hg-index: A New Index to Characterize the Scientific Output of Researchers Based on the h- and g- Indices

S. Alonso *, F.J. Cabrerizo†, E. Herrera-Viedma‡, F. Herrera‡

November 25, 2008

Abstract

To be able to measure the scientific output of researchers is an increasingly important task to support research assessment decisions. To do so, we can find several different measures and indices in the literature. Recently, the h-index, introduced by Hirsch in 2005, has got a lot of attention from the scientific community for its good properties to measure the scientific production of researchers. Additionally, several different indicators, for example, the g-index, have been developed to try to improve the possible drawbacks of the h-index. In this paper we present a new index, called hg-index, to characterize the scientific output of researchers which is based on both h-index and g-index to try to keep the advantages of both measures as well as to minimize their disadvantages.

Keywords: h-index, g-index, bibliometric indicators, research evaluation

*Sergio Alonso (zerjioi@ugr.es) is with the Software Engineering Department, University of Granada. C/ Periodista Daniel Saucedo Aranda s/n. 18071 Granada, Spain.

†Francisco Javier Cabrerizo (cabrerizo@issi.uned.es) is with the Department of Software Engineering and Computer Systems, Distance Learning University of Spain (UNED). 28040 Madrid, Spain.

‡Enrique Herrera-Viedma (viedma@decsai.ugr.es) and Francisco Herrera (herrera@decsai.ugr.es) are with the Department of Computer Science and Artificial Intelligence, University of Granada. C/ Periodista Daniel Saucedo Aranda s/n. 18071 Granada, Spain.

1 Introduction

Nowadays, to measure the scientific output of researchers is an increasingly important task to support research assessment decisions as accepting research projects, contracting researchers or awarding scientific prizes.

To do so, there exist several different indicators that allow to quantify both the production of scientists and the impact of their publications. It is usually desirable to use a combination of those different indicators in order to obtain a global view of the scientific output of the researcher being evaluated [22, 29].

Some of the most commonly used indicators to measure the scientific output of researchers that we can find in the literature are [9, 15]:

- Production indicators: total number of published papers and number of papers published in a certain period of time.
- Impact indicators (usually based on the received citations): total number of citations (including or excluding self citations), average number of citations per paper, number and pecentage of significant papers (papers with more than a certain amount of cites) and number of citations of the most significant papers.
- Indicators based on the impact of the jounals: median impact factor of the journals where the papers are published, relative citation rates (document citations compared with the average citations of the papers in the journal) and normalized position of the journals (computed according to the location of the publication journals in the ranking of journals ordered by impact factor).

In the last few years, the scientific community has paid a lot of attention to a new index, introduced by Hirsch in 2005 and called the h-index [15]. It presents several good properties (for example, it is simple to compute and it takes into account both the quantity and impact of the publications). Many papers have been published about it [1, 3, 8, 10, 14, 23, 24, 30]. A comprehensive list of h-index related publications can be found at [7]. Additionally, some new indicators based on the h-index that try to overcome its limitations have been developed [4, 5, 13, 18, 19, 20, 27, 28]. Among them, we can find the g-index [11, 12].

The aim of this paper is to present a new index (called hg-index) to characterize the scientific output of researchers. This index is based on both the h- and g- indices and tries to keep the advantages of both measures while minimizing their disadvantages.

To do so, the paper is set as follows. In section 2 we introduce both the h- and g- indices as well as we point out some of their most interesting properties and drawbacks. In section 3 we present the new hg-index and we discuss its properties. Section 4 presents a practical example in which the new index is applied and where some of its benefits are shown. Finally, in section 5 we point out our conclusions.

2 Preliminaries: the h- and g- indices

The h-index was originally presented by Hirsch in 2005 [15]. The original definition was:

Definition 1: A scientist has index h if h of his/her N_p papers have at least h citations each, and the other $(N_p - h)$ papers have no more than h citations each.

One of its main advantages is that it measures both the quantity and the impact of the author's papers in a single measure, aspects that traditionally has been measured with several different indicators. Another benefit of this indicator is that it is quite simple to compute from the citation data available through the Web of Science of the ISI Web of Knowledge [17]. The h-index has been proven to be robust in the sense that it is insensitive to a set of lowly cited papers [30]. Additionally, the difficulty of increasing the h-index grows exponentially as all the most cited papers of the researcher have to receive new cites to obtain a higher index. Moreover, the h-index is insensitive to one or several outstandingly highly cited papers (which is usually considered as a drawback).

