A Download $h^{(2)}$-Index as a Meaningful Usage Indicator of Academic Journals

Ping-huan Hua[1], Ronald Rousseau [2, 3], Xiu-kun Sun[1], Jin-kun Wan[1]

[1]*gfhs@cnki.net*
China Scientometrics and Bibliometrics Research Center of Tsinghua University,
Beijing 100084, China

[2] *ronald.rousseau@khbo.be*
[2]KHBO (Association K.U.Leuven), Industrial Sciences and Technology, B-8400, Oostende, Belgium
[3]K.U.Leuven, Dept. Mathematics, Celestijnenlaan 200B, 3001 Leuven (Heverlee), Belgium

Abstract

The number of downloads can be considered as representing usage as well as social impact of scientific journals. In this context the download $h^{(2)}$ index – an adaptation of Kosmulski's $h^{(2)}$ index - is suggested as a new indicator. The download $h^{(2)}$ index of six thousand Chinese academic journals has been determined. We compare the download $h^{(2)}$ with other indicators such as the number of downloads and the download immediacy index and highlight advantages of using the download $h^{(2)}$ index. It is shown that this new index is a number that agrees well with high volumes of data.

Introduction

In the era of electronically available journals, Open Access (OA) or not, academic articles can easily be scanned and downloaded by readers. When the contents of an article have been absorbed by the reader the article is cited (or not) in later publications, in the same way as paper based articles. Clearly, because of their web presence it is nowadays easier than ever before for academic journals to exert scholarly and social impact.

Proposing simple transformations of existing indicators for the evaluation of web impact of academic journals is not difficult, cf. the definition of the web impact factor, although the practical application of such proposals may lead to unforeseen difficulties (Ingwersen, 1998; Noruzi, 2006). Some indicators do not even have to be transformed at all: a classical journal impact factor stays the same, whether or not the journal publisher makes articles available on paper, in electronic form, or both. An example of a simple transformation, or adaptation, is the use of a download immediacy index as studied in (Wan *et al.*, 2009). Recall that the journal download immediacy index (DII) is defined as the number of downloads of a journal's articles within one publication year, divided by the number of published articles by that journal in that same year (Wan *et al.*, 2007). In this adaptation downloads simply replace citations. Other download impact factors can similarly be defined.

However, using a simple analogue of a classical indicator may sometimes be questionable. Although clear parallels exist between downloads and citations, the two are not the same. Of course, it is well-known that downloads can be a first step towards citations (Moed, 2005). Downloading and citing of electronically available

journals and articles are different acts. The main differences between the two are that downloading occurs on a much higher scale than citing and with a much shorter response time (response to the act of making article content public). Because of these differences we propose a new *h*-type index for article downloads, complementing the journal *h*-index (based on citations).

Downloading in the CNKI

Download data used in this paper are obtained from the Chinese National Knowledge Infrastructure (CNKI) (Wan & Zhang, 2004). A short overview of the CNKI is given in (Wan *et al.*, 2009). The main body of the CNKI database are the *China Knowledge Resources Integral Databases*, containing information related to more than seven thousand Chinese academic journals. At present, CNKI contains more than 22 million journal papers. In 2006 each day about 360,000 papers published in the period 2004─2006 were downloaded from the central servers at CNKI, where downloads at mirror sites are not even taken into account. In that same year over 1.78 million papers published in 2006 and included in the database have been downloaded a total of 48.87 million times. This means that one such article is downloaded on average 27 times. Obviously the number of downloads is an order of magnitude higher than the number of citations received, see also (Wan *et al.*, 2009).

Besides sheer numbers of downloads, another characteristic of web-based journal data is the response time, which is typically much shorter than for citations. Finally the temporal distribution of downloads and citations differs. This is illustrated in Fig.1, based on data collected in 2006 (downloading occurred during the year 2006). The numbers on the y-axis *y* are relative numbers of actions (downloads or citations), expressed as percentages; the x-axis refers to publication years.

