Journal of Informetrics 3 (2009) 341–347

Contents lists available at ScienceDirect

Journal of Informetrics

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

New seniority-independent Hirsch-type index

Marek Kosmulski

Department of Electrochemistry, Lublin University of Technology, Nadbystrzycka 38, PL-20618 Lublin, Poland

ARTICLE INFO

Article history:
Received 8 March 2009
Received in revised form 5 May 2009
Accepted 5 May 2009

Keywords:
Citation metrics
Hirsch-index
Hirsch-type indices
Research output

ABSTRACT

The following seniority-independent Hirsch-type index has been defined. A scientist has index hpd if hpd of his/her papers have at least hpd citations per decade each, and his/her other papers have less than hpd + 1 citations per decade each. In contrast with the original *h*-index, which steadily increases in time, hpd of a mature scientist is nearly constant over many years, and hpd of an inactive scientist slowly declines. Therefore hpd is suitable to compare the scientific output of scientists in different ages.

1. Introduction

The systems of assessment of the scientific output of scientists and of institutions are based on the assessment of individual publications. This obvious truth is seldom pronounced. The total number of citations is by far the most common measure of success of a scientific paper. Such an approach favors old papers over recent papers, because the number of citations of a paper cannot decrease, and usually it increases, although the quality of a scientific paper does not improve on aging. I argue that the average number of citations per year cpy defined as total number of citations divided by (1 + present year − publication year) is more suitable as a measure of success of a scientific paper than its total number of citations. It will be shown below that cpy remains nearly constant over time periods comparable with the length of scientific career of a scientist. This is in line with the intuitive feeling that the quality of a paper does not change in time. The properties of cpy will be studied in this present paper. The cpy of individual publications can be used to assess the scientific output of scientists, scientific institutions, etc. The assessment is analogous to the methods based on total number of citations of individual papers. A Hirsch-type cpy-based index will be studied in detail.

2. The assessment of individual publications

The present paper is focused on assessment of active scientists in context of their selection (recruitment), advancement or award of funds. The cpy is probably not suitable to compare contemporary scientists with Darwin or Newton.

Different aspects of citation histories of scientific publications have been studied (Exner & Kunz, 1995; Glanzel, 2007; Glanzel, Schlemmer, & Thijs, 2003; Glanzel & Schoepflin, 1994; Maricic, Spaventi, Pavicoic, & Pifat-Mrzljak, 1998; van Dalen & Henkens, 2004).

The analysis of citation history of a paper in this section was limited to 20 years, which is a typical time horizon to that considered in previous studies. It was arbitrarily assumed that the period of 20 years is relevant to practical purposes, that is, papers substantially older than 20 years are seldom taken into account in competition between scientists for positions,

E-mail address: mkosmuls@hektor.umcs.lublin.pl.

1751-1577/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.joi.2009.05.003

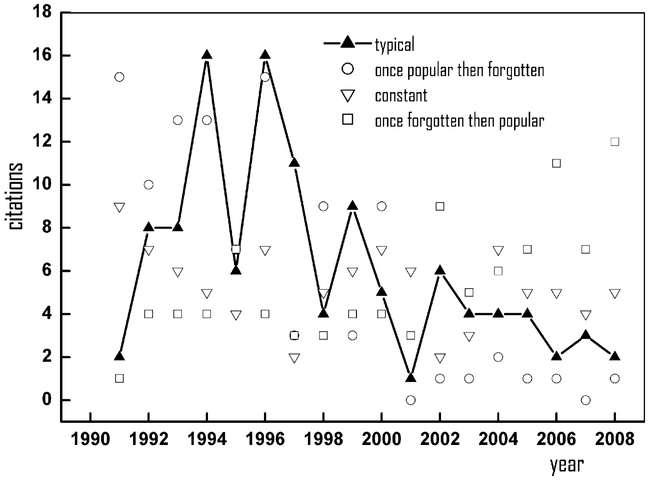

Fig. 1. Citation histories of 4 papers.

grants, etc. The representative sample of original journal papers (no review papers) consists of 24 papers from Journal of Applied Physics representing physics, 29 papers from Journal of Molecular Biology representing life sciences, and 34 papers from Journal of Organic Chemistry representing chemistry. The journals were selected arbitrarily as representatives of 3 main branches of science, and all papers from these journals were considered, which were published in 1990 and had received 90–120 citations till February 2009, according to the Web of Science® database, Thomson Reuters©. Such a citation record is equivalent to an average of 4.5–6 citations per year. They belong to the top-20 % but not to the top-2% most cited papers published by each journal in 1990. Thus, the publications examined below have been rather successful, but not extremely successful.