However, the h-index presents some drawbacks that have been pointed out in the literature [2, 4, 9, 16, 21, 25]. To overcome these issues several authors have proposed several variants of the h-index, each of them usually centering its attention on an specific aspect of the index. For example, the A-index [6, 18],

tries to incorporate the number of cites of the called Hirsch Core papers (the h most cited papers of the author), the AR-index [18, 19] which also introduces the age of the papers into the equation as the total number of cites of a paper is very sensitive to its age or the Dynamic h-index [13] which introduce some variations to make the h-index time-dependent.

One of the h- related indices that has got more attention is the called g-index. This index, presented by Egghe in 2006 [11, 12] was designed to provide more importance to the most cited papers of the author, as in the case of the h-index, it does not matter if a paper has more than h cites when computing the measure.

Example 1: Supposse that we want to compare the scientific production of two different researchers. The first one has published 30 papers. His 20 most cited papers have received 20 cites each. The second researcher has also published 30 papers but his 20 most cited papers have received 50 cites each and the rest less than 20 cites. According to the Hirsch definition, both have a h-index of 20 whilst it is obvious that the production of the second researcher has a higher impact factor.

The g-index is defined as follows:

Definition 2: A set of papers has a g-index g if g is the highest rank such that the top g papers have, together, at least g^2 citations. This also means that the top $g+1$ papers have less than $(g+1)^2$ cites.

It is easy to prove that $g \geq h$ [12]. However, although the g-index is successful in evaluating the production of a researcher incorporating the actual citations of his papers it also presents some drawbacks that have to be taken into account. For example, the g-index may be greately influenced by a very successful paper.

Example 2: Supposse that we want to compare the scientific production of two different researchers. The first researcher has published 30 papers but only one of those publications has been successful receiving 500 cites (we can think of a successful general review paper) and the rest have not received any cites. The second researcher has published 50 papers and all of them have received 10 cites (all her publications have good visibility). The g-index for the first researcher is 22 ($22^2 = 484 < 500$ [the cites of the best 22 papers], $23^2 = 529 > 500$ [the cites

of the best 23 papers]) whilst the g-index of the second one is 10 ($10^2 = 100$ [the cites of the best 10 papers], $11^2 = 121 > 110$ [the cites of the best 11 papers]). In this case both authors have the same total number of cites and the second one receive cites for all her papers, which can be interpreted as that all her work has bigger visibility and produces more interest in the scientific community. However, her g-index is much less than the g-index of the first researcher that only achieved a big hit paper but whose production (which is also lower than the second resercher's one) is almost unknown to the scientific community.

3 A New Index to Characterize Scientific Output of Researchers

In [26] Rousseau states:

> *"As to the h- and the g-index: they do measure different aspects of a scientist's publication list. Certainly the h-index does not tell the full story, and, although a more sensitive indicator than the h-index, neither does the g-index. Taken together, g and h present a concise picture of a scientist's achievements in terms of publications and citations."*

We do agree that both measures incorporate several interesting properties about the publications of a researcher and that both should be taken into account to measure the scientific output of scientists.

Therefore, we present a combined index, that we call the hg-index that tries to fuse all the benefits of both previous measures and that tries to minimize the drawbacks that each one of them presented.

Definition 3: The hg-index of a researcher is computed as the geometric mean of his h- and g- indices, that is:

$$hg = \sqrt{h \cdot g}$$

It is trivial to demonstrate that $h \leq hg \leq g$ and that $hg - h \leq g - hg$, that is, the hg-index corresponds to a value nearer to h than to g. This property