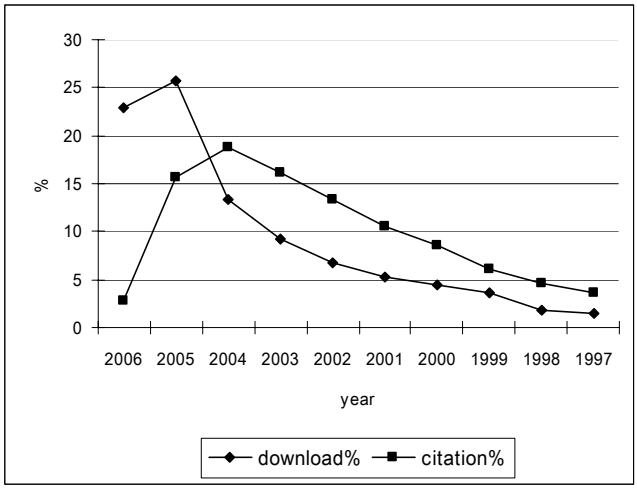

Figure 1. Comparison between the distribution of the number of downloads and the total cites distribution (in percentages)

Figure 1 clearly shows that the peak of the download distribution curve is higher and comes earlier than the peak of the citation curve. The most recent papers are the ones that are most downloaded, illustrating the quick response time phenomenon. This

corroborates the observations made in (Wan *et al.*, 2005). A reviewer pointed out that, although downloading and citing can both be considered as forms of use, the former is also a form of information seeking behaviour, while the latter reflects integration of the existing literature into one's own work. These are two totally different forms of use.

Definition and calculation of the download $h^{(2)}$ index

The *h*-index proposed by J.E. Hirsch (Hirsch, 2005) has already attracted an immense interest from the informetric community in the world. Introduced as a single scientist's total career indicator, the *h*-index concept has in recent years been applied to scientific achievements of groups of scientists, academic journals, organizations, countries, patents, and even outside a publication-citation context (Liu & Rousseau, 2007, 2009). Let us recall that a journal's *h*-index for the publication year Y is equal to *h* if the *h* most-cited articles received each at least *h* citations, while the article ranked *h+1* received at most *h* citations. Replacing citations by downloads in the definition of a journal's *h*-index yields this journal's download *h*-index.

As a bibliometric indicator, the *h*-index has its shortcomings and limitations; consequently several modifications have been proposed. For a review we refer to (Egghe, 2009). Among these modifications we mention Kosmulski's $h^{(2)}$-index. A scientist's $h^{(2)}$ index is defined as the highest natural number k such that his k most-cited publications received each at least k^2 citations (Kosmulski, 2006). Similarly, we define the download $h^{(2)}$ index of a journal in the year Y as the number $h^{(2)}$ such that $h^{(2)}$ is the highest rank such that the first $h^{(2)}$ articles published in this journal in the year Y, have been downloaded each at least $[h^{(2)}]^2$ times. Note that this definition is of the 'immediacy' type. Other versions of download $h^{(2)}$ indices can similarly be defined.

The calculation of a journal's download $h^{(2)}$ index is relatively easy. We explain it by considering the example of the journal *Modernizing Agriculture*, one of the journals included in the CNKI database. In 2006, 377 papers were published in this journal. Among these, forty were not downloaded at all. The other ones have been downloaded a total of 5,330 times. Table 1 shows the number of downloads of the most downloaded articles (n_{down}). Ranks (No) and squared ranks (No^2) are shown too. Inspecting the data of Table 1 from the top down, it is easily seen that the download $h^{(2)}$ index is 10, as there are 10 papers, each of which have been downloaded at least 100 times, and no 11 articles that each have been downloaded at least $11^2 = 121$ times. Consequently, this journal's download $h^{(2)}$ index is 10. Table 1 also shows that *Modernizing Agriculture*'s download *h*-index is 33.

