A new Hirsch-type index saves time and works equally well as the original h-index

Marek Kosmulski

Department of Electrochemistry, Lublin University of Technology
PL-20618 Lublin (Poland)
e-mail: mkosmuls@hektor.umcs.lublin.pl

Abstract

A scientist's h(2) index is defined as the highest natural number such that his h(2) most-cited papers received each at least $[h(2)]^2$ citations. The advantage of h(2) as the index to characterize the scientific output of an individual over the original h index (HIRSCH, 2005) is that less work is required to verify the authorship of the relevant papers.

(HIRSCH, 2005) introduced an index h defined as the number of papers of certain author with citation number of at least h, and calculated the values of the h-index of Nobel-winners and of other top physicists and biologists. Typical h-values in this group were about 50, and the highest h of an individual was 191. The h-index gained popularity after a publication in Nature (BALL, 2005). BORNMANN & DANIEL (2005) discussed the usefulness of the h-index in assessment of relatively young scientists (h-index of about 3), but in the opinion of the present author the h-index is more suitable for assessment of mature scientists who have published at least 50 papers and have h-indexes of at least 10.

The h-indexes of all full professors affiliated at a department of chemistry of one university in Poland have been calculated and analyzed in detail. Sufficient information was available to assess if the papers found in the database (Thomson Scientific) were authored by the individual of interest of by someone else who happens to have the same family name and first name initial(s). Even the maiden names of the professors who took their husbands' names after marriage were know, but (in this particular case) the pre-marriage publications did not improve their h-indexes. For 14 out of 19 professors, the automatic search of the database immediately produced the correct h-index. For the other 5 professors, the automatic search of the database produced an overestimated h-index. This was because the other scientists, who accidentally have the same family name and first name initial(s) have also published frequently cited papers. The verification of the authorship was time-consuming, and without sufficient knowledge of the scientific CVs of the individuals of interest, the results would be very uncertain. In other words, the calculation of the h-index would be very easy provided that each scientist had an unique combination of family name and initials, but obviously this is not the case. This problem was not mentioned by HIRSCH (2005), although he must have encountered it. For example the automatic search for M.L.Cohen produces a combined h-index of the physicist from California (third-highest h-index among physicists, according to HIRSCH) and of another M.L.Cohen representing medicine, who also has an outstanding citation record. Hopefully HIRSCH solved this problem properly, but for a non-expert user of the database it is even difficult to establish how many different M.L.Cohens contributed to the fantastic result produced by the automatic search of the database.

The verification of the authorship is the most difficult and time consuming step in the calculation of the h-index. The time required for the analysis can be reduced when a new index is used rather than h. A scientist's h(2) index is defined as the highest natural number such that his h(2) most-cited papers received each at least $[h(2)]^2$ citations. The h(2)-index can be established by looking at the list of papers of an individual ordered by number of citations in the Thomson Scientific database. Most natural scientists and engineers know the squares of natural numbers up to 16 by heart, and h(2)>16 occurs very seldom. For example a h(2) of 10 denotes that 10 papers were cited at least 100 times each. The total number of citations of a scientist is at least $[h(2)]^3$, and usually it is higher than $[h(2)]^3$ by a factor of about 5 (analogous problem for the original h-index was discussed in detail by HIRSCH). Thus, the h(2)-index is roughly proportional to the cube root of the total number of citations.

The h(2) of 25 was obtained by automatic search for citations of E.Witten (the highest h-index among physicists, according to HIRSCH), and the same search produced h=112. As expected $[h(2)]^3$ is on the same order of magnitude as h^2. Witten's h(2) is lower than his h by a factor of 4.5, and so is the time necessary to verify the authorship of papers contributing to the corresponding indexes. Yet, the original h-index still needs much less time for the verification of the authorship than the total number of citations, and in the opinion of the present author this is the most significant advantage of the h-index over the total number of citations.

Table 1 presents the ranking of 19 chemistry professors from the one university (*vide ultra*) in terms of the following factors

max – the number of citations of the most cited paper. This is not necessarily the best method to assess the scientific output of an individual, but in contrast with other methods considered by HIRSCH (total number of papers, total number of citations, etc.), it can be quickly established.

h(2) – index. The average h(2) of an individual was 5.11 (st. dev. 1.10).

ch(2) – h(2) index corrected for self citations. For each frequently cited paper, the number of self-citations was subtracted from the number of citations. Then the papers were ranked according to the corrected number of citations and h(2) was established as described above. It should be emphasized that citations corrected for self-citations are not necessarily independent citations, because citations by a coauthor were not corrected for. The average h(2)-index dropped by 18 % after correction for self-citations.

h – index. The average h of an individual was 15.42 (st. dev. 5.39).

ch - h index corrected for self-citations as described above for h(2). The average h-index dropped by 26 % after correction for self-citations.

The shared ranks are expressed by averages, e.g., shared 1st and 2nd place = rank 1.5.

