Journal of Informetrics 6 (2012) 285–288

Contents lists available at SciVerse ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Short communication

HistCite analysis of papers constituting the *h* index research front

Lutz Bornmann[a,*], Werner Marx[b]

[a] Max Planck Society, Administrative Headquarters, Hofgartenstr. 8, D-80539 Munich, Germany
[b] Max Planck Institute for Solid State Research, Heisenbergstr. 1, D-70569 Stuttgart, Germany

ARTICLE INFO

Article history:
Received 24 August 2011
Received in revised form 27 October 2011
Accepted 7 November 2011

Keywords:
h index
Research front
HistCite

ABSTRACT

In the Essential Science Indicators (Thomson Reuters), a research front exists to the *h* index (entitled "GOOGLE SCHOLAR H-INDEX; SCIENCE CITATION INDEX; GENERALIZED HIRSCH H-INDEX; H INDEX; GOOGLE SCHOLAR CITATIONS") consisting of a group of highly cited papers. We used HistCite to analyze the structure and relationships of the 45 papers forming the *h* index research front. Since we were interested in the topics of research on the *h* index at the front, we classified each paper according to its main topic. Six topics (inductively generated) were sufficient to classify the 45 papers: (1) citation database, (2) empirical validation study, (3) new application, (4) theoretical analysis, (5) new index development, and (6) literature review.

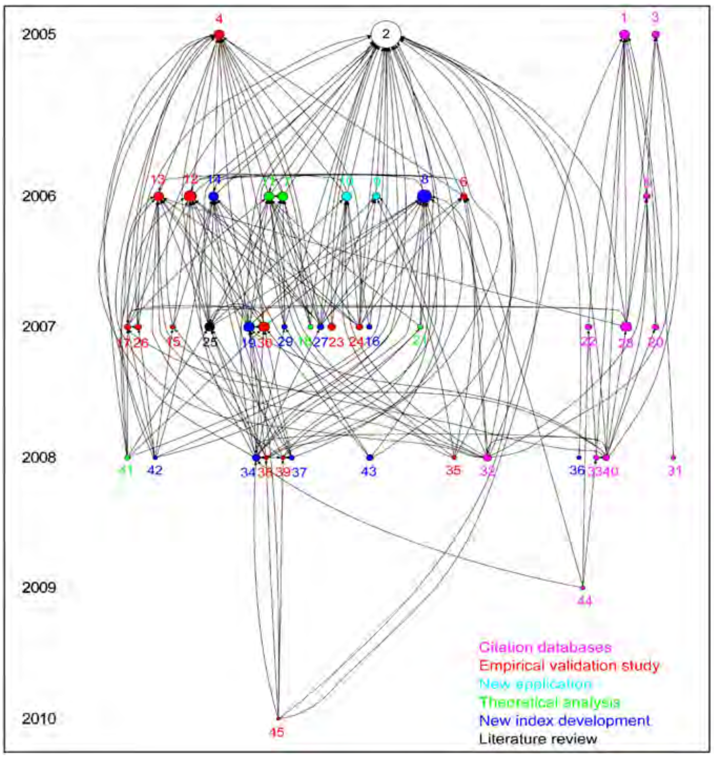
In 2005, Hirsch (2005) introduced the *h* index for quantifying the research output of scientists (Bornmann & Daniel, 2007, 2009). The index was proposed as an alternative to standard bibliometric indicators (such as total citation counts) and is defined as follows: "A scientist has index h if h of his 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 *h* index is based on a scientist's lifetime citedness, which incorporates productivity as well as citation impact. The *h* index is approximately proportional to the square root of the total citation counts (Franceschini & Maisano, 2010). Up to the end of 2010, the paper of Hirsch (2005) has been cited about 660 times, which shows that the index is fascinating to many people. The considerable impact of the *h* index on both bibliometricians and the wider scientific community can be traced to its simplicity and intuitive meaning (Franceschini & Maisano, 2010).