The papers published in 1990 (especially in the second half of 1990) had limited chance of being cited in 1990. Thus, the number of citations obtained by a paper in the year of its publication is unlikely to be representative for that paper, and it strongly depends on the exact publication date. To avoid distortion in the statistical analysis by one atypical year, only the sum of citations obtained by a paper in 1990 and in 1991 was taken into account (and assigned to 1991).

The citation histories of 4 papers (out of 87) are explicitly shown in Fig. 1. The citations obtained in year 2009 (till February, 0 or 1 citation for most publications) are ignored in Fig. 1 and in further analysis. Although the studied publications represent different branches of science, many common properties of their citation histories (typical history is represented by a polygonal chain in Fig. 1) are observed. For most papers their citation rate increased over the first 5 years after publication, and then slowly decreased to reach in 2008 about 1/2 (J. Org. Chem.), 1/3 (J. Appl. Phys.) or 1/4 (J. Mol. Biol.) of the citation rate from 1995. The citation history had an oscillatory character as indicated by the polygonal chain in Fig. 1. A typical standard deviation in the number of citations in particular years was about 0.7 in J. Appl. Phys. and J. Mol. Biol. and 0.6 in J. Org. Chem. of the average citation rate over the period 1991–2008, the highest standard deviation was 1.1, and the lowest standard deviation was 0.3 of the average value.

Fig. 1 shows also 3 examples of atypical citation histories. Citation histories similar to the typical graph shown in Fig. 1 occur for most publications, and atypical histories are rare. Extreme cases of atypical histories were selected for Fig. 1, and most other atypical histories are intermediate between the extreme cases and normal behavior. Several papers received many citations over the first few years after the publication, and then their citation rate dropped to almost zero. Very few papers showed almost constant citation rate over the period 1990–2008. Finally, very few papers (“sleeping beauties”; van Raan, 2004) were seldom cited in the first decade after publication, and more often cited in the second decade.

2.1. Definition of cpy

The oscillatory character of the citation rate of an individual paper makes it rather unattractive as a measure of the quality of a scientific paper, because a selection of particular year may randomly favor or disfavor certain papers. For example, each of 4 papers shown in Fig. 1 had its “good years” (more citations than any other paper) and “bad years”.

The oscillations are substantially damped when the average citation rate

$$\text{cpy} = \frac{(\text{citations over the period 1990 to year } x)}{(x-1989)}, \quad (1)$$

is plotted (Fig. 2a, the same papers as in Fig. 1). All 4 polygonal chains representing the citation histories are relatively smooth. Again an increase in cpy in typically observed over a few first years after publication, but the maximum occurs later (about 7 years after publication), and it is less sharp than the maximum in the citation history shown in Fig. 1. The cpy still shows a systematic change and oscillations, but they are less significant than those in the number of citations in each year.