The download *h*-index and $h^{(2)}$ index can also be illustrated visually. This is done in Fig. 2. The download *h*-index is the abscissa of the intersection of the download curve

and the line $y = x$; the download $h^{(2)}$ index is the value of the abscissa of the intersection of the download curve and the curve $y = x^2$.

Table 1 Calculating the download *h*-index and $h^{(2)}$-index

No	$(No)^2$	n_{down}
1	1	254
2	4	218
3	9	170
4	16	169
5	25	139
6	36	127
7	49	122
8	64	121
9	81	118
10	100	117
11	121	112
……		
31	961	35
32	1024	34
33	1089	34
34	1156	32
35	1225	32

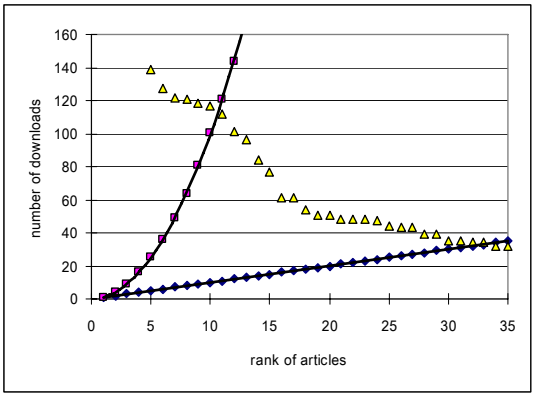

Figure 2. The download *h*-index and $h^{(2)}$ index of *Modernizing Agriculture*

The download $h^{(2)}$ index of Chinese academic journals

The results of determining the download $h^{(2)}$ index of 6,400 Chinese academic journals are shown in Figure 3. It is clear that the distribution of the $h^{(2)}$-values follows approximately a normal curve, with top near $h^{(2)} = 8$ (see Fig. 4). Table 2 gives data regarding the download *h*-index, the download $h^{(2)}$-index, the number of downloads (n_{down}, in the same year) and the download immediacy index (DII) of 39 agricultural journals with $h^{(2)}$-value larger than 10. This means: we show those journals that have each at least 11 articles that have been downloaded each at least 121 times.

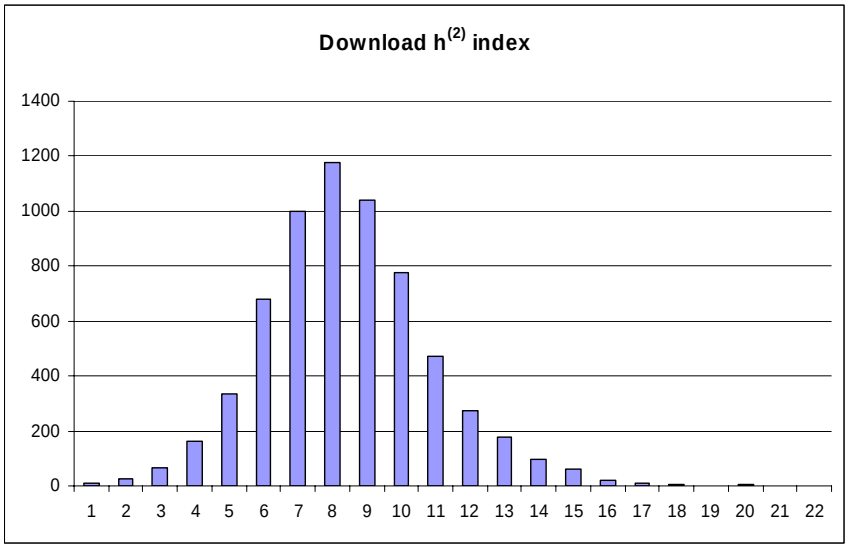

Figure 3. The distribution of the download $h^{(2)}$ index of Chinese academic journals