In the Essential Science Indicators (Thomson Reuters), a research front exists to the *h* index (entitled "GOOGLE SCHOLAR H-INDEX; SCIENCE CITATION INDEX; GENERALIZED HIRSCH H-INDEX; H INDEX; GOOGLE SCHOLAR CITATIONS") consisting of a group of highly cited papers. "A research front is a group of highly cited papers, referred to as core papers, in a specialized topic defined by a cluster analysis. A measure of association between highly cited papers is used to form the clusters. That measure is the number of times pairs of papers have been co-cited, that is, the number of later papers that have cited both of them. Clusters are formed by selecting all papers that can be linked together by a specified co-citation threshold. Research fronts can reveal emerging areas of science through citation patterns" (Thomson Reuters, 2008). The 45 papers shown in Fig. 1 form the core papers of the *h* index research which were published between 2005 and 2010 with more than 3000 citations in total (date of search in Web of Science: May 2011). For each paper the percentage of the Global Citation Score (GSC%) was calculated: It is the percentage difference between the average annual citation counts of a paper in the core and the mean of the average annual citation counts over all 45 papers ($m = 15$). For example, Jacso (2005) (paper no. 1 in Fig. 1) was cited 15 percentage points below the average over all 45 papers.

We used HistCite (Garfield, 2004, 2009; http://garfield.library.upenn.edu/algorithmichistoriographyhistcite.html) to analyze the structure and relationships of the 45 papers forming the *h* index research front. This paper is similar to the paper

* Corresponding author.
E-mail address: bornmann@gv.mpg.de (L. Bornmann).

1751-1577/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.joi.2011.11.001



No.	Research front paper	GCS%
1.	Jacso P, 2005, CURR SCI, V89, P1537	85
2.	Hirsch JE, 2005, PROC NAT ACAD SCI USA, V102, P16569	804
3.	Noruzi A, 2005, LIBRI, V55, P170	51
4.	Bornmann L, 2005, SCIENTOMETRICS, V65, P391	101
5.	Jacso P, 2006, ONLINE INF REV, V30, P297	48
6.	Saad G, 2006, SCIENTOMETRICS, V69, P117	57
7.	Egghe L, 2006, SCIENTOMETRICS, V69, P121	112
8.	Egghe L, 2006, SCIENTOMETRICS, V69, P131	228
9.	Banks MG, 2006, SCIENTOMETRICS, V69, P161	65
10.	Braun T, 2006, SCIENTOMETRICS, V69, P169	115
11.	Glänzel W, 2006, SCIENTOMETRICS, V67, P315	109
12.	Van Raan AFJ, 2006, SCIENTOMETRICS, V67, P491	180
13.	Cronin B, 2006, J AM SOC INF SCI TECHNOL, V57, P1275	123
14.	Batista PD, 2006, SCIENTOMETRICS, V68, P179	117
15.	Vinkler P, 2007, J INFORM SCI, V33, P481	50
16.	Schubert A, 2007, SCIENTOMETRICS, V70, P201	45
17.	Oppenheim C, 2007, J AM SOC INF SCI TECHNOL, V58, P297	73
18.	Egghe L, 2007, J AM SOC INF SCI TECHNOL, V58, P452	52
19.	Jin BH, 2007, CHIN SCI BULL, V52, P855	164
20.	Norris M, 2007, J INFORMETR, V1, P161	57
21.	Burrell QL, 2007, J INFORMETR, V1, P170	50
22.	Kousha K, 2007, J AM SOC INF SCI TECHNOL, V58, P1055	71
23.	Schubert A, 2007, J INFORMETR, V1, P179	86
24.	Costas R, 2007, J INFORMETR, V1, P193	65
25.	Bornmann L, 2007, J AM SOC INF SCI TECHNOL, V58, P1381	136
26.	Vanclay JK, 2007, J AM SOC INF SCI TECHNOL, V58, P1547	62
27.	Sidiropoulos A, 2007, SCIENTOMETRICS, V72, P253	71
28.	Meho LI, 2007, J AM SOC INF SCI TECHNOL, V58, P2105	170
29.	Iglesias JE, 2007, SCIENTOMETRICS, V73, P303	42
30.	Hirsch JE, 2007, PROC NAT ACAD SCI USA, V104, P19193	173
31.	Neuhaus C, 2008, J DOC, V64, P193	45
32.	Bar-Ilan J, 2008, SCIENTOMETRICS, V74, P257	121
33.	Kousha K, 2008, SCIENTOMETRICS, V74, P273	61
34.	Bornmann L, 2008, J AM SOC INF SCI TECHNOL, V59, P830	128
35.	Molinari JF, 2008, SCIENTOMETRICS, V75, P163	43
36.	Ruane F, 2008, SCIENTOMETRICS, V75, P395	41
37.	Schreiber M, 2008, J INFORMETR, V2, P211	52
38.	Schreiber M, 2008, J AM SOC INF SCI TECHNOL, V59, P1513	43
39.	Schreiber M, 2008, SCIENTOMETRICS, V76, P187	45
40.	Meho LI, 2008, J AM SOC INF SCI TECHNOL, V59, P1711	73
41.	Woeginger GJ, 2008, MATH SOC SCI, V56, P224	54
42.	Anderson TR, 2008, SCIENTOMETRICS, V76, P577	43
43.	Radicchi F, 2008, PROC NAT ACAD SCI USA, V105, P17268	93
44.	Harzing AW, 2009, J AM SOC INF SCI TECHNOL, V60, P41	62
45.	Schreiber M, 2010, J AM SOC INF SCI TECHNOL, V61, P169	32

