Jointly published by Akadémiai Kiadó, Budapest and Springer, Dordrecht

Scientometrics, Vol. 81, No. 3 (2009) 779–785
DOI: 10.1007/s11192-009-2290-1

A multidimensional extension to Hirsch's h-index

MIGUEL A. GARCÍA-PÉREZ

Departamento de Metodología, Facultad de Psicología, Universidad Complutense, Campus de Somosaguas, 28223 Madrid, Spain

The h-index is becoming a reference tool for career assessment and it is starting to be considered by some agencies and institutions in promotion, allocation, and funding decisions. In areas where h indices tend to be low, individuals with different research accomplishments may end up with the same h. This paper proposes a multidimensional extension of the h index in which the conventional h is only the first component. Additional components of the multidimensional index are obtained by computing the h-index for the subset of papers not considered in the immediately preceding component. Computation of the multidimensional index for 204 faculty members in Departments of Methodology of the Behavioral Sciences in Spain shows that individuals with the same h can indeed be distinguished by their values in the remaining components, and that the strength of the correlation of the second and third components of the multidimensional index with alternative bibliometric indicators is similar to that of the first component (i.e., the original h).

Introduction

The *h* index [HIRSCH, 2005] has gained recognition as a measure of research accomplishments, perhaps because it considers directly the impact of each of the papers authored by an individual. A scientist's *h* is the highest number of his/her papers that have each received at least that number of citations. Thus, a scientist with an *h* of 20 has 20 published papers each of which has received at least 20 citations. The *h* index was conceived and tested in disciplines where *h* tends to be high, and it has been shown to have predictive power in those areas [BORNMANN & DANIEL, 2005; HIRSCH, 2007]. As a result, the *h* index is starting to be considered by some agencies and institutions as an aid in decisions for promotion, allocation, or funding [RODRÍGUEZ NAVARRO & IMPERIAL RÓDENAS, 2007].

The *h* index has also been shown to have some drawbacks [BATISTA, CAMPITELI, KINOUCHI & MARTINEZ, 2006; KELLY & JENNIONS, 2006, 2007; LEHMANN, JACKSON & LAUTRUP, 2007; ROUSSEAU, 2008; SCHREIBER, 2007; VINKLER, 2007; WENDL, 2007], which has led several authors to propose variants or generalizations [ANDERSON, HANKIN & KILLWORTH, 2008; BATISTA & AL., 2006; BURRELL, 2007; EGGHE, 2006A, 2006B; IGLESIAS & PECHARROMÁN, 2007; IMPERIAL & RODRÍGUEZ-NAVARRO, 2007;

Received January 21, 2009; Published online April 17, 2009

Address for correspondence:
MIGUEL A. GARCÍA-PÉREZ
E-mail: miguel@psi.ucm.es

0138–9130/US $ 20.00

KOSMULSKI, 2006; LEHMANN & AL., 2007, 2008; LIANG, 2006; RUANE & TOL, 2008; SCHUBERT, 2009; SIDIROPOULOS, KATSAROS & MANOLOPOULOS, 2007; TABER, 2005]. In non-mainstream areas where *h* indices tend to be low and where self-citations may thus play an important role, comparison of the research accomplishments of different individuals through their *h* can be misleading in that individuals with qualitatively and quantitatively different careers may end up having the same *h* index [GARCÍA-PÉREZ, 2009]. In these cases, *h* appears insufficient as a criterion in promotion, allocation, or funding decisions.

This paper proposes a multidimensional extension to the *h* index and illustrates its capabilities to differentiate the research accomplishments of individuals in an area where *h* indices are low and many individuals have the same *h*.