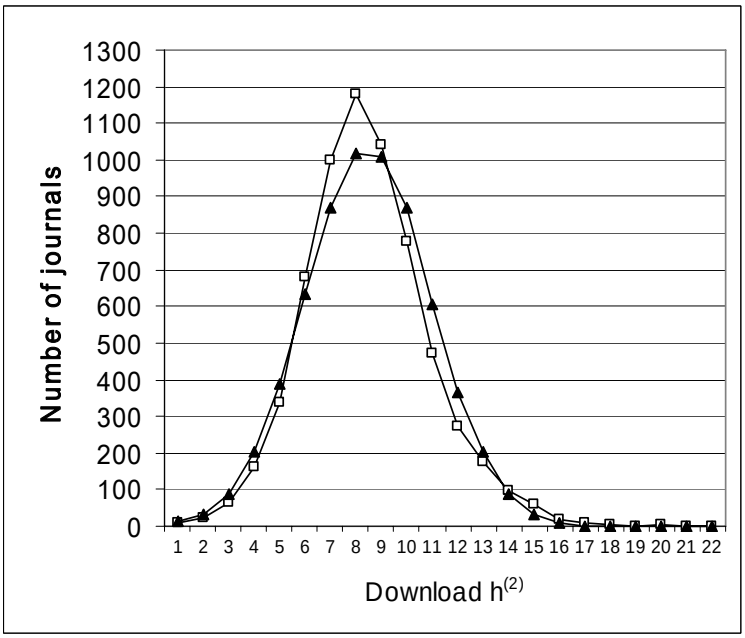

Figure 4. Best fitting normal curve (open squares) for the download data shown in Fig. 3

Table 2. 39 agricultural academic journals ($h^{(2)}$ > 10), 2006 data

journal	$h^{(2)}$	n_{down}	h	DII	ranks			
					$h^{(2)}$	n_{down}	h	DII
Chinese Agricultural Science Bulletin	15	68,925	112	48.6	1	3	1	24
Gansu Nongye	14	94,882	108	35.1	2	1	2	62
Journal of Agricultural Mechanization Research	14	34,843	86	43	2	5	8	35
Journal of Anhui	13	74,975	96	25.8	4	2	5	120

journal	$h^{(2)}$	n_{down}	h	DII	ranks			
					$h^{(2)}$	n_{down}	h	DII
Agricultural Sciences								
Transactions of the Chinese Society of Agricultural Engineering	13	36,098	98	61.8	4	4	4	8
Scientia Agricultura Sinica	13	29,299	101	74.4	4	6	3	2
Agro-Environmental Protection	13	27,669	90	58.1	4	7	6	13
Acta Horticulturae Sinica	13	22,169	89	71.3	4	8	7	4
Chinese Journal of Eco-agriculture	13	16,434	71	58.1	4	16	11	14
Chinese Journal of Soil Science	13	15,656	75	67.2	4	18	10	7
Acta Agronomica Sinica	12	21,345	86	69.3	11	9	9	5
Research of Soil and Water Conservation	12	20,499	71	37.1	11	10	12	51
Progress In Veterinary Medicine	12	16,746	71	44.1	11	15	13	32
Journal of Soil and Water Conservation	12	14,041	70	58.5	11	21	14	11
World Agriculture	12	13,282	67	45.5	11	28	18	29
Transactions of the Chinese Society of Agricultural Machinery	11	20,412	69	36	16	11	16	58
Feed Industry	11	18,924	69	36.7	16	12	17	52
Xiandai Nongye Keji	11	18,815	55	14.3	16	13	37	263
Chinese Journal of Animal Science	11	16,041	63	31.2	16	17	20	86
Journal of Henan Agricultural Sciences	11	15,562	67	34.1	16	19	19	68
Livestock and Poultry Industry	11	13,722	55	22.5	16	24	38	152
Journal of Maize Sciences	11	13,639	61	40.4	16	26	25	41
Journal of Fruit Science	11	12,889	70	56.3	16	29	15	17
Seed	11	12,514	61	32.3	16	31	26	79
China Animal Husbandry & Veterinary Medicine	11	12,480	54	31.7	16	32	42	84
Agricultural Research in the Arid Areas	11	11,963	60	42.7	16	33	27	36
Jiangsu Agricultural Sciences	11	11,824	56	34.5	16	35	32	66
Pesticides	11	11,742	58	38.6	16	36	29	44
Shaanxi Journal of Agricultural Sciences	11	11,195	55	30.3	16	37	39	94
Acta Agriculturae Boreali-occidentalis Sinica	11	10805	54	31.2	16	40	43	87
China Vegetables	11	10,188	51	30.8	16	42	50	89
Chinese Countryside Well-off Technology	11	9,936	47	23.7	16	45	64	142

