Journal of Informetrics 2 (2008) 298–303

Contents lists available at ScienceDirect

Journal of Informetrics

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

A symmetry axiom for scientific impact indices☆

Gerhard J. Woeginger

Department of Mathematics and Computer Science, TU Eindhoven, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

ARTICLE INFO

Article history:
Received 27 June 2008
Received in revised form 1 September 2008
Accepted 2 September 2008

Keywords:
Scientific impact measures
Citation analysis
Axiomatization

ABSTRACT

We provide a new axiomatic characterization of the Hirsch-index. This new characterization is based on a simple and appealing symmetry axiom which essentially imposes that the number of citations and the number of publications should be treated in the same way and should be measured in the same scale.

1. Introduction

Over the last decade, citation analysis has become the most popular method for research assessment. The prevalent idea is that research assessment should be done by using simple and objective methods. Whereas peer review suffers from possible subjectivity, citation analysis is based on simple numbers. Hence, citation analysis seems to be inherently more accurate than peer review.

Hirsch (2005) proposed the *h-index* (or *Hirsch-index*) as a simple tool for quantifying the scientific productivity and the scientific impact of an individual researcher. The *h*-index is based on the researcher's most cited articles and on the number of citations that they have received in other publications: "*A scientist has index h, if h of his or her n articles have at least h citations each, whereas the other n-h articles have at most h citations each.*" This *h*-index has many advantages: It is simple to understand and easy to compute. It can be applied to any level of aggregation. It is a robust indicator; see Rousseau (2007). It is hardly influenced by heavily cited publications. It is not influenced at all by unimportant (almost never cited) publications. Although the *h*-index is a fairly primitive indicator, it has attracted a lot of attention among scientometricians and information scientists, and it has been applied to a variety of areas; see for instance Ball (2005), Bornmann and Daniel (2005), Bornmann and Daniel (2007), Cronin and Meho (2006), Glänzel (2006), Hirsch (2007), Liu and Rousseau (2007), Oppenheim (2007) and van Raan (2006).

Of course there are more scientific impact indices, including variants that take into account the number of authors or the age of publications. Here we only want to mention three other impact indices.

- The *m-index* of a researcher is his/her *h*-index divided by the number of years since his/her first publication. This index was also proposed by Hirsch (2005), and its idea is to compensate young researchers at the beginning of their carreer, who did not yet have the time to publish many papers or to gather many citations.

☆ This research has been supported by the Netherlands Organisation for Scientific Research (NWO), Grant 639.033.403; by DIAMANT (an NWO mathematics cluster); and by BSIK Grant 03018 (BRICKS: Basic Research in Informatics for Creating the Knowledge Society).

E-mail address: gwoegi@win.tue.nl.

1751-1577/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.joi.2008.09.001

- The *g-index* of a researcher is the largest integer g such that his/her top g papers received together at least g^2 citations. This index was proposed by Egghe (2006a). The h-index does not take into account that some papers may have extraordinarily many citations, and the g-index tries to compensate for this; see also Egghe (2006b) and Tol (2008).
- The *w-index* of a researcher is the largest integer w such that his/her publication list contains w papers with respectively at least $1, 2, 3, 4, \ldots, w$ citations. This index was proposed by Woeginger (2008). The h-index tends to cluster many scientists into the same index value, whereas the w-index leads to a somewhat finer ranking.

In this article, we provide new axiomatic characterizations for the h-index and the w-index. Both indices have already been characterized in Woeginger (2008). Marchant (2008) provides an axiomatic characterization of the ranking of scientists that results from the h-index. For axiomatic characterizations of many other concepts in mathematical decision making, we refer the reader to Moulin (1988). The advantage of our new characterizations is that they are based on a new, simple and appealing axiom that imposes a symmetric handling of the number of citations and the number of publications of a scientist.

The article is organized as follows: Section 2 summarizes and recalls some basic definitions around scientific impact indices. Section 3 introduces and discusses our new symmetry axiom. Section 4 discusses the old characterizations of Woeginger (2008), and formulates the new characterizations that are based on the symmetry axiom. Sections 5 and 6 provide the proofs for these characterizations.

2. Basic definitions

The following set-up is mainly taken from Woeginger (2008). A researcher with $n \geq 0$ publications is represented by a vector $x = (x_1, \ldots, x_n)$ with non-negative integer components $x_1 \geq x_2 \geq \cdots \geq x_n$; the k th component x_k of this vector states the total number of citations to this researcher's k th-most important publication. We denote by X the set of all such vectors with non-increasing components. We say that a vector $x = (x_1, \ldots, x_n)$ is *dominated* by a vector $y = (y_1, \ldots, y_m)$, if $n \leq m$ holds and if $x_k \leq y_k$ for $1 \leq k \leq n$.

Definition 2.1. A scientific impact index is a function f from the set X into the set $\mathbb{N}$ of non-negative integers that satisfies the following three conditions:

- If x is the empty vector, then $f(x) = 0$.
- If $x = (x_1, \ldots, x_n)$ and $y = (x_1, \ldots, x_n, 0)$, then $f(x) = f(y)$.
- Monotonicity: If x is dominated by y, then $f(x) \leq f(y)$.

These three conditions are natural and fundamental: A researcher without output has no impact. Publications without citations have no impact, and hence cannot influence the impact of a researcher. If the citations to the output of researcher Y dominate the citations to the output of researcher X publication by publication, then Y has more impact than X.

The following definition provides a formal mathematical description of the h-index introduced by Hirsch (2005) and the w-index introduced by Woeginger (2008). An h-index of at least k means that there are k distinct publications that all have at least k citations. And a w-index of at least k means that there are k distinct publications that have at least $1, 2, 3, 4, \ldots, k$ citations, respectively.

Definition 2.2. The h-index is the scientific impact index $h : X \to \mathbb{N}$ that assigns to vector $x = (x_1, \ldots, x_n)$ the value $h(x) := \max\{k : x_k \geq k\}$.

The w-index is the scientific impact index $w : X \to \mathbb{N}$ that assigns to vector $x = (x_1, \ldots, x_n)$ the value $w(x) := \max\{k : x_m \geq k - m + 1 \textit{for all}\, m \leq k\}$.

Furthermore we mention the g-index, which assigns to vector $x = (x_1, \ldots, x_n)$ the value $g(x) := \max\{k : \sum_{i=1}^{k} x_i \geq k^2\}$. Finally, we introduce six other (perhaps somewhat artificial) scientific impact indices that will be useful in the rest of the paper.

Definition 2.3. Let $x = (x_1, \ldots, x_n)$ be a vector in X. Consider the following scientific impact indices.

- The maximum-index $f_{\max} : X \to \mathbb{N}$ is defined by $f_{\max}(x) := x_1$.
- The index $f_{nz} : X \to \mathbb{N}$ assigns to x the number of its non-zero components.
- The zero-index $f_0 : X \to \mathbb{N}$ is defined by $f_0(x) := 0$.
- The index $f_1 : X \to \mathbb{N}$