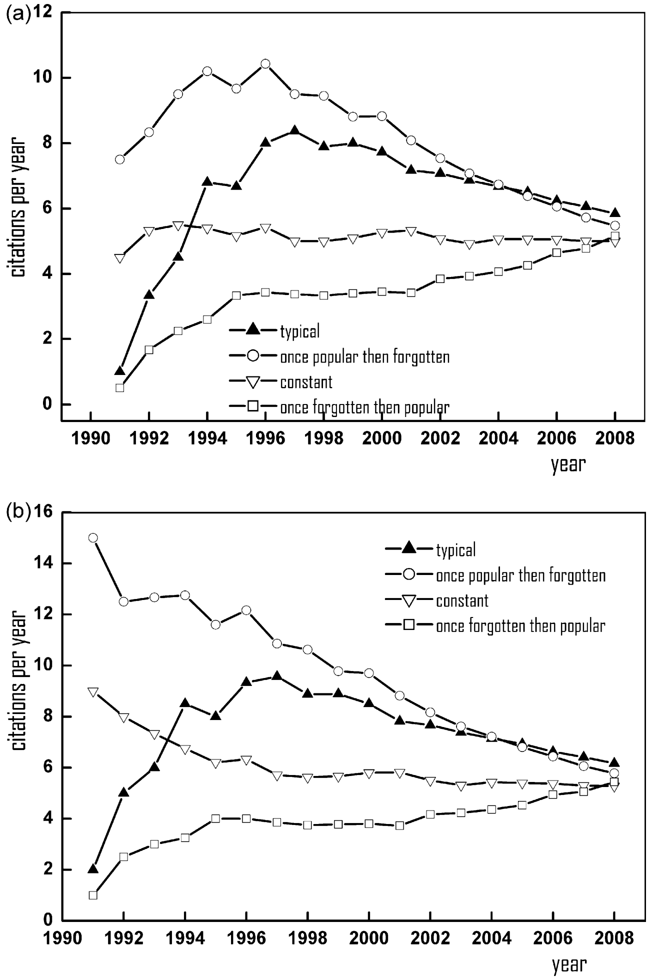

Fig. 2. The cpy of the same 4 papers as in Fig. 1. (a) The cpy* (Eq. (2)) of the same 4 papers as in Fig. 1. (b)

The definition of cpy in Eq. (1) implies a sensitivity of the course of the cpy in the first few years after publication to the exact publication date of the paper of interest (e.g., January 1990 vs. December 1990). This problem can be reduced when a modified definition is used

$$\mathrm{cpy}^* = \frac{(\text{citations over the period 1990 to year } x)}{(x - 1990)}, x > 1990 \tag{2}$$

(two first years “merged”) or

$$\mathrm{cpy}^{**} = \frac{(\text{citations over the period 1991 to year } x)}{(x - 1990)}, x > 1990 \tag{3}$$

(the citations obtained in the year of publication ignored). The cpy* is plotted in Fig. 2b for the same papers as cpy in Fig. 2a. The course of cpy** is similar, because 3 of 4 papers of interest received zero citations in 1990. The difference between cpy and cpy* or cpy** is substantial in the first few years after publication, and it is rather immaterial for papers which are more than 10 years old.

2.2. Properties of cpy

The cpy in 2008 reached on average about 0.7 (J. Org. Chem. and J. Appl. Phys.) or 0.6 (J. Mol. Biol.) of the highest cpy in the paper’s history, with the ratios ranging from 0.3 to 1.

A typical standard deviation in cpy in particular years was about 0.2 of the cpy in 2008 for J. Org. Chem. and 0.3 for J. Mol. Biol. and J. Appl. Phys., the highest standard deviation was 0.6, and the lowest standard deviation was 0.1 of the cpy in 2008. The standard deviation in cpy is chiefly due to sudden changes in cpy during the first few years after publication, and in papers older than 10 years, the cpy is quite stable.

The system of evaluation based on cpy slightly favors papers, which are about 5 years old. This has limited impact on evaluation of scientists, when all competitors have similar distributions of the publication ages. Lower credit given to older papers can even be considered as an advantage of the evaluation system of individuals based on cpy, because it favors those who are currently active over those who enjoy high citation numbers of their old papers—just opposite to the system based on total numbers of citations.

The current cpy of papers is calculated and displayed in popular software (Web of Science®, Publish or Perish (A computer program, which retrieves and analyzes academic citations from the Web. Available from www.harzig.com/pop.htm.)). Otherwise it can be easily calculated using standard spreadsheet software.

3. A cpy-based assessment of individuals

By analogy with the indices based on total numbers of citations (Hirsch, 2005), the following parameters for assessment of individuals can be proposed.

1. Sum of cpy of all individual papers.
2. Average cpy per paper.
3. Number of papers having cpy higher than certain number, say 5.
4. Sum of cpy of several, say 5, top-cpy papers.
5. Hirsch-type approach, which will be discussed later.

The obvious advantage of cpy-based indices with respect to the indices based on total numbers of citations is that the former do no favor old scientists over young scientists.