journal	$h^{(2)}$	n_{down}	h	DII	ranks			
					$h^{(2)}$	n_{down}	h	DII
Acta Pedologica Sinica	11	9,424	63	68.3	16	48	21	6
Plant Nutrition and Fertilizer Science	11	9,227	63	59.5	16	50	22	10
Soil and Water Conservation In China	11	9,128	53	36.7	16	51	46	53
Research of Agricultural Modernization	11	7,865	54	74.2	16	70	44	3
Acta Phytopathologica Sinica	11	6,682	56	79.5	16	83	33	1
Chinese Journal of Rice Science	11	6,166	51	56.6	16	95	51	16
Forest Engineering	11	5,477	42	44.5	16	114	77	31

Comparison of the download $h^{(2)}$ with other bibliometric indicators

As a first step we calculate the Pearson correlation between the download $h^{(2)}$-index, the number of downloads, the citation h-index and the download immediacy index (DII), for the 39 agricultural journals shown in Table 2. Results are shown in Table 3.

Table 3. Pearson correlation coefficients

	$h^{(2)}$	n_{down}	h	DII
$h^{(2)}$	1	0.768	0.887	0.273
n_{down}		1	0.818	-0.103
h			1	0.308
DII				1

Clearly, the download $h^{(2)}$-index, the number of downloads and the h-index are highly correlated. The DII clearly measures another property of journal publications.

We comment now on some relations between indicators

-) The download $h^{(2)}$-index and the number of downloads in the current year (n_{down})

The download $h^{(2)}$ is an indicator combining quality (or at least visibility, or attractivity) with quantity, while the number of downloads (n_{down}) is an attractivity indicator for the journal as a whole. Hence, these two indices contain an element of attractivity. The number of downloads, however, makes no difference between journals that have published a quite different number of articles. Consequently, it is heavily influenced by the number of published articles, and hence it is mainly a quantity (of articles) indicator. In order to determine the download $h^{(2)}$-index of a journal one needs the complete number of downloads for all articles published in the journal (otherwise one cannot rank them). Yet, once ranked it is very easy to determine the $h^{(2)}$-index. It is, moreover, well-known that the h-index (and hence also the $h^{(2)}$-index) is very robust for small errors (Vanclay, 2007; Rousseau, 2007). On the one hand, articles in the tail (with no or very few downloads) do not play a role in the

determination of the $h^{(2)}$-index. Hence no checks are necessary to see it the number of downloads is exactly 1 or maybe 2. On the other hand, it does not matter if the most downloaded article is downloaded 5,000 times or 5,010 times. Also this has no influence at all on the final result.

-) Comparison of the download $h^{(2)}$-index and the download immediacy index, DII.
The download immediacy index, DII, is a quality (visibility) indicator for the journal as a whole. It reflects the average number of downloads in the current year. Two journals may have the same DII but, due to the difference in published number of articles, a large difference in the number of downloads. It may be argued that the one with the higher number of downloads has a larger impact on current scientific activities. For example, the two journals *Scientia Agricultura Sinica* and *Research of Agricultural Modernization* in Table 2, have almost the same download immediacy index, but the ratio of the number of downloads in the current year is about 4 to 1. Also the Pearson correlation coefficient given in Table 3 shows that the download $h^{(2)}$-index and DII measure different journal characteristics.

