IV. An h-index type indicator in the reference framework

Using the same construction method as for the original *h*-index [4] we define a journal *h*-index based on publication and reference data, denoted as Hr. Concretely, Hr for journal J in publication year Y is defined as the natural number Hr such that Hr publications have at least Hr references while the other journal publications have no more than Hr reference**s**. Similarly a page-based *h*-index, denoted as Hp, can be defined as the number Hp such that Hp publications are at least Hp pages long while the other journal publications are not more than Hp pages long. Pages can be measured as natural numbers or as rational numbers. In our examples we will use natural numbers.

V. Examples

A. Reference factors. Using a two year window (w=2) over the years [Y-2 ; Y-1] , i.e. taking s=1, we calculated the relative reference factor for the year 2008 for twenty Chinese journals (those with the highest impact factor**s**) included in the Web of Science and compare these with the standard two-year impact factor, denoted as IF, see Table II.

Somewhat surprisingly the Pearson correlation between the standard impact factor and this reference factor is quite high, namely 0.88. As we know that, as a general rule, journals in fast-developing fields have higher impact factors than journals in more stable and slowly developing fields, this fact probably explains the high correlation.

We also calculated this reference factor for three famous multidisciplinary journals, see Table III. These journals' impact factors are much higher than for the Chinese journals, and also the reference factors are somewhat higher.

TABLE II. Chinese journals

Journal	IF	REF	REF/IF
CELL RESEARCH	4.535	14.569	3.212
COMMUN COMPUT PHYS	2.330	5.061	2.172
FUNGAL DIVERS	2.279	7.627	3.347
WORLD J GASTROENTERO	2.081	7.319	3.517
ASIAN J ANDROL	2.059	6.620	3.215
TRANSPORTMETRICA	2.043	4.571	2.238
EPISODES	1.870	6.236	3.335
ACTA PHARMACOL SIN	1.676	5.850	3.491
ACTA GEOL SIN-ENG	1.431	3.788	2.647
ACTA PHYS SIN-CH ED	1.165	3.437	2.950
ACTA BIOCH BIOPH SIN	1.086	6.667	6.139
SCI CHIN SER G	0.973	3.335	3.428
CHINESE J CHEM	0.945	4.266	4.514
J MATER SCI TECHNOL	0.869	4.678	5.383
PEDOSPHERE	0.865	1.789	2.068
J INTEGR PLANT BIOL	0.859	5.112	5.951
CHINESE MED J-PEKING	0.858	4.923	5.738
J COMPUT MATH	0.765	2.582	3.375
ACTA MECH SINICA	0.764	3.514	4.599
CHINESE PHYS LETT	0.743	3.369	4.534

Because of this their REF to IF ratio is much smaller than for the Chinese journals.

TABLE III. Multidisciplinary journals

Journal	IF	REF	REF/IF
NATURE	31.434	10.328	0.329
SCIENCE	28.103	9.607	0.342
PNAS	9.380	8.404	0.896

TABLE IV. H-type indices for the journal *Scientometrics*

year	Group 1: Counting based on all documents			Group 2: Counting based on articles, proceeding papers and reviews		
	P	Hr	Hp	P	Hr	Hp
1978	7	5	5	5	5	5
1979	23	11	13	20	11	13
1980	39	15	12	33	14	12
1981	37	13	13	29	13	13
1982	37	13	13	25	13	13
1983	48	14	12	25	14	12
1984	44	16	11	25	15	11
1985	62	19	17	53	19	17
1986	50	17	14	40	17	14
1987	69	17	16	46	16	16
1988	61	18	17	49	18	17
1989	79	20	16	64	20	16
1990	70	19	18	61	19	17
1991	83	21	20	78	20	20
1992	82	20	20	76	20	20
1993	62	21	20	58	21	20
1994	96	22	19	59	22	19
1995	83	21	20	71	21	19
1996	99	21	20	87	21	20
1997	81	22	21	75	22	21
1998	89	21	20	84	21	20
1999	140	23	21	128	23	20
2000	89	26	22	82	26	22
2001	115	24	20	91	24	20
2002	87	26	19	84	26	19
2003	94	27	21	83	27	21
2004	101	28	19	89	28	19
2005	129	32	23	113	32	23
2006	145	36	22	136	35	22
2007	129	36	25	129	36	25
2008	131	32	22	128	32	22
2009	192	36	22	189	36	22

B. Hirsch indices. We calculated yearly Hr and Hp indices for the journal *Scientometrics*. Table IV shows two groups of indicators: one based on all published documents and one based on articles, proceedings papers and reviews. There is almost no difference between the two groups of data. The reason is that other types of publications such as editorial material, book reviews, notes and bibliographies are usually short and contain few references. Table V. gives correlation coefficients between the P, Hr and Hp time sequences, while Fig. 1 illustrates the time evolution of the Hr and Hp sequences. This figure illustrates the well known fact that reference lists grow much faster than the number of pages of articles.