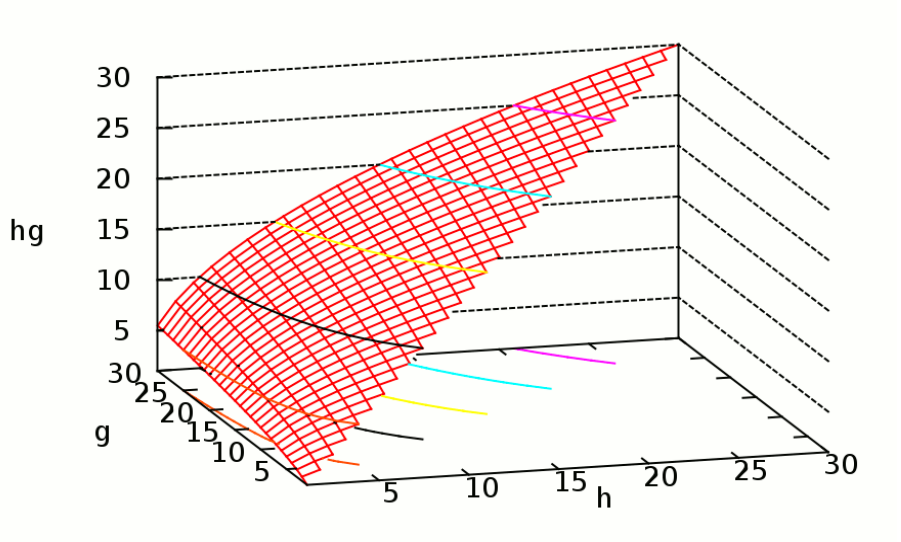

Figure 1: The growth in the hg-index as a function of h and g

can be seen as a penalization of the g-index in the cases of a very low h-index, thus avoiding the problem of the big influence that a very successful paper can introduce in the g-index. In figure 1 there is a representation of the growth of the hg-index as a function of h and g. From the figure it can be seen how the hg-index softens the influence of a high g-index when the h-index is low.

It is interesting to note that the hg-index can be interpreted in terms of geometry as the square root of the area of the rectangle with side lengths h and g.

In figure 2 we represent the hg-index of three different researchers. We can see that both Researcher A and Researcher B have the same hg-index ($hg_A = hg_B = 14.97 = \sqrt{h_A \cdot g_A} = \sqrt{h_B \cdot g_B}$) whilst Researcher C has a slightly bigger hg-index ($hg_C = 16.58 = \sqrt{h_C \cdot g_C}$).

Some additional the benefits of this new index are the following:

- It is very simple to compute once the h- and g- indices have been obtained.
- It provides more granularity than the h- and g- indices. This is specially interesting when compared with the h-index. As we have previously mentioned, to increase the h-index is difficult (more when the h-index is high) and it is usual to find that many different researchers have the same h-

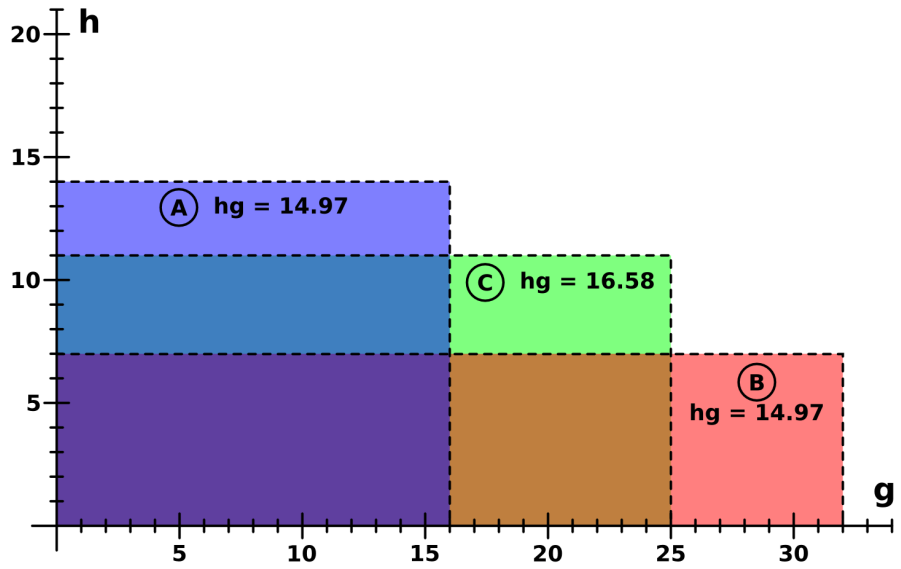

Figure 2: Geometrical interpretation of the hg-index

index with a very different number of publications and cites. The hg-index provides a more fine-grained way to compare scientists.

- The hg-index is valued in the same scale as both h- and g- indices (both represent the number of papers that comply with a condition about their cites). Thus, the hg-index it is easy to understand and to compare with those existing indices.

- It takes into account the cites of the highly cited papers (the h-index is insensitive to highly cited papers) but it significantly reduces the impact of single very high cited papers (a drawback of the g-index), thus achieving a better balance between the impact of the majority of the best papers of the author and very highly cited ones.

  Example 3: We part from example 2. The hg-index of the first researcher is 4.7 ($\sqrt{1 \cdot 22} = 4.7$) and the hg-index of the second researcher is 10 ($\sqrt{10 \cdot 10} = 10$). It can be seen how the hg-index has drastically minimized the effect of the very highly cited paper for the first researcher as the rest of his production has a very low impact. However, the hg-index of the second researcher maintains a good value as her production has a very

	h-**index**	g-**index**	**g/h**	hg-**index**
Braun	25	38	1.52	30.82
Egghe	13	19	1.46	15.72
Garfield	27	59	2.19	39.91
Glänzel	18	27	1.50	22.05
Ingwersen	13	26	2.00	18.38
Leydersdorff	13	19	1.46	15.72
Martin	16	27	1.69	20.78
Moed	18	27	1.50	22.05
Narin	27	40	1.48	32.86
Rousseau	13	15	1.15	13.96
Schubert	18	30	1.67	23.24
Small	18	39	2.17	26.50
Van Raan	19	27	1.42	22.65
White	12	25	2.08	17.32

Table 1: List of scientists with their h-, g-, g/h and hg- indices

constant citation rate. As it can be seen from the example, we believe that the hg-index provides a much more balanced measure of the impact of the researcher's papers.