The problem of *h*-index favoring old over young scientists has been recognized in the original paper by Hirsch (2005), who introduced the *m* quotient defined as *h* divided by the number of years passed since the first paper was published. The *m* quotient unduly favors young scientists. For example, there are many scientists who have $h=1$ in the first year of their careers or $h=2$ in the second year of their careers, and rather few, who have $h=30$ in the 30th year of their careers. Several other attempts to correct the original *h*-index for the age of paper/scientist have been published (Egghe & Rousseau, 2007; Jin, Liang, Rousseau, & Egghe, 2007; Sidiropoulos, Katsaros, & Manolopoulos, 2006). It should be emphasized that favoring old over young scientists is not the only, and probably even not the most important drawback of the original Hirsch index. Self-citations, multi-author papers, and comparison of scientists working in different fields are other problems, which have been widely discussed (Bornmann & Daniel, 2009; Egghe, 2008; Schreiber, 2007, 2008).

3.1. *Definition of hpy*

In this present study a new Hirsch-type cpy-based index is proposed. In order to introduce the idea and to study the properties of the new index, an authentic publication record of a scientist will be used. The entire analysis is based on the data taken from Web of Science®. The Eminent Scientist (ES) has published 211 papers over the period 1959-2006. The average age of his papers is 28 years. The papers received 2636 citations (12.5 per paper), and his *h*-index is 23.

The top-cpy papers of ES are summarized in Table 1.

Table 1
Top-cpy papers of ES in 2008.

Rank in cpy	cpy	Year published	Citations	Rank in citations
1	8.18	1969	327	1
2	4.84	1990	92	4
3	3.38	1977	108	3
4	3.29	2002	23	21
5	2.96	1962	139	2
6	2.86	2002	20	33
7	2.55	1987	56	5
8	2.17	2003	13	57
9	2.14	2002	15	46
10	2.14	2002	15	47
11	2.11	2000	19	34
12	1.86	2001	13	58
13	1.83	2003	11	67
14	1.47	1977	47	7
15	1.44	2000	13	59
16	1.43	2002	10	70
17	1.33	1976	44	8
18	1.32	1968	54	6
19	1.27	1998	14	48
20	1.27	1994	19	35

Among the top-cpy papers, there are several relatively recent papers with low total numbers of citations (e.g., papers ranked 67 and 70 in total number of citations). On average the papers in Table 1 (top-cpy) are more recent than the most-cites papers, and papers ranked 9–20 in total number of citations are beyond the top-20 in cpy. The recent publication activity of ES had limited effect on his total number of citations (the most recent paper in the top-20 was published in 1990), but it had a substantial effect on his cpy (every 2nd paper in top 10 was published in XXI century).

By analogy to the original *h*-index, the following seniority-independent Hirsch-type index can be defined. A scientist has index hpy if hpy of his/her papers have at least hpy citations per year each, and his/her other papers have less than hpy + 1 citations per year each. The data from Table 1 produce hpy equal to 3.

3.2. Properties of hpy

This is typical for scientists with *h* of about 20, that only a few of their papers have substantially more than 5 citations per year. Therefore, such scientists usually have hpy of 1–5 with a few exceptions. The most successful scientists ($h > 50$) have hpy in excess of 10. The disadvantage of the hpy index is that many scientists have identical hpy of 1–5. This problem can be avoided by introduction of a scaling factor. In typical citation records, the number of citations of the most cited paper is in the same range the number of papers published. For example, among 16 citation records analyzed in (Schreiber, 2007), in 13 cases the ratio of the number of citations of the most cited paper to the number of papers published was in the range from 0.6 to 2.6, and the extreme values were 0.23, and 4.4 (a factor of 4, but still the same order of magnitude). For about 30% of papers in a typical citation record, the number of citations of a paper and the rank of that paper in the list of the most cited papers differ by a factor <10. In contrast, the cpy of the most cited paper is lower than the number of papers published by an order of magnitude. The role of the scaling factor is that the ranked numbers assume values on the same order of magnitude as their ranks.

