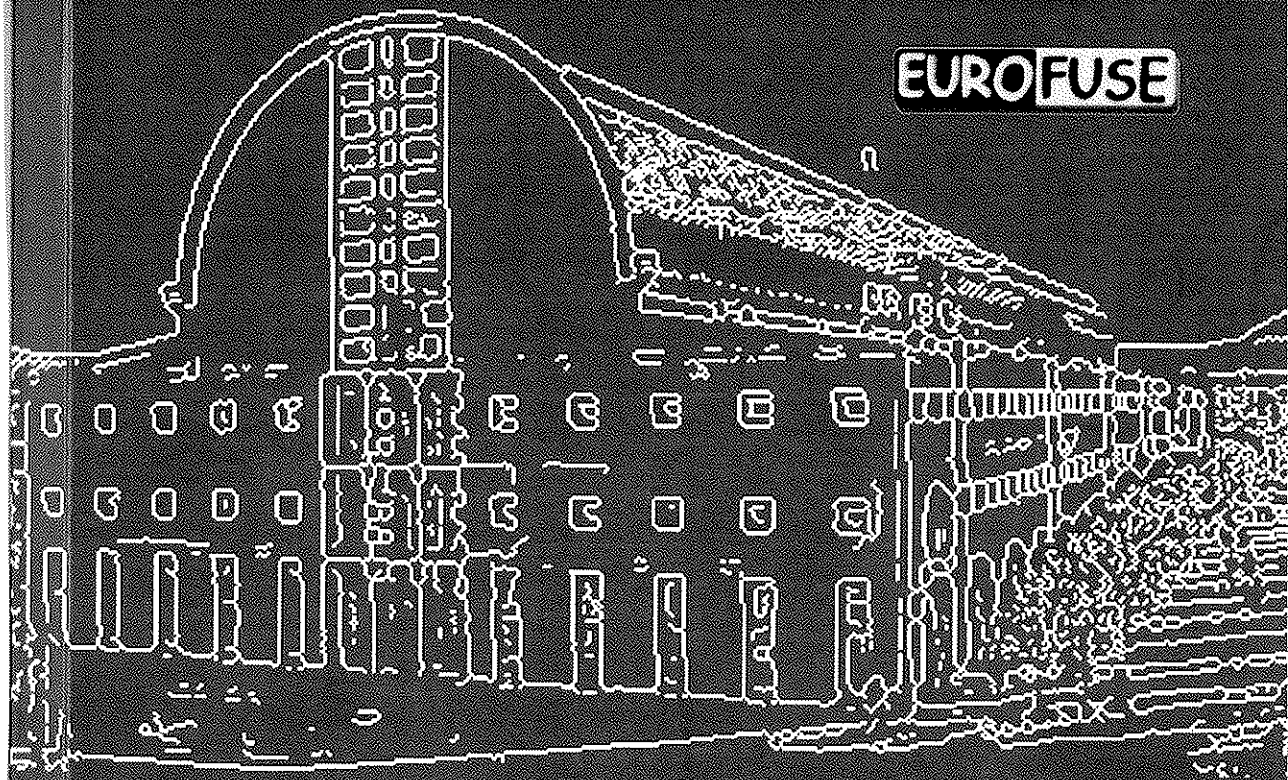
EUROFUSE



EUROFUSE WORKSHOP 09
PREFERENCE MODELLING AND DECISION ANALYSIS

September 16 – 18, 2009
Pamplona

Burillo, Pedro
Bustince, Humberto
De Baets, Bernard
Fodor, János
(editors)

Title: *EUROFUSE WORKSHOP'09. Preference Modelling and Decision Analysis*

Editors: Pedro Burillo, Humberto Bustince, Bernard De Baets and János Fodor

Edited by: Public University of Navarre

Printed by: Lankopi

D.L.: BI-2394/2009

ISBN: 978-84-9769-242-7

Distributed by: Sección de Publicaciones
Universidad Pública de Navarra
Campus de Arrosadía
31006 Pamplona
Fax: 948 169 300
e-mail: publicaciones@unavarra.es

Organising Committee

Chairmen

B. De Baets J. Fodor
P. Burillo H. Bustince

Members

E. Barrenechea M. Pagola
F.J. Fernández M. Galar
J. Sanz C. López-Molina
A. Jurío D. Paternain
C. Guerra A. Burusco
N. Frago A. Marín
V. Mohedano