The multidimensional *h* index

Let M be the total number of papers that an individual has published, let N be the number of those papers that have been cited at least once (and note that $N \leq M$), and let $\boldsymbol{C} = (c_1, c_2, ..., c_N)$ be an N-dimensional vector of citation counts whose components indicate the number of citations received by each of those N papers. The citation counts in $\boldsymbol{C}$ are assumed ordered so that $c_i \geq c_{i+1}$ for all $1 \leq i < N$. With this notation and conventions, the h index is the largest i (for $1 \leq i \leq N$) satisfying $c_i \geq i$, and we will define $h_1 = h$ to be the first component of the multidimensional h index. Naturally, the $N - h_1$ remaining papers published by this individual have also received citations and, thus, they are worth considering. The second component, h_2, of the multidimensional h index is then obtained by applying the same logic to those $N - h_1$ papers. Specifically, find the largest i (but now for $h_1 + 1 \leq i \leq N$) satisfying $c_i \geq i - h_1$ and define $h_2 = i - h_1$. This process iterates to obtain subsequent h_j until the entire vector $\boldsymbol{C}$ has been exhausted, and the process yields a total number of J components. In general, and for all $1 \leq j \leq J$, the j-th component of the multidimensional index is defined as

$$h_j = i - \sum_{k=0}^{j-1} h_k ,$$

where i is the largest integer (for $\sum_{k=0}^{j-1} h_k + 1 \leq i \leq N$) satisfying

$$c_i \geq i - \sum_{k=0}^{j-1} h_k$$

and where $h_0 = 0$.

The multidimensional h index is thus defined as $\boldsymbol{H} = (h_1, h_2, ..., h_J)$. The iterative process that yields all the h_j can be stopped early so that only components satisfying $h_j > m$ for some reasonable minimum value m are included, or so that only K components

(regardless of their value) are obtained. We will refer to the trimmed vector obtained in these latter conditions as $\widetilde{\boldsymbol{H}}$.

Note that the number J of components of the multidimensional index $\boldsymbol{H}$ will vary among individuals in any given sample. For the purpose of comparing individuals, it is then useful to extend the definition of $\boldsymbol{H}$ so that it has any desired number of components with the convention that $h_j = 0$ for $j > J$.

The multidimensional index $\boldsymbol{H}$ has some properties that are worth pointing out:

1) $J \leq N$, with equality holding when $h_1 = 1$ so that $h_j = 1$ also for all $1 \leq j \leq J$. In general, J will be substantially smaller than N unless the number of papers receiving a single citation is large.

2) $h_j \geq h_{j+1}$ for all $1 \leq j < J$. This is natural because none of the papers contributing to h_{j+1} has received more than h_j citations.

3) $\sum_{j=1}^{J} h_j = N$. This is also natural because the components of $\boldsymbol{H}$ represent counts of papers that have been cited at least once, the overall number of which is N.

4) The number of papers that have been cited at least k times is given by

$$\sum_{h_j \geq k} h_j .$$

This sum conveys a useful generalization of the concept embodied in the h index, which represents the number of papers that have been cited at least h times and is given by this expression when $k = h_1$ provided $h_2 < h_1$.

Although the next section will present an empirical illustration of the performance of $\boldsymbol{H}$, a simple example is worth presenting now. Consider an individual who has published $M = 43$ papers of which $N = 38$ have been cited at least once with citation counts $\boldsymbol{C}$ = (42, 13, 11, 11, 10, 10, 10, 10, 9, 8, 7, 7, 7, 6, 5, 5, 5, 5, 5, 4, 4, 4, 4, 3, 3, 3, 3, 3, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1). Assuming that the minimum reasonable h_j is $m = 3$, it can easily be seen that $\widetilde{\boldsymbol{H}}$ =(9, 5, 5, 4, 3) whereas $\boldsymbol{H}$ = (9, 5, 5, 4, 3, 2, 2, 2, 2, 1, 1, 1, 1), yielding $J = 13$. All of the properties of $\boldsymbol{H}$ that were listed above can easily be verified in this example (for instance, $\sum_{j=1}^{13} h_j = 38 = N$; also, the number of papers that have received at least 5 citations is

$$\sum_{h_j \geq 5} h_j = 9 + 5 + 5 = 19.$$

Illustration

Data for this illustration come from a recent analysis of the publication records of 204 professors of Methodology of the Behavioral Sciences in Spain [GARCÍA-PÉREZ, 2009], a sample that virtually exhausts the population of Spanish professors in that field. Figure 1 shows the distribution of h indices in this sample, which reveals that h indices are certainly too low to allow fine distinctions among individuals as regards their research accomplishments. In addition, many individuals share the same h index

although it is unlikely that their publication records actually match quantitatively or qualitatively. The multidimensional index ***H*** might help differentiate these individuals.