4 Example of Application

In the following we present a more realistic example of the use of the hg-index in the evaluation of the scientific output of researchers. We part from the example given in [12] where some scientists where compared using the h- and g- indices and the g/h quotient.

We part from the h- and g- indices and the g/h quotient about each researcher and we additonally compute the hg-index. We show these data in table 1 (alphabetically ordered).

In the following tables we rank the different scientists according to the different measures that we have presented. Table 2 shows the rank of the researchers according to their h-index, table 3 according to the g-index, table 4 according

to the quotient g/h, table 5 according to their hg-index and table 6 according to a lexicographical order on the h-index and g-index.

	h-**index**
Garfield	27
Narin	27
Braun	25
Van Raan	19
Glänzel	18
Moed	18
Schubert	18
Small	18
Martin	16
Egghe	13
Ingwersen	13
Leydersdorff	13
Rousseau	13
White	12

Table 2: Scientists ranked by their h-index

	g-**index**
Garfield	59
Narin	40
Small	39
Braun	38
Schubert	30
Glänzel	27
Martin	27
Moed	27
Van Raan	27
Ingwersen	26
White	25
Egghe	19
Leydersdorff	19
Rousseau	15

Table 3: Scientists ranked by their g-index

The first thing to notice in the example is that the hg-index (as well as the g/h quotient and the lexicographical order) provides more granularity than any of the h- and g- indices separately. This is an advantage as it allows to provide a better rank between the researchers.

If we pay attention to the g/h quotient ranking we can see that White, who was the researcher with a lower h-index and also a low g-index is the third in the rank. That is because the g/h quotient cannot directly be used to rank the researchers as it is just a measure of how the h- and g- indices relate to each other. In general the g/h quotient can be used to identify the scientist with a greater disparity in both indices (which means that only a few of the publications receive many cites) and the scientists with similar h- and g- indices (all the best publications have an almost constant amount of cites).

The lexicographical order provides the same granularity as the hg-index but, in our oppinion, it overestimates the importance of the h-index. For example,

	g/h
Garfield	2.19
Small	2.17
White	2.08
Ingwersen	2.00
Martin	1.69
Schubert	1.67
Braun	1.52
Glänzel	1.50
Moed	1.50
Narin	1.48
Egghe	1.46
Leydersdorff	1.46
Van Raan	1.42
Rousseau	1.15

Table 4: Scientists ranked by their g/h quotient

	hg**-index**
Garfield	39.91
Narin	32.86
Braun	30.82
Small	26.50
Schubert	23.24
Van Raan	22.65
Glänzel	22.05
Moed	22.05
Martin	20.78
Ingwersen	18.38
White	17.32
Egghe	15.72
Leydersdorff	15.72
Rousseau	13.96

Table 5: Scientists ranked by their hg-index

	h**-index**	g**-index**
Garfield	27	59
Narin	27	40
Braun	25	38
Van Raan	19	27
Small	18	39
Schubert	18	30
Glänzel	18	27
Moed	18	27
Martin	16	27
Ingwersen	13	26
Egghe	13	19
Leydersdorff	13	19
Rousseau	13	15
White	12	25

Table 6: Scientists ranked by their h- and g- indices (lexicographical order)

in the case of comparing Van Raan and Small, the lexicographical order gives a bigger rank to Van Raan just because his h-index is one point higher, completely ignoring that the g-index of Small is much higher (meaning that his best publications have received together much more cites). In this case, the hg-index gives a more balanced rank between them, placing Small two positions higher than Van Raan in the rank.

From the example, we can say that generally the new hg-index provides a more balanced view of the scientific output of researchers than the h- and g-indices separately and that it provides a more fine-grained measurement that allows to compare scientists more efficiently.

5 Conclusions

In the last years the h-index, a measure of the scientific output of researchers based on both the quantity and impact of publications, has received great attention from the scientific community. Many papers have dealt with this index and have proposed new variations of the h-index (for example, the g-index) to overcome its drawbacks.