Program Committee

U. Bodenhofer (Austria)
I. Bloch (France)
T. Calvo (Spain)
C. Carlsson (Finland)
J. Chamorro (Spain)
F. Chiclana (UK)
I. Couso (Spain)
S. Cubillo (Spain)
G. Deschrijver (Belgium)
S. Díaz (Spain)
J. Dombi (Hungary)
D. Dubois (France)
M. Fedrizzi (Italy)
R. Fuentes (Spain)
J.L. García-Lapresta (Spain)
I. Georgescu (Romania)
M. Grabisch (France)
F. Herrera (Spain)
E. Herrera-Viedma (Spain)
E. Hüllermeier (Germany)
E. Induráin (Spain)
R. Jensen (UK)
J. Kacprzyk (Poland)
E.P. Klement (Austria)
E. Kerre (Belgium)
A. Kolesárová (Slovak Republic)
M. T. Lamata (Spain)
J. Liu (UK)
V. Loia (Italy)
L. Martínez (Spain)
G. Mayor (Spain)
P. Melo-Pinto (Portugal)
R. Mesiar (Slovakia)
J. Montero (Spain)
J. A. Olivas (Spain)
E. Pap (Serbia)
G. Pasi (Italy)
J.I. Peláez (Spain)
H. Prade (France)
D. Radojevic (Serbia)
H. Rommelfanger (Germany)
D. Ruan (Belgium)
S. Sessa (Italy)
E. Szmidt (Poland)
P. Sobrevilla (Spain)
V. Torra (Spain)
J. Torrens (Spain)
E. Trillas (Spain)
J.L. Verdegay (Spain)
Z.S. Xu (China)
S. Zadrozny (Poland)

Bibliometry

Interdisciplinary Applications 2

The Aggregation of Bibliometric Indices to Evaluate the Scientific Output of Researchers: A Case of Study in the Fuzzy Community

F.J. Cabrerizo
Dept. Software Engineering and Computer Systems
Distance Learning University of Spain (UNED)
28040, Madrid, Spain
cabrerizo@issi.uned.es

S. Alonso
Dept. Software Engineering
University of Granada
18071, Granada, Spain
zerjioi@ugr.es

E. Herrera-Viedma
F. Herrera
Dept. Computer Science and A.I.
University of Granada
18071, Granada, Spain
{viedma,herrera}@decsai.ugr.es

Abstract

This paper examines how the aggregation of bibliometric indices is an appropriate way to combine information integrating the best qualities of every aggregated index. To do so, we analyze a new index, the *hg*-index, based on the aggregation of the Hirsch's h-index and Egghe's g-index and using the geometric mean as aggregation operator, and how it provides results that integrate the information from both indices, allowing the combination of their best qualities. Moreover, we study its correlation with the h- and g- indices and we analyze the results over a set of researchers specialized in fuzzy theories.

Keywords: Aggregation, Research Evaluation, Bibliometric Indices, h-index, g-index, Geometric Mean.

1 Introduction

To measure scientific output of researchers is an increasingly important task for the scientific community. In fact, nowadays, almost every research assessment decision (accepting research projects, contracting researchers or awarding scientific prizes) depends to a great extent upon the scientific merits of the involved researchers. To do so, the computation of bibliometric measures has attracted significant interest, due to the benefits of obtaining an unbiased and fair criterion [3, 19].

There exist several different indicators that allow the quantification of both the production of scientists and the impact of their publications. In 2005, J.E. Hirsch presented the h-index [14], which, in a short period of time, has became extremely popular. The original definition was:

> *"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."*

The h-index has attracted a lot of attention among scientometricians and information scientists, and it has been applied to a variety of areas [6, 8, 9] and it has been analyzed in some studies [12, 18]. Furthermore, Egghe [13] and Alonso et al. [2] have developed two review papers about the h-index, and a comprehensive list of h-index related publications can be found at the web page: `http://sci2s.ugr.es/hindex/biblio.php`.