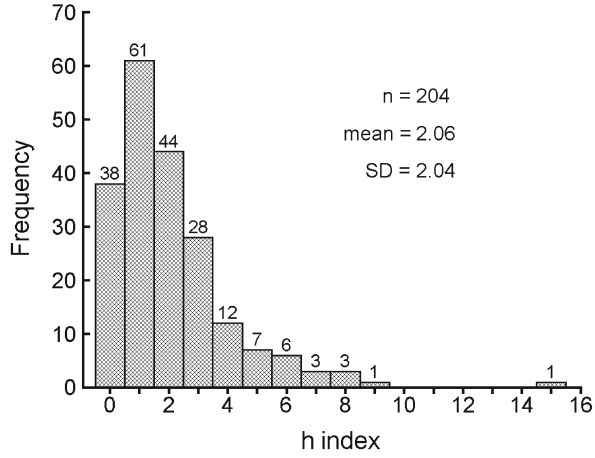

Figure 1. Distribution of the *h* index among tenured professors of Methodology of the Behavioral Sciences in Spain. Sample size, mean, and standard deviation (SD) are given in the inset. Numerals at the top of the bars indicate case frequencies

Raw data from this sample were thus re-analyzed to obtain the multidimensional index ***H*** for each individual, and Figure 2 shows tabulated scatterplots of h_2 against h_1 (Figure 2a), h_3 against h_1 (Figure 2b), and h_3 against h_2 (Figure 2c). This multidimensional approach allows making distinctions that were not possible with the simple *h* index. For instance, of the three individuals with an *h* of 8 (i.e., $h_1 = 8$), one has $h_2 = 2$ whereas the two other have $h_2 = 4$ (see Figure 2a); also, of the six individuals with $h_1 = 6$, three have $h_2 = 2$, two have $h_2 = 3$ and only one has $h_2 = 4$ (see Figure 2a); similarly, of the seven individuals with $h_1 = 5$, two have $h_3 = 0$, four have $h_3 = 2$, and only one has $h_3 = 3$ (see Figure 2b). Note also that there are 21 individuals with $h_1 \geq 5$ (i.e., with an *h* of 5 or higher), but only two of them have $h_2 \geq 5$ (see Figure 2a) and only one has $h_3 \geq 5$ (see Figure 2b).

Scatterplots in the panels of Figure 2 indicate that the first three components of ***H*** are positively correlated, which is natural given that they are subject to the order restriction $h_j \geq h_{j+1}$ for $1 \leq j < J$. The correlation between h_1 and h_2 is 0.862, the correlation between h_2 and h_3 is identically valued at 0.862, and the correlation between h_1 and h_3 is only slightly lower and valued at 0.780. In other words, consecutive components of ***H*** seem to be more strongly correlated than components that are further apart. In any case, these relatively high correlations do not lessen the utility of the additional dimensions of ***H*** because they indeed allow distinguishing individuals who are identical in terms of the preceding components (as was illustrated above).