In this paper we have presented a new index, called the hg-index, which is based on the h- and g- indices and that fuses both measures in order to obtain a more balanced view of the scientific production of researchers and that minimizes some of the problems that they present. An empirical example shows the good behaviour of this measure.

References

[1] P. Ball, Index aims for fair ranking of scientists. *Nature* 436(7053) (2005) 900.

[2] J. Bar-Ilan, Which h-index? - A comparison of WoS, Scopus and Google Scholar. *Scientometrics* 74(2) (2008) 257–271.

[3] L. Bornmann, H.-D. Daniel, What Do We Know About the h Index? *Journal of the American Society for Information Science and Technology* 58(9) (2007) 1381–1385.

[4] L. Bornmann, R. Mutz, H.-D. Daniel, Are There Better Indices for Evaluation Purposes than the h Index? A Comparison of Nine Different Variants of the h Index Using Data from Biomedicine. *Journal of the American Society for Information Science and Technology* 59(5) (2008) 830–837.

[5] T. Braun, W. Glnzel, A. Schubert, A Hirsch-type index for journals. *The Scientist* 19(22) (2005) 8.

[6] Q.L. Burrell, On the h-index, the size of the Hirsch core and Jin's A-index. *Journal of Informetrics* 1(2) (2007) 170–177.

[7] h-index and Related Publications, maintained by F.J. Cabrerizo. `http://sci2s.ugr.es/bibliography/bib.php?id=3`. Accessed 2.06.2008.

[8] A. Cho, Your career in a number. *Science Now*. `http://sciencenow.sciencemag.org/cgi/content/full/2005/812/1` (2005). Accessed 26.06.2008.

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

[10] B. Cronin, L. Meho, Using the h-index to rank influential information scientists. *Journal of the American Society for Information Science and Technology* 57(9) (2006) 1275–1278.

[11] L. Egghe, An improvement of the h-index: The g-index. *ISSI Newsletter* 2(1) (2006) 8–9.

[12] L. Egghe, Theory and practise of the g-index. *Scientometrics* 69(1) (2006) 131–152.

[13] L. Egghe, Dynamic h-index: The Hirsch index in function of time. *Journal of the American Society for Information Science and Technology* 58(3) (2007) 452–454.

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

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

[16] J.E. Iglesias, C. Pecharromán, Scaling the h-index for different scientific ISI fields. *Scientometrics* 73 (2007) 303–320.

[17] Isi Web of Knowledge. `http://www.isiwebofknowledge.com/` Accessed 26.06.2008.

[18] B. Jin, The AR-index: complementing the h-index. *ISSI Newsletter* 3(1) (2007) 6.

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

[20] J.M. Levitt, M. Thelwall, Two new indicators derived from the h-index for comparing citation impact: Hirsch Frequencies and the Normalised Hirsch Index. In D. Torres-Salinas, H.F. Moed (Eds), *Proceedings of the 11th Conference of the International Society for Scientometrics and Informetrics* 2 (2007) pp. 876–877. Madrid, Spain: Spanish Research Council (CSIC).

[21] L. Leydesdorff, Caveats for the use of citation indicators in research and journal evaluations. *Journal of the American Society for Information Science and Technology* 59(2) (2008) 278–287.

[22] B.R. Martin The use of multiple indicators in the assessment of basic research. *Scientometrics* 36(3) (1996) 343–362.

[23] A. Molinari, J.F. Molinari, Mathematical aspects of a new criterion for ranking scientific institutions based on the h-index. *Scientometrics* 75(2) (2008) 339–356.

[24] C. Oppenheim, Using the h-index to rank infuential British researchers in information science and librarianship. *Journal of the American Society for Information Science and Technology* 58(2) (2007) 297–306.

[25] A. Purvis, The h index: playing the numbers game. *Trends in Ecology & Evolution* 21(8) (2006) 422–422.

[26] R. Rousseau, New developments related to the Hirsch index. *Science Focus* 1(4) (2006) 23–25 (in Chinese). English version available at: E-LIS: code 6376, http://eprints.rclis.org/archive/00006376/. Accessed 26.06.2008.

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

[28] A. Sidiropoulos, D. Katsaros, Y. Manolopoulos, Generalized Hirsch h-index for disclosing latent facts in citation networks. *Scientometrics* 72(2) (2007) 253–280.

[29] T.N. Van Leeuwen, M.S. Visser, H.F. Moed, T.J. Nederhof, A.F.J. Van Raan, The holy grail of science policy: Exploring and combining bibliometric tools in search of scientific excellence. *Scientometrics* 57(2) (2003) 257–280.

[30] J.K. Vanclay, On the Robustness of the h-Index. *Journal of the American Society for Information Science and Technology* 58(10) (2007) 1547–1550.