3.3. Definition of hpd

Using a scaling factor of 10, a new seniority-independent Hirsch-type index can be defined. The scaling factor of 10 may be criticized as an arbitrary number, which randomly favors or disfavors individuals. On the other hand, a year is also an arbitrary period of time, that is, a number of citations per decade cpd is an equally arbitrary measure of a quality of a paper as a number of citations per year. A scientist has index hpd if hpd of his/her papers have at least hpd citations per decade each, and his/her other papers have less than hpd + 1 citations per decade each. The data from Table 1 produce hpd equal to 14, which is on the same order of magnitude as *h* of the same scientist.

3.4. Properties of hpd

The citation history of ES's papers has been analyzed in order to examine the properties of the hpd index. Table 2 presents the citation record of the top (at that time)-cpy ES's papers in 1989.

Table 2 shows a similar trend as Table 1, that relatively recent papers with low total number of citations may have a high cpd. Obviously in 1989 ES had fewer citations and lower *h*-index than in 2009, but Table 2 indicates that his hpd in 1989 was 20, that is, it was substantially higher than in 2009. This result in intuitively correct, namely ES has published his most-cited

Table 2
Top-cpy papers of ES in 1989.

Rank in cpy in 1989	Cpy in 1989	Year published	Citations	Rank in citations in 1989	Rank in cpy in 2008
1	7.24	1969	152	1	1
2	4.77	1977	62	3	3
3	3	1982	24	15	22
4	2.89	1981	26	12	32
5	2.77	1977	36	6	14
6	2.6	1975	39	5	24
7	2.57	1976	36	7	17
8	2.5	1978	30	9	26
9	2.5	1986	10	38	27
10	2.5	1986	10	40	38
11	2.5	1986	10	42	45
12	2.33	1978	28	10	7
13	2.33	1987	7	55	21
14	2.29	1962	64	2	5
15	2.25	1978	27	11	33
16	2.2	1985	11	34	36
17	2.18	1979	24	16	30
18	2	1968	44	4	18
19	2	1977	26	13	41
20	2	1987	6	60	54

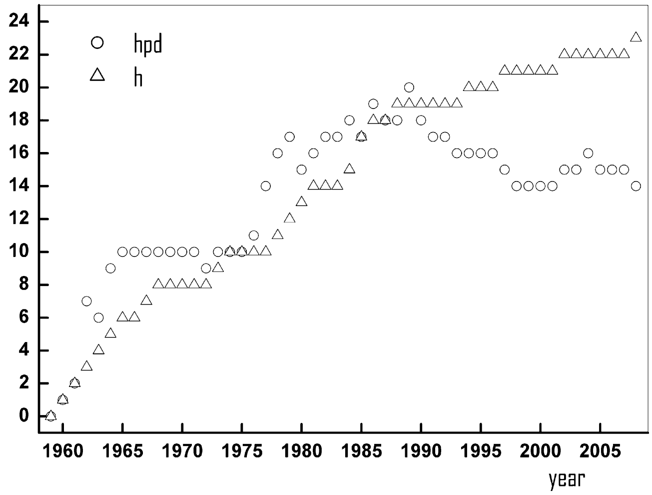

Fig. 3. The history of hpd- and *h*-index of ES.

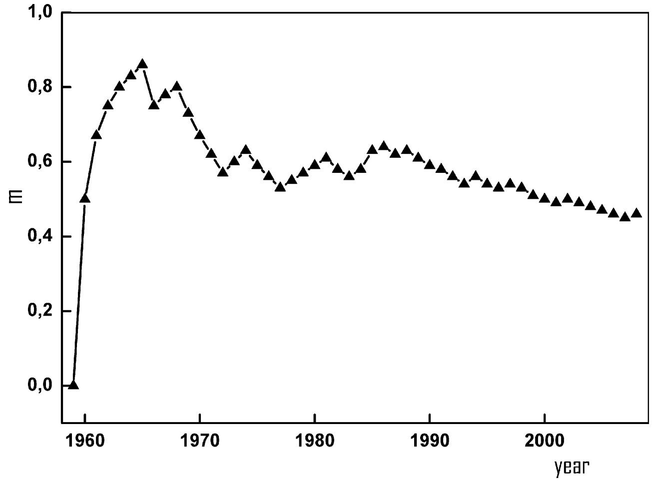

Fig. 4. The history of *m*-index of ES.

papers before 1990, and the contribution of papers published after 1990 to his high citation record is rather insignificant. Thus, a higher credit received by ES in 1989 than in 2008 is entirely fair. The last column in Table 2 indicates, that only 7 out of top-20 papers in 1989 held their positions in top-20 till 2008, and a few papers dropped severely in their ranks (e.g., from 3rd to 22nd place). On the other hand, a few cases of advancement of relatively old papers (e.g., from 14th to 5th place) occurred.