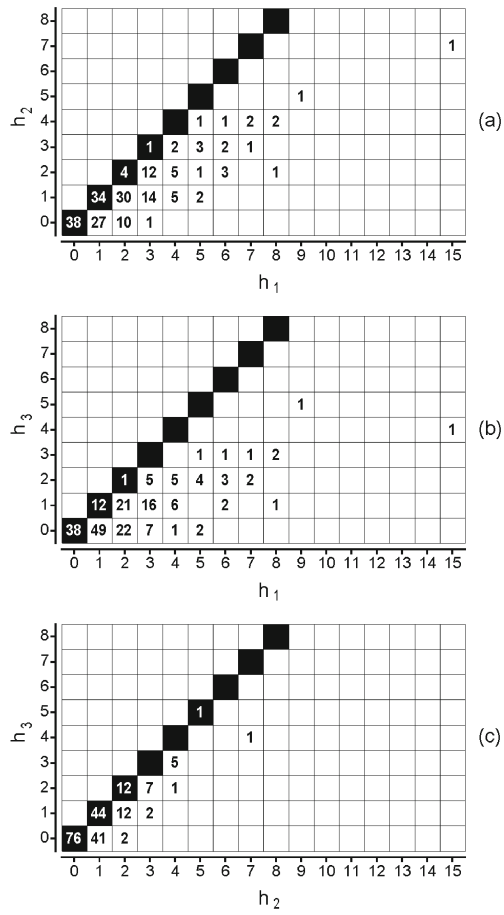

Figure 2. Tabulated scatterplots of h_2 against h_1 (a), h_3 against h_1 (b), and h_3 against h_2 (c) for the total sample of 204 professors. Numerals within the cells indicate frequencies; cells with zero frequencies are left blank. Equality of the two indices in each plot occurs along the diagonal (shaded cells)

It is also interesting to note that the additional components of ***H*** are also correlated with alternative bibliometric indicators that have been shown to correlate with Hirsch *h*. For instance, the correlation between *h* and *M* (i.e., the total number of papers published) has been reported to be 0.697 by VAN RAAN [2006] and the correlation between log *h* and log *M* has been reported by COSTAS & BORDONS [2008] to be 0.903. For the sample used in this paper, the correlation between *M* and h_1 (i.e., *h*) is 0.823, whereas the correlations between *M* and h_2 or h_3 are, respectively, 0.884 and 0.868. These three correlations are similar but they account for different aspects as indicated by the fact that the multiple correlation between *M* on the one hand and h_1, h_2, and h_3, on the other, is 0.913.

On the other hand, SAAD [2006] reported a correlation of 0.87 or 0.83 (depending on the source that provided the data) between h and total number N_c of citations,

$$N_c = \sum_{i=1}^{N} c_i \ ,$$

(with the notation introduced above), whereas VAN RAAN [2006] reported a slightly higher correlation valued at 0.938 and COSTAS & BORDONS [2008] reported an even higher correlation valued at 0.964 between log h and log N_c. Given the nonlinear relationship between h and N_c, the strength of their relation is better measured by the correlation between log h and log N_c as computed by COSTAS & BORDONS [2008]. From our data, the correlation between log N_c and log h_1 (i.e., log h) is 0.839, whereas the correlations between log N_c and log h_2 or log h_3 are, respectively, 0.689 and 0.675. Again, these three correlations account for different aspects as indicated by the fact that the multiple correlation between log N_c on the one hand and log h_1, log h_2, and log h_3, on the other, is 0.868.

Conclusion

This paper has proposed a multidimensional extension to Hirsch h index and has shown that the additional components are useful to distinguish individuals with the same h. The multidimensional extension uses the same logic as the original h and provides additional information under the same principles. The multidimensional extension should be useful in fields where h values are generally low, and it is perhaps more reasonably computed by removing self-citations which may substantially inflate h indices undeservedly in some circumstances [GARCÍA-PÉREZ, 2009; ZHIVOTOVSKY & KRUTOVSKY, 2008].

*

This research was supported by grant SEJ2005-00485 (Ministerio de Educación y Ciencia, Spain).

References

ANDERSON, T. R., HANKIN, R. K. S., KILLWORTH, P. D. (2008), Beyond the Durfee square: Enhancing the h-index to score total publication output. *Scientometrics*, 76 : 577.

BATISTA, P. D., CAMPITELI, M. G., KINOUCHI, O., MARTINEZ, A. S. (2006), Is it possible to compare researchers with different scientific interests? *Scientometrics*, 68 : 179.

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

BURRELL, Q. L. (2007), Hirsch index or Hirsch rate? Some thoughts arising from Liang's data. *Scientometrics*, 73 : 19.