TABLE V. Correlation coefficients

based on all documents			based on articles, proceeding papers and reviews		
(Hr,Hp)	(Hr,P)	(Hp,P)	(Hr,Hp)	(Hr,P)	(Hp,P)
0.88	0.92	0.83	0.89	0.92	0.83

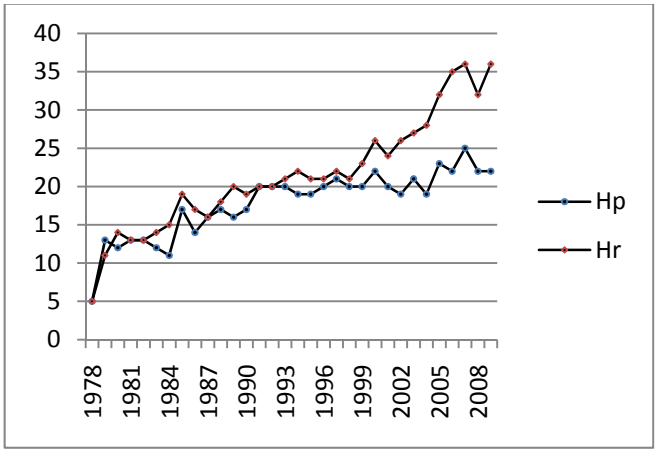

Figure 1. Comparison of Hr and Hp
(counting based on articles, proceedings papers and reviews)

VI. Conclusion

This presentation surely contributes to the profusion of measures [20], leading to the question why reference based measures are useful and what they exactly represent [20]. We have provided some answers to these questions but much more needs to be done. Yet, we are convinced that our mirror image of classical citation analysis will contribute to a better understanding of the structure of science.

Acknowledgment

Liming Liang's work is sponsored by the NSFC project 70973118 .

REFERENCES

[1] E. Spinak, Diccionario Enciclopédico de Bibliometría, Cienciometría e Informetría: Caracas: UNESCO-CII/II, 1996.

[2] R. Rousseau, "Publication and citation analysis as a tool for information retrieval," in Social Information Retrieval Systems. Emerging technologies and applications for searching the web effectively, D.Goh and S. Foo, Eds. Hershey (PA): Information Science Reference (IGI Global), 2008, pp. 252-267.

[3] T.F. Frandsen and R. Rousseau, "Article impact calculated over arbitrary periods," Journal of the American Society for Information Science & Technology, vol.56, 2005, pp.58-62.

[4] J.E. Hirsch, "An index to quantify an individual's scientific research output," Proceedings of the National Academy of Sciences of the USA, vol. 102, 2005, pp. 16569-16572.

[5] T. Braun, W. Glänzel and A. Schubert, "A Hirsch-type index for journals," The Scientist, vol.19, issue 22, 2005, p.8.

[6] L. Egghe, "An improvement of the H-index: the G-index," ISSI Newsletter, vol.2, issue 1, 2006, pp.8-9.

[7] BH. Jin, "The AR-index: complementing the h-index," ISSI Newsletter, vol. 3, issue 1, 2007, p.6.

[8] BH. Jin, LM. Liang, R. Rousseau and L. Egghe, "The R- and AR-indices: complementing the h-index," Chinese Science Bulletin, vol. 52, 2007, pp.855-863.

[9] R.M. Fano, "Information theory and the retrieval of recorded information,"in Documentation in Action, J. H. Shera, A. Kent, andJ.W. Perry Eds. New York: Reinhold Publ. Co. ,1956, pp. 238–244.

[10] M.M. Kessler, "Bibliographic coupling between scientific papers," American Documentation, vol.14, 1963, pp.10-25.

[11] H. Small, "Co-citation in the scientific literature: a new measure of the relationship between two documents," Journal of the American Society for Information Science, vol. 24, 1973, pp.265-269.

[12] I.V. Marshakova. "System of document connections based on references (in Russian), "Nauchno-Tekhnicheskaya Informatsiya, ser.2, vol. 6, 1973, pp.3-8.

[13]. DZ. Zhao and A.Strotmann, "Evolution of research activities and intellectual influences in information science 1996-2005: introducing author bibliographic coupling analysis," Journal of the American Society for Information Science and Technology, vol. 59, 2008, pp. 2070-2086.

[14] M. Zitt and H. Small, "Modifying the journal impact factor by fractional citation weighting: The audience factor," Journal of the American Society for Information Science and Technology, vol. 59, 2008, pp.1856-1860.

[15] J. Nicolaisen and T.F.Frandsen, "The reference return ratio," Journal of Informetrics, vol.2, 2008, pp.128-135.

[16] H.F. Moed, "Measuring contextual citation impact of scientific journals," Journal of Informetrics, vol. 4, 2010, pp.265-277.

[17] R. Guns and R. Rousseau, "New journal impact indicators take references into account: a comparison," ISSI Newsletter, vol.6, issue1, 2010, pp. 9-14.

[18] LM. Liang and R. Rousseau, "Measuring a journal's input rhythm based on its publication-reference matrix," Journal of Informetrics, vol.4, 2010, pp. 201-209.

[19] LM. Liang, "The R-sequence: a relative indicator for the rhythm of science," Journal of the American Society for Information Science & Technology, vol. 56, 2005, pp. 1045-1049.

[20] R. Van Noorden, "A profusion of measures," Nature, vol. 465, 2010, pp. 864-866.