Table 1 Ranking of 19 individuals in terms of different criteria

Individ.	max	h(2)	ch(2)	h	ch
A	1	1.5	2.5	8.5	5
B	2	5	2.5	3.5	3.5
C	3	5	2.5	2	2
D	4	1.5	2.5	1	1
E	5	5	5.5	6	7.5
F	6	5	5.5	5	3.5
G	7	10.5	10.5	12	10.5
H	8	10.5	10.5	10	10.5
I	9	5	10.5	3.5	7.5
J	10	16	10.5	14	14
K	11	10.5	10.5	8.5	7.5
L	12	10.5	10.5	12	14
M	13	10.5	16.5	12	16.5
N	14.5	16	10.5	15	12
O	14.5	16	16.5	17	16.5
P	16	10.5	10.5	7	7.5
Q	18	16	16.5	17	14
R	17	16	16.5	17	18
S	19	19	19	19	19

Table 2 presents the correlations between the rankings based on different criteria.

Table 2 Correlation coefficients

	max	h(2)	ch(2)	h	ch
max	1.0000	0.8738	0.9145	0.7871	0.8183
h(2)	0.8738	1.0000	0.8643	0.9138	0.8786
ch(2)	0.9145	0.8643	1.0000	0.8497	0.9266
h	0.7871	0.9138	0.8497	1.0000	0.9289
ch	0.8183	0.8786	0.9266	0.9289	1.0000

Tables 1 and 2 indicate that different criteria produce a similar order with a few exceptions. The time-consuming correction for self-citation has induced rather insignificant changes in the rankings. Interestingly, although the contribution of self-citations to h (26 %) is substantially higher than to h(2) (18 %), the h-ch correlation coefficient is higher than the h(2)-ch(2) correlation coefficient. This result is against expectations. In other words, replacement of h by h(2) did not result in reduction of the effect of self citations on the record of the individual. For instance two individuals (I and M) clearly improved their position by frequent auto-citation in terms of h- and h(2)-index as well.

The most substantial difference between the rankings based on the h- and h(2)-index is in relatively better correlation of the later with the max-based ranking. Clearly h favors a type of "hard worker" (many papers with moderate number of citations per paper, the individual P is an example) over a type of "genius" (few papers with a high number of citations per paper, the individual A is an example). This problem was discussed by EGGHE (2006), who proposed the g-index, which gives even more credit to a "genius" than the present h(2)-index.

The popularity of the h-index may be due to the fact that "hard-workers" and more numerous than "genii".

Although the h(2)-index failed in elimination of self-citations as the means to artificially improve own record, it succeeded in reducing the number of papers in the sample of interest from 293 (h-index) to 97 (h(2)-index), and the time necessary to verify their authorship.

Certainly the h(2)-index has the same intrinsic disadvantages as the other indexes based on the number of citations that is:

- the indexes of scientists working in different fields are not comparable
- the indexes of scientists of different age are not comparable
- own record can be easily improved by self-citations or mutual citations
- own record can be easily improved by publishing review papers

Nevertheless, the h(2)-index offers an attractive alternative to the h-index and to the total number of citations as the means to assess the scientific output of a chemist.

The idea coined in the present paper can be further generalized, by defining a h(x) index as the number of papers of certain author with citation number of at least $[h(x)]^x$. The original h-index corresponds to x=1, the h(2)-index introduced in the present paper corresponds to x=2, and the total number of papers corresponds to $x=-\infty$. The original h-index is probably appropriate in the fields, where the typical number of citations per article is relatively low, e.g., in mathematics or astronomy. The h(2)-index is favored in chemistry and physics. In medicine and biology, where the typical number of citations per article is higher than in chemistry, x=2.5 may be more appropriate.

Acknowledgement

An anonymous referee is acknowledged for helpful suggestions.

References

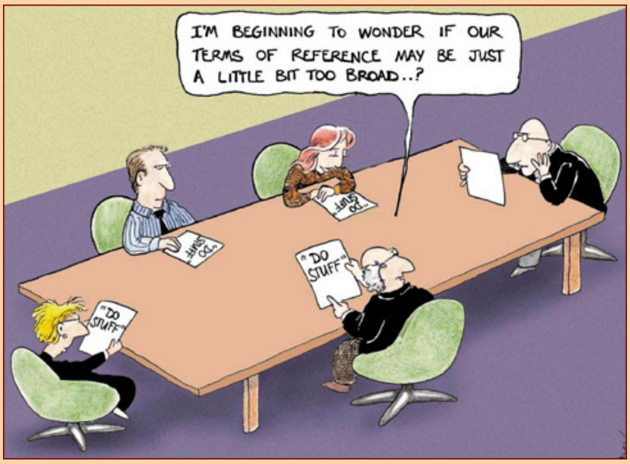
BALL, P. (2005), Index aims for fair ranking of scientists. *Nature,* 436 : 900.

BORNMANN, L. DANIEL H.D. (2005), Does the h-index for ranking of scientists really work? *Scientometrics*, 65: 391-392.

EGGHE, L. (2006), An improvement of the h-index: the g-index. *ISSI Newsletter,* 2:8-9.

HIRSCH, J.E. (2005), An index to quantify an individual's scientific research output. *P.Natl. Acad.Sci.USA,* 102: 16569-16572.

CARTOON



© Nick Kim (Nearing Zero). Reproduced with the permission of the author.