-) Comparison of the download $h^{(2)}$ and the download *h-index*
The citation *h-index* is a good bibliometric indicator for evaluating the citation impact of journals and in situations where the *h*-index is rather small, the $h^{(2)}$ – index, being even smaller, does not discriminate among sources (Liu & Rousseau, 2007). In the case of downloads, however the situation is different. Download *h*-indices are sometimes quite high. In such cases the download $h^{(2)}$-index has the advantage of yielding smaller values, which are, perhaps, easier to remember or seem more meaningful. Another advantage of the download $h^{(2)}$-index is the fact that a journal's *h*-index cannot be larger than the number of published articles. As downloads are often quite high this happens occasionally, as shown in Table 4. The download $h^{(2)}$-index has no problem with such situations. Recall than in cases the *h*-index is equal to the number of published articles, the number of downloads of the h^{th} ranked article is called a pseudo *h*-index (Rousseau *et al*., 2008). Recall also that we have considered downloads over one year. If we extended the download window the problem of having to cope with pseudo *h*-indices would increase even more.

Table 4. Journals for which the download *h*-index is equal to the number of published papers

A: Journal name	B: # publications	C: # downloads	D: h-index	E: h(2) index
Chinese Criminal Science	37	3,848	37	10
Accounting Research	65	25,448	65	20
Economic Research Journal	54	24,465	54	22

A: Journal name; B: number of published articles; C: n_{down};
D: download *h*; E: download $h^{(2)}$

-) If it were possible to link to individual articles, a variant of the idea of a web impact factor (Ingwersen, 1998; Noruzi, 2006) could be used to characterize the impact of web-based articles. As this is either impossible, or not done on a large scale, this approach does not (yet) lead to a useful indicator.

Advantages of the download $h^{(2)}$-index

- Like most other *h*-type indices it combines quantity and quality.
- The precision problem (needing the exact number of downloads for a large number of articles) is small, and smaller than for the download *h*-index.
- As the number of downloads can be very high the resulting number is smaller and more easily remembered than that of the corresponding download *h*-index. One may say that it is a number that agrees well with high volumes of data.

The download $h^{(2)}$-index is field dependent

Every two months, the *Essential Science Indicators* lists a new crop of what it calls hot papers in science. Hot papers are selected by virtue of being cited among the top 0.1% in a current period of two months. Papers are selected in each of 22 fields of science and must be published within the latest two years. In a similar vein we define 'hot downloads' as articles that have been being downloaded among the top 0.1% in a current year period. In 2006, more than 1.78 million papers have been published; hence the number of hot papers, according to this definition, is 1,780. The top 12 papers are shown in Table 5. The most downloaded article was downloaded 1,417 times; the 12th 831 times.

Table 5. Hot downloads in 2006

Rank	Title	Journal	n_{down}
1	Incomplete Contracting Theory: A Survey	Economic Research Journal	1,417
2	Approaches to Literature Survey	Economic Research Journal	1,261
3	The Survey About Theory Of The Optimal Shareholder Structure Design And Power Balance With Shareholder Structure	Accounting Research	1,117
4	Summarization and Application on the Development of Artificial Intelligence Technology	Programmable Controller & Factory Automation	1,086
5	Comments on Reform of RMB Exchange Rate Regime	Studies of International Finance	1,006
6	Several Discriminations in Rural Tourism, Agriculture Tourism and Folklore Tourism	Tourism Tribune	996
7	Civil Society in China: Concepts, Classification and Institutional Environment	Social Sciences in China	894
8	Corporate Governance and Entrepreneurship	Economic	885

		Research Journal	
9	Compensation System Design: How to Connect With Employee Motivation?	New Capital	870
10	Collective Action, Free-rider Problem, and the Application of Formal Models in Social Sciences	Sociological Studies	868
11	The Influence of the Renminbi Appreciation on International Trade and Countermeasures	Journal of the Postgraduate of Zhongnan University of Economics and Law	857
12	Reflections on the newly Issued Accounting Standards	Accounting Research	831