Fig. 1. Historiograph of the 45 papers forming the *h* index research front. The nodes represent the 45 papers, the circle diameters are proportional to the GCS and the arrows indicate the citation direction. GSC% is the percentage difference between the average annual citation counts of a paper in the core and the mean of the average annual citation counts over all 45 papers ($m = 15$).

of Zhang, Thijs, and Glänzel (2011) who “trace the diffusion of *h*-related literature over a five-year period beginning with the introduction of the *h* index.” Whereas Zhang et al. (2011) included all relevant literature devoted to the *h* index ($n = 755$ papers) we focused on the *h* index research front papers. The historiography in Fig. 1 displays the linkages between these papers. The pathway arrows indicate which later papers cited which earlier papers. The circle diameters are proportional to the number of citations (GCS) until May 2011. Each of the papers was cited at least 5 times. The graph illustrates how the nodes are positioned on the pathways between the papers (the so-called centrality). The interconnectedness helps to identifying central publications. Visual overviews like those in Fig. 1 can help researchers and analysts interested in a certain research field to identify relations between key papers and the development of topics (Aris, Shneiderman, Qazvinian, & Radev, 2009).

As the figure shows most of the papers in the *h* index research front were published between 2006 and 2008. All these papers have a high degree of interconnectedness. The most cited papers measured by GCS% are the introductions of the *h* index itself (Hirsch, 2005) (paper no. 2) and the *g* index by Egghe (2006) (paper no. 8). Since we were interested in the topics of research on the *h* index at the front, we classified each paper in Fig. 1 according to its main topic. Six topics (inductively generated) were sufficient to classify the 45 papers: (1) citation database, (2) empirical validation study, (3) new application, (4) theoretical analysis, (5) new index development, and (6) literature review. Eleven studies deal with the citation databases (Web of Science, Scopus, Google Scholar) which can be used for the calculation of the *h* index. The main research question of these studies is whether the databases provide different index values for the same unit (e.g., one and the same scientist). The comparison of the feeless Google Scholar with the fee-based Web of Science and Scopus databases (for the *h* index calculation) is of particular interest here.

The disadvantages discussed in connection with the *h* index (Bornmann & Daniel, 2007, 2009) have led to the development of a now large number of *h* index variants. These are new indices based on the *h* index (Alonso, Cabrerizo, Herrera-Viedma, & Herrera, 2009; Bornmann, Mutz, Hug, & Daniel, 2011) that are supposed not to have one or more disadvantages. Eleven studies in Fig. 1 have a focus on the development of new variants. For example, the *g* index proposed by Egghe (2006) (paper no. 8) places more weight on the citation performance of a set of papers than the *h* index does. Further variants related to the *h* index research fronts are the *r* index and *ar* index (Jin, Liang, Rousseau, & Egghe, 2007) (paper no. 19) and successive *h* indices (Schubert, 2007) (paper no. 16). The indices at aggregated levels – group, research facility or country – can be calculated analogously to those of individual researchers or as successive *h* indices at higher aggregate levels (Kosmulski, 2006; Prathap, 2006). Only a small part of the ever introduced *h* index variants are included in Fig. 1. For a recently published meta-analysis on the *h* index, Bornmann et al. (2011) identified 37 *h* index variants in total. Their study shows a high correlation between the *h* index and its variants with an overall mean value between .8 and .9. This means that there is redundancy between most of the *h* index variants (also those introduced in the core papers in Fig. 1) and the *h* index.