COSTAS, R., BORDONS, M. (2008), Is g-index better than h-index? An exploratory study at the individual level. *Scientometrics*, 77 : 267.

EGGHE, L. (2006A), An improvement of the h-index: The *g*-index. *ISSI Newsletter*, 2 (1) : 8.
EGGHE, L. (2006B), Theory and practise of the *g*-index. *Scientometrics*, 69 : 131.
GARCÍA-PÉREZ, M. A. (2009), The Hirsch *h* index in a non-mainstream area: Methodology of the Behavioral Sciences in Spain. *Spanish Journal of Psychology*, in press.
HIRSCH, J. E. (2005), An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the U.S.A.*, 102 : 16569.
HIRSCH, J. E. (2007), Does the *h* index have predictive power? *Proceedings of the National Academy of Sciences of the U.S.A.*, 104 : 19193.
IGLESIAS, J. E., PECHARROMÁN, C. (2007), Scaling the *h*-index for different scientific ISI fields. *Scientometrics*, 73 : 303.
IMPERIAL, J., RODRÍGUEZ-NAVARRO, A. (2007), Usefulness of Hirsch's *h*-index to evaluate scientific research in Spain. *Scientometrics*, 71 : 271.
KELLY, C. D., JENNIONS, M. D. (2006), The *h* index and career assessment by numbers. *Trends in Ecology and Evolution*, 21 : 167.
KELLY, C. D., JENNIONS, M. D. (2007), H-index: Age and sex make it unreliable. *Nature*, 449 : 403.
KOSMULSKI, M. (2006), A new Hirsch-type index saves time and works equally well as the original h-index. *ISSI Newsletter*, 2 (3) : 4.
LEHMANN, S., JACKSON, A. D., LAUTRUP, B. E. (2007), Measures for measures. *Nature*, 444 : 1003.
LEHMANN, S., JACKSON, A. D., LAUTRUP, B. E. (2008), A quantitative analysis of indicators of scientific performance. *Scientometrics,* 76 : 369.
LIANG, L. (2006), *h*-index sequence and *h*-index matrix: Constructions and applications. *Scientometrics*, 69 : 153.
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 : 491.
RODRÍGUEZ NAVARRO, A., IMPERIAL RÓDENAS, J. (2007), *Índice h. Guía Para la Evaluación de la Investigación Española en Ciencia y Tecnología Utilizando el Índice h*. Madrid: Consejería de Educación de la Comunidad de Madrid. Retrieved January 8, 2009, from http://www.madrimasd.org /informacionidi/biblioteca/publicacion/doc/33_indiceh.zip.
ROUSSEAU, R. (2008), Reflections on recent developments of the h-index and h-type indices. *Collnet Journal of Scientometrics and Information Management*, 2 : 1.
RUANE, F., TOL, R. S. J. (2008), Rational (successive) *h*-indices: An application to economics in the Republic of Ireland. *Scientometrics*, 75 : 395.
SAAD, G. (2006), Exploring the h-index at the author and journal levels using bibliometric data of productive consumer scholars and business-related journals respectively. *Scientometrics*, 69 : 117.
SCHREIBER, M. (2007), A case study of the Hirsch index for 26 non-prominent physicists. *Annalen der Physik*, 16 : 640.
SCHUBERT, A. (2009), Using the h-index for assessing single publications. *Scientometrics*, in press.
SIDIROPOULOS, A., KATSAROS, D., MANOLOPOULOS, Y. (2007), Generalized Hirsch *h*-index for disclosing latent facts in citation networks. *Scientometrics*, 72 : 253.
TABER, D. F. (2005), Quantifying publication impact. *Science*, 309 : 2166.
VINKLER, P. (2007), Eminence of scientists in the light of the *h*-index and other scientometric indicators. *Journal of Information Science*, 33 : 481.
WENDL, M. C. (2007), H-index: However ranked, citations need context. *Nature*, 449 : 403.
ZHIVOTOVSKY, L. A., KRUTOVSKY, K. V. (2008), Self-citation can inflate *h*-index. *Scientometrics*, 77 : 373.