Table 5 shows that the hot downloads belong mostly to the social sciences. This fact shows that the values of the download $h^{(2)}$-index are subject dependent. The average download $h^{(2)}$-indices of various subject areas are shown in Figure 5. From this figure we can see that the download $h^{(2)}$-index of finance and economics journals is the highest (13.1), while a small subfield such as silkworm and beekeeping journals has the lowest (only 4.8). One might have expected that this list would be dominated by biomedical articles. That this is not the case is a result of the fact that the CNKI database does not contain many biomedical journals (and it does not include the top Chinese biomedical journals).

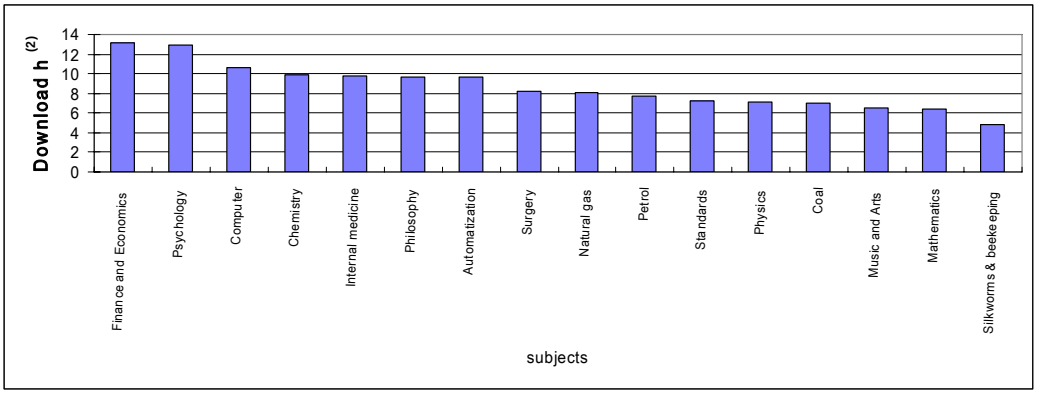

Figure 5. The average download $h^{(2)}$-index in various subjects

h-Type indices are field dependent, and the download $h^{(2)}$-index is no exception. Maybe it might be interesting to try to apply the Molinari approach (Molinari & Molinari, 2008) or Radicchio *et al.*'s idea (Radicchi *et al.*, 2008) to a download context. We leave this as a suggestion for further research.

Another suggestion is to calculate rational download $h^{(2)}$-indices. The rational *h*-index has been introduced by Ruane and Tol (Ruane & Tol, 2008) and further studied, e.g. in (Guns & Rousseau, 2009). It helps to discriminate between sources with the same *h-index*. As shown in Table 2, also here there are many journals with the same download $h^{(2)}$-index. If a journal's download $h^{(2)}$-index is equal to $h^{(2)}$, then a rational download $h^{(2)}$-index is defined as $h^{(2)}_{\text{rat}} = h^{(2)} + 1 - \text{m}/(3\ (h^{(2)})^2 + 3\ h^{(2)} + 1)$, where 3

$(h^{(2)})^2 + 3\ h^{(2)} + 1$ is the extra number of downloads needed for the extreme case that the first $h^{(2)}$ articles have each been downloaded exactly $(h^{(2)})^2$ times and all other articles are not downloaded at all, and m is the actual number of extra downloads needed to reach a download $h^{(2)}$-index of $h^{(2)}+1$. For instance if 10, 8, 6, 4 is the number of downloads of the four most downloaded articles of a journal then its download $h^{(2)}$-index is 2. If the number of downloads for the second and the third article become 9 then the download $h^{(2)}$-index becomes 3. Hence four extra downloads are needed. This leads to $h^{(2)}_{\text{rat}} = 2 + 1 - 4/(12+6+1) = 53/19 \approx 2.79$.