The *h* index is no longer being used only as a measure to quantify the research output of scientists, but also to qualify journals (Braun, Glänzel, & Schubert, 2006) (paper no. 10) as an alternative to the Impact Factor for journals (Garfield, 1972) and to evaluate ensembles of papers related to keyword or compound specific research topics (Banks, 2006) (paper no. 9). The latter can help to differentiate between hot topics and older topics in a discipline. Both (new) applications of the *h* index are part of the *h* index research front in Fig. 1 and seem to be the most important new areas of application as opposed to others, e.g., the library's *h* index proposed by Liu and Rousseau (2009).

In recent years a lot of studies have analyzed different aspects of the *h* index and its variants which are summarized by Alonso et al. (2009), Bornmann and Daniel (2007, 2009), Egghe (2010), Norris and Oppenheim (2010), Panaretos and Malesios (2009), Thompson, Callen, and Nahata (2009), and Zhang et al. (2011). The first published literature overview of Bornmann and Daniel (2007) (paper no. 34) is part of the citation network in Fig. 1.

Two further groups of papers are included in the research front: the one group (named theoretical analysis in Fig. 1) contains mathematical–statistical descriptions of the *h* index (or its variants) (Glänzel, 2006) (paper no. 11) or informetric models (Egghe & Rousseau, 2006) (paper no. 7) for the *h* index (or its variants). The other group of papers present the results of empirical studies which investigated the validity of the *h* index and its variants (like the *g* index). Most of the studies analyzed the convergent validity while they compared the *h* index (and its variants) with standard bibliometric indicators (e.g., citations per paper, total citation counts) (Costas & Bordons, 2007) (paper no. 24) and with the assessment by peers (Bornmann & Daniel, 2005; van Raan, 2006) (paper no. 4, 12). van Raan (2006) found, e.g., that the "*h* index . . . relate[s] in a quite comparable way with peer judgments" (p. 491) for 147 Dutch research groups in chemistry. Other papers under this topic deal with the robustness of the *h* index with regard to self-citations (Schreiber, 2008) (paper no. 39) and to perturbations in the tail of the citation distributions (Vanclay, 2007) (paper no. 26). Hirsch (2007) (paper no. 30) investigated whether the *h* index is better than other standard bibliometrics indicators in predicting future scientific achievement.

Since the *h* index is a fast-growing field of research, it would be interesting to see whether the *h* index research front will develop within the framework of the six topics in Fig. 1 or whether other topics will become important. If the research follows directions proposed, e.g., by Mingers (2009) the framework will not change substantially. According to Mingers (2009) "more research is needed to study the *h*-index for larger and more diverse groups of researchers, especially those earlier in their career; carry out more comparisons both across and within social science disciplines; undertake comparisons with the more sophisticated bibliographic measures to validate the reliability of the *h*-index; develop rigorous stochastic models to understand its theoretical properties" (p. 1151). We would like to add the application of the *h* index to topic-specific ensembles of papers (Banks, 2006). Since the selection (retrieval) of the corresponding papers based on search terms like keywords or materials (specific chemical substances or extensive classes of compounds) is a real challenge, more research is needed here.

Further interesting questions for future research on the *h* index research front would be: When did the front appear for the first time? What is the typical life time of a paper in the front? What is the typical life time of a topic in the front? What are possible reasons for the fact that certain topics emerged or disappeared?

References

Alonso, S., Cabrerizo, F. J., Herrera-Viedma, E., & Herrera, F. (2009). *h*-Index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of Informetrics*, *3*(4), 273–289. doi:10.1016/j.joi.2009.04.001

Aris, A., Shneiderman, B., Qazvinian, V., & Radev, D. (2009). Visual overviews for discovering key papers and influences across research fronts. *Journal of the American Society for Information Science and Technology*, *60*(11), 2219–2228. doi:10.1002/asi.21160

Banks, M. G. (2006). An extension of the Hirsch index: Indexing scientific topics and compounds. *Scientometrics*, *69*(1), 161–168. doi:10.1007/s11192-006-0146-5

Bornmann, L., & Daniel, H.-D. (2005). Does the *h*-index for ranking of scientists really work? *Scientometrics*, *65*(3), 391–392. doi:10.1007/s11192-005-0281-4

Bornmann, L., & Daniel, H.-D. (2007). What do we know about the *h* index? *Journal of the American Society for Information Science and Technology*, *58*(9), 1381–1385. doi:10.1002/asi.20609

Bornmann, L., & Daniel, H.-D. (2009). The state of *h* index research. Is the *h* index the ideal way to measure research performance? *EMBO Reports*, *10*(1), 2–6. doi:10.1038/embor.2008.233

Bornmann, L., Mutz, R., Hug, S. E., & Daniel, H. D. (2011). A meta-analysis of studies reporting correlations between the *h* index and 37 different *h* index variants. *Journal of Informetrics*, *5*(3), 346–359. doi:10.1016/j.joi.2011.01.006

Braun, T., Glänzel, W., & Schubert, A. (2006). A Hirsch-type index for journals. *Scientometrics*, *69*(1), 169–173. doi:10.1007/s11192-006-0147-4

Costas, R., & Bordons, M. (2007). The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. *Journal of Informetrics*, *1*(3), 193–203.