The main advantage of the h-index is that it combines a measure of quantity and impact in a single indicator, aspects that traditionally have been measured separately by using different indices. However, the h-index presents other drawbacks that have been pointed out in the literature [4, 7]. To overcome these issues, several authors have proposed variants of the h-index, each of them usually centering its attention on a particular aspect of the h-index [5, 11, 15, 19]. One of the h- related indices that has got more attention is the so called g-index, presented by Egghe in 2006 [10]. The g-index is defined as follows:

> *"A scientist has 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 $(g+1)^2$ citations."*

As Bornmann et al. [5] point out, the different indices stand for very different dimensions of the scientist's research output, but they can complement each other very well. Therefore, the combination of different indices using some aggregation operator will provide us a more complete evaluation of the scientific production of researchers.

One of the aggregation operators, which can be used to combine the information provided by different bibliometric indices, is the geometric mean, because, among its properties, it takes into account all the aggregated values and it is not influenced by extremely high values, obtaining a value which fuses the information provided by every aggregated value.

In [17], 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."*

Following this idea, in [1], Alonso et al. present a combined index, called hg-index, based on the aggregation of the h- and g- indices and using the geometric mean as aggregation operator. This index tries to fuse all the benefits of both previous measures while minimizing the drawbacks that each one of them presented.

The aim of this paper is to examine how the aggregation of bibliometric indices, which measure different aspects of the scientific production of researchers, can provide us with a more complete evaluation of the scientific output of researchers than if only one index is used. To do so, we show how the hg- index fuses the information that the h- and g- indices provide separately. Furthermore, we study the correlation among the h-, g- and hg- indices and we analyze the results over a set of researchers specialized in fuzzy theories.

The paper is set out as follows. Section 2 introduces the geometric mean and some of its properties. In Section 3, we present the hg- index. In Section 4, the hg-index is applied in an example where some authors specialized in fuzzy theories are compared. In addition, the correlation among the h-, g- and hg- indices is studied. Finally, some concluding remarks are pointed out in Section 5.

2 The Geometric Mean

The geometric mean is a type of mean or average, which indicates the central tendency or typical value of a set of numbers. It is defined as the n-th root of the product of a set of n numbers. The geometric mean can be understood in terms of geometry as the geometric mean of two numbers, a and b, is simply the side length of the square whose area is equal to that of a rectangle with side lengths a and b.

Some properties of the geometric mean are the following:

- It is the only one.
- It takes into account all the values of the set of numbers.
- It only applies to positive numbers.
- It is the center of gravity of the set of numbers in multiplicative terms.
- It is more robust than the arithmetic mean to high values, but not to low values.
- It is more than or equal to the minimum value of the set of numbers and less than or equal to the arithmetic mean of the set of numbers.

In Figure 1, there is a representation of the growth of the geometric mean as function of a and b in $[0,1] \times [0,1]$. It can be seen how the geometric mean of two numbers, a and b, softens the influence of a high value of b when the value of a is low.

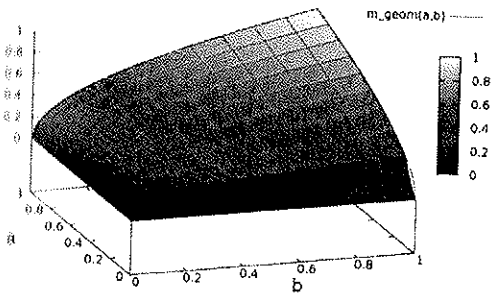


Figure 1: The growth of the geometric mean as a function of a and b in $[0,1] \times [0,1]$

3 The hg-Index

The h- and g- indices incorporate several interesting properties about the publications of a researcher and, therefore, both should be taken into account to measure the scientific output of researchers:

- The h-index mainly reflects the number of most cited articles (h-core) of a researcher, but the actual number of citations does not influence its value.
- The g-index combines the number of most cited articles of a researcher with the intensity of their citations.

In [1], Alonso et al. present the hg-index, that tries to fuse the different aspects of evaluation of both previous measures.

Definition 1. The hg-index of a researcher is computed as the geometric mean of his/her 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.

Some benefits of this index are the following:

- It is very simple to compute once the h- and g- indices have been obtained.
- 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 citations). Thus, the hg-index is easy to understand and to compare with those existing indices.
- It takes into account the number of citations 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.
- It provides more granularity than the h- and g- indices. This is specially interesting when compared with the h-index, because increasing 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-index with a very different number of publications and citations.

4 Case of Study

In this section, we analyze the behavior of the hg-index in comparison with the h- and g- indices in an example where some authors specialized in fuzzy theories are compared, and we study the correlation among these indices.