Conclusion

The download $h^{(2)}$-index can be used as a bibliometric indicator for the evaluation of the visibility and short-term use of electronically available academic journals. We consider it to be the indicator that is best suited for the specific purpose of characterizing downloads.

Acknowledgements

The authors thank Li Si-zhi, Du Jian and other collaborators for support. RR's work has been supported by the National Natural Science Foundation of China (NSFC Grant No 70673019). They also thank two anonymous reviewers for some useful suggestions.

References

Egghe, L. (2009). The Hirsch-index and related impact measures. *Annual Review of Information Science and Technology*, to appear.

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

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

Ingwersen, P. (1998). The calculation of Web impact factors. *Journal of Documentation*, 54, 236-243.

Kosmulski, M. (2006). A new Hirsch-type index saves time and works equally well as the original *h*-index. *ISSI Newsletter*, 2(3): 4-6.

Liu, YX. & Rousseau, R. (2007). Hirsch-type indices and library management: the case of Tongji University Library. In: *Proceedings of ISSI 2007* (D. Torres-Salinas & H. F. Moed, eds.). Madrid: CINDOC-CSIC, pp. 514-522.

Liu, YX. & Rousseau, R. (2009). Properties of *h*-type indices: the case of library classification categories. *Scientometrics*, 79(2), DOI:10.1007/s11192-009-0415-1 (to appear)

Moed, H.F. (2005). Statistical relationships between downloads and citations at the level of individual documents within a single journal. *Journal of the American Society for Information Science and Technology*, 56, 893-902.

Molinari, J.-F. & Molinari, A. (2008). A new methodology for ranking scientific institutions. *Scientometrics*, 75, 163-174.

Noruzi, A. (2006). The Web impact factor: a critical review. *The Electronic Library*, 24,

490-500.

Radicchi, F., Fortunato, S. & Castellano, C. (2008). Universality of citation distributions: toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences of the USA*, 105, 17268-17272.

Rousseau, R. (2007). The influence of missing publications on the Hirsch-index. *Journal of Informetrics*, 1, 2-7.

Rousseau, R., Guns, R. & Liu YX (2008). The *h*-index of a conglomerate. *Cybermetrics*, 12(1), Paper 2. http://www.cindoc.csic.es/cybermetrics/papers/v12i1p2.html

Ruane, F., & Tol, R. S. J. (2008). Rational (successive) *h*-indices: An application to economics in the Republic of Ireland. *Scientometrics*, 75(2), 395-405.

Thomson/Reuters Corporation. New hot papers. http://www.esi-topics.com/nhp/

Vanclay, J. K. (2007). On the robustness of the *h*-index. *Journal of the American Society for Information Science and Technology*, 58, 1547-1550.

Wan, JK., Hua, PH., Du, J. & Song, HM. (2007). Focus on the front of scientific evaluation, putting new indicators into practice: Three new bibliometric indicators adopted by China Academic Journals Comprehensive Citation Report. *Digital Library Forum* (in Chinese), issue 3, 36–41.

Wan, JK., Hua, PH., Rousseau, R. & Sun, XK. (2009). The journal download immediacy index (DII): Experiences using the CNKI full-text database. To appear.

Wan, JK., Hua, PH. & Sun, XK. (2005). Bibliometric analysis on cited frequency and download frequency of journal papers. *New Technology of Library and Information Service*, (4):58-62. (in Chinese).

Wan, JK. & Zhang, ZH. (2004). *Chinese Academic Journals Webometrics Test Report*. Beijing: CAJ-CD Publishing House of Tsinghua University (in Chinese).

Published as:

Ping-huan Hua, Ronald Rousseau, Xiu-kun Sun, Jin-kun Wan (2009). A download $h^{(2)}$-index as a meaningful usage indicator of academic journals. In: ISSI 2009; 12th International Conference on Scientometrics and Informetrics, (Birger Larsen & Jacqueline Leta, eds.), Rio de Janeiro: BIREME & Federal University of Rio de Janeiro, p. 587-596.