Egghe, L. (2006). Theory and practise of the *g*-index. *Scientometrics*, *69*(1), 131–152. doi:10.1007/s11192-006-0144-7

Egghe, L. (2010). The Hirsch index and related impact measure. *Annual Review of Information Science and Technology*, *44*, 65–114.

Egghe, L., & Rousseau, R. (2006). An informetric model for the Hirsch-index. *Scientometrics*, *69*(1), 121–129.

Franceschini, F., & Maisano, D. A. (2010). Analysis of the Hirsch index's operational properties. *European Journal of Operational Research*, *203*(2), 494–504. doi:10.1016/j.ejor.2009.08.001

Garfield, E. (1972). Citation analysis as a tool in journal evaluation: Journals can be ranked by frequency and impact of citations for science policy studies. *Science*, *178*(4060), 471–479.

Garfield, E. (2004). Historiographic mapping of knowledge domains literature. *Journal of Information Science*, *30*(2), 119–145.

Garfield, E. (2009). From the science of science to scientometrics visualizing the history of science with HistCite software. *Journal of Informetrics*, *3*(3), 173–179.

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. doi:10.1007/s11192-006-0102-4

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. doi:10.1073/pnas.0507655102

Hirsch, J. E. (2007). Does the *h* index have predictive power? *Proceedings of the National Academy of Sciences of the United States of America*, *104*(49), 19193–19198. doi:10.1073/pnas.0707962104

Jacso, P. (2005). As we may search – Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. *Current Science*, *89*(9), 1537–1547.

Jin, B., Liang, L., Rousseau, R., & Egghe, L. (2007). The R- and AR-indices: Complementing the h-index. *Chinese Science Bulletin*, *52*(6), 855–863.

Kosmulski, M. (2006). I – A bibliometric index. *Forum Akademickie*, (11), 31.

Liu, Y., & Rousseau, R. (2009). Properties of Hirsch-type indices: The case of library classification categories. *Scientometrics*, *79*(2), 235–248. doi:10.1007/s11192-009-0415-1

Mingers, J. (2009). Measuring the research contribution of management academics using the Hirsch-index. *Journal of the Operational Research Society*, *60*(9), 1143–1153. doi:10.1057/jors.2008.94

Norris, M., & Oppenheim, C. (2010). The *h*-index: A broad review of a new bibliometric indicator [Review]. *Journal of Documentation*, *66*(5), 681–705. doi:10.1108/00220411011066790

Panaretos, J., & Malesios, C. (2009). Assessing scientific research performance and impact with single indices. *Scientometrics*, *81*(3), 635–670. doi:10.1007/s11192-008-2174-9

Prathap, G. (2006). Hirsch-type indices for ranking institutions' scientific research output. *Current Science*, *91*(11), 1439.

Schreiber, M. (2008). The influence of self-citation corrections on Egghe's *g* index. *Scientometrics*, *76*(1), 187–200.

Schubert, A. (2007). Successive *h*-indices. *Scientometrics*, *70*(1), 201–205. doi:10.1007/s11192-007-0112-x

Thompson, D. F., Callen, E. C., & Nahata, M. C. (2009). New indices in scholarship assessment [Review]. *American Journal of Pharmaceutical Education*, *73*(6).

Thomson Reuters. (2008). *Using bibliometrics: A guide to evaluating research performance with citation data (chapter: spezialisation)*. Philadelphia, PA, USA: Thomson Reuters.

Vanclay, J. K. (2007). On the robustness of the *h*-index. *Journal of the American Society for Information Science and Technology*, *58*(10), 1547–1550.

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. doi:10.1556/Scient.67.2006.3.10

Zhang, L., Thijs, B., & Glänzel, W. (2011). The diffusion of H-related literature. *Journal of Informetrics*, *5*(4), 583–593. doi:10.1016/j.joi.2011.05.004