4.1 Example Based on Researchers in Fuzzy Logic and Soft Computing

In the following, we present an example of use of the hg-index in the evaluation of the scientific output of researchers. In [16], the top authors with the word fuzzy in the ISI Web of Knowledge are shown. We part from the fifteen most cited researchers and we compute the h-, g- and hg- indices about each one of them. This information has been collected on 01-05-2009 and it is shown in Table 2.

Table 1: Researchers ranked by their h-, g-, and hg- indices.

	h		g		hg
H. Prade	45	L.A. Zadeh	156	L.A. Zadeh	72.83
R.R. Yager	41	M. Sugeno	83	H. Prade	58.48
D. Dubois	41	R.R. Yager	76	R.R. Yager	55.82
J.C. Bezdek	39	H. Prade	76	D. Dubois	55.45
F. Herrera	38	D. Dubois	75	J.C. Bezdek	52.25
L.A. Zadeh	34	J.M. Mendel	74	J.M. Mendel	49.42
J.M. Mendel	33	J.C. Bezdek	70	F. Herrera	47.35
W. Pedrycz	33	F. Herrera	59	W. Pedrycz	41.82
S.K. Pal	28	W. Pedrycz	53	M. Sugeno	41.75
H. Ishibuchi	28	S.K. Pal	53	S.K. Pal	38.52
J.J. Buckley	27	H. Ishibuchi	51	H. Ishibuchi	37.79
N.R. Pal	22	J.J. Buckley	51	J.J. Buckley	37.11
M. Sugeno	21	N.R. Pal	48	N.R. Pal	32.50
D.A. Linkens	21	R. Lowen	40	R. Lowen	28.28
R. Lowen	20	D.A. Linkens	32	D.A. Linkens	25.92

Table 2: List of researchers with their h-, g- and hg- indices.

	h	g	hg
L.A. Zadeh	34	156	72.83
R.R. Yager	41	76	55.82
H. Prade	45	76	58.48
D. Dubois	41	75	55.45
M. Sugeno	21	83	41.75
J.M. Mendel	33	74	49.42
J.C. Bezdek	39	70	52.25
W. Pedrycz	33	53	41.82
S.K. Pal	28	53	38.52
N.R. Pal	22	48	32.50
R. Lowen	20	40	28.28
F. Herrera	38	59	47.35
H. Ishibuchi	28	51	37.79
J.J. Buckley	27	51	37.11
D.A. Linkens	21	32	25.92

If we pay attention to Table 2, we can see how the researchers obtain significant variations among the values of their indices. It proves that these indices stand for very different aspects of the scientist's research output: whereas the h-index reflects the core of most cited papers, but it does not take into account the actual number of citations, the g-index provides more importance to the most cited papers of the researcher. Finally, the hg-index takes into account both aspects of the scientific output of each researcher and, therefore, provides us a more balanced view of the scientific output of researchers.

In fact, if we rank the different researchers according to the different indices (Table 1), we can see that the hg-index provides more granularity than any of the h- and g- indices separately. This is an advantage of the aggregation of bibliometric indices as it is more likely to provide a complete ordering of researchers because it takes into account the information provided by every aggregated index.

If we compare M. Sugeno with respect to D.A. Linkens, we see that they present the same h-index (21), while their g-indices are very different (83 and 32, respectively). This is detected by the hg-index, which awards M. Sugeno with respect to D.A. Linkens. Similarly, if we compare W. Pedrycz with respect to S.K. Pal, we see that they present the same g-index (53), while their h-indices are very different (33 and 28, respectively). This also is detected by the hg-index, which awards W. Pedrycz with respect to S.K. Pal.

The problem is that the h- and g- indices measure different dimensions of the scientific research output. However, the hg-index di-

tinguishes better among researcher because it takes into account the information provided by the h- and g- indices.

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

4.2 Correlation Among the Indices Based on the Example

The fact that the h- and g- indices measure different dimensions of the scientist's research output can be proved by the weak correlation between these indices. To quantify it, as it is not clear whether the values of the indices follow a normal distribution, we have computed Spearman's rank-order correlation coefficients ρ (Table 3).

Table 3: Spearman's rank-order correlation coefficients ρ.

ρ	h	g	hg
h	1.000	0.637	0.886
g	-	1.000	0.917
hg	-	-	1.000

In this case, we show the Spearman's rank-order correlation coefficients among all the studied indices. These data speak for themselves. The correlation between the h- and g-indices is 0.637, which is low and can be explained because the h-index can be low while the g-index can be high due to that a few papers receive many citations. However, the correlation between the hg- and h- indices is 0.886 and between the hg- and g- indices is 0.917, i.e., the correlation between the hg-index and the h- and g- indices is high. To visualize it, Figure 2 shows the h- and g- indices in dependence on the hg-index.

Hence, taking into account the results drawn in the above section and the high correlation with the h- and g- indices, we can state that the hg-index is homogeneous, robust, has a good behavior and allows one to obtain an

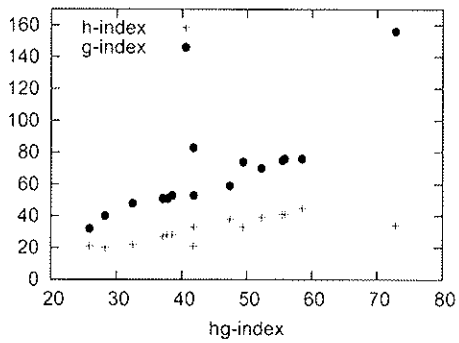


Figure 2: Scatter plot of h- and g- indices versus hg-index

adequate ranking among researchers. We see how the aggregation of bibliometric indices provides us a more complete ordering than if we use the indices separately.

5 Concluding Remarks

In this work, we have analyzed how the aggregation of bibliometric indices in an appropriate way can provide us a more complete view of the scientific output of researchers. Using the hg-index, which integrates the best qualities of the h- and g- indices using the geometric mean as aggregation operator, we have shown how it is possible to allow an index more complete without increasing the complexity in its computation. Therefore, the aggregation of bibliometric indices provide us significant advantages as a more granularity to compare researchers more efficiently and a more balanced view of the scientific output of researchers. However, although the hg-index shares the same scale that the h- and g- indices, it certainly does not have the ease of interpretation of either the h- or g- index.

Acknowledgements

This paper has been developed with the financing of FEDER funds with the PETRI project (PET2007-0460) and FUZZYLING project (TIN2007-61079).

References

[1] S. Alonso, F.J. Cabrerizo, E. Herrera-Viedma and F. Herrera. *hg*-index: A new index to characterize the scientific output of researchers based on the *h*- and *g*- indices. *Scientometrics* (2009). In press.

[2] S. Alonso, F.J. Cabrerizo, E. Herrera-Viedma and F. Herrera. *h*-index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of Informetrics* (2009). In press.

[3] J. Bar-Ilan. Informetrics at the beginning of the 21st century - A review. *Journal of Informetrics* 2(1) (2008) 1–52.

[4] L. Bornmann and 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.

[5] L. Bornmann, R. Mutz and 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.

[6] T. Braun, W. Glänzel and A. Schubert. A Hirsch-type index for journals. *Scientometrics* 69(1) (2006) 169–173.

[7] R. Costas and 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.

[8] B. Cronin and L. Meho. Using the *h*-index to rank influential information scientist. *Journal of the American Society for Information Science and Technology* 57(9) (2007) 1275–1278.

[9] E. Csajbók, A. Berhidi, L. Vasas and A. Schubert. Hirsch-index for countries based on Essential Science Indicators data. *Scientometrics* 73(1) (2007) 91–117.

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

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

[12] L. Egghe. Examples of simple transformations of the *h*-index: Qualitative and quantitative conclusions and consequences for other indices. *Journal of Informetrics* 2(2) (2008) 136–148.

[13] L. Egghe. The hirsch-index and related impact measures. *Annual Review of Information Science and Technology* (2009). In press.

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

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

[16] J.M. Merigó and A.M. Gil-Lafuente. An overview of fuzzy research in the ISI Web of knowledge. *Proceedings of the WCE*, London, England (2009). In press.

[17] R. Rousseau. New developments related to the Hirsch index. *Science Focus* 1(4) (2006) 23–25 (in Chinese). English version available at `http://eprints.rclis.org/6376/`.

[18] R. Rousseau. Woeginger's axiomatisation of the *h*-index and its relation to the *g*-index, the $h^{(2)}$-index and the r^2-index. *Journal of Informetrics* 2(4) (2008) 26[illegible] 372.

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