The analysis of the citation record of ES during his entire career in terms of *h*-index and of hpd is presented in Fig. 3.

The *h*-index steadily increased, although since 1988 the increase has been rather slow. The hpd index peaked in 1989, but the changes in hpd over the period 1982–1992 have been rather insignificant (range 17–20), and oscillatory rather than systematic. Fig. 4 shows the *m* quotient (Hirsch, 2005), which peaked in 1965 (when ES was young), and steadily decreased since then. The course of m of ES confirms the above allegation that the m quotient unduly favors young scientists.

ES received relatively consistent hpd (range 14–20) over the period 1977–2008 (32 years). The citation history of only one scientist (ES) has been examined in detail. However, the stable value of cpy of individual papers demonstrated in Section 2 suggests, that also in other successful scientists, the hpd should rather be stable once they reach certain level of "scientific maturity".

4. Conclusions

1. The present hpd index is a useful parameter, which can be applied to compare the citation records of scientists in different ages. The hpd can be easily calculated from easily available data.

2. The hpd can be further modified, e.g., for multi-author papers the individual cpy of each paper can be divided by the number of co-authors to produce the contribution of single co-author.
3. The hpd can also be used in assessment of the scientific journals and institutions.

Acknowledgment

The idea of combining the citations in 1990 and in 1991 into one number (Section 2) was suggested by an anonymous referee.

References

Bornmann, L., & Daniel, H. D. (2009). The state of *h* index research. *EMBO Reports*, *10*(1), 2.
Egghe, L. (2008). Mathematical theory of the *h*- and *g*-index in case of fractional counting of authorship. *Journal of American Society Information Science Technology*, *59*(10), 1608–1616.
Egghe, L., & Rousseau, R. (2007). An *h*-index weighted by citation impact. *Information Processing and Management*, *44*(2), 770–780.
Exner, O., & Kunz, M. (1995). Citation histories of related papers in the field of chemical correlation analysis. *Scientometrics*, *32*(1), 3–10.
Glanzel, W. (2007). Characteristic scores and scales. A bibliometric analysis of subject characteristics based on long-term citation observation. *Journal of Informetrics*, *1*, 92–102.
Glanzel, 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.
Glanzel, W., & Schoepflin, U. (1994). A stochastic model for the ageing of scientific literature. *Scientometrics*, *30*(1), 49–64.
Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of National Academy of Sciences of United States of America*, *102*(46), 16569–16572.
Jin, B. H., Liang, L., Rousseau, R., & Egghe, L. (2007). The R- and AR-indices: Complementing the h-index. *Chinese Science Bulletin*, *52*(6), 855–863.
Maricic, S., Spaventi, J., Pavicoic, L., & Pifat-Mrzljak, G. (1998). Citation context versus the frequency counts of citation histories. *Journal of the American Society for Information Science*, *49*(6), 530–540.
Schreiber, M. (2007). Self-citation corrections for the Hirsch index. *EPL*, *70*, 30002.
Schreiber, M. (2008). A modification of the *h*-index: The h_m index accounts for multi-authored manuscripts. *Journal of Informetrics*, *2*(3), 211–216.
Sidiropoulos, A., Katsaros, D., & Manolopoulos, Y. (2006). Generalized *h*-index for disclosing latent facts in citation networks. *Scientometrics*, *72*(2), 253–280.
van Dalen, H. P., & Henkens, K. (2004). Demographers and their journals: Who remains uncited after ten years. *Population and Development Review*, *30*(3), 489–506.
van Raan, A. (2004). Sleeping beauties in science. *Scientometrics*, *59*(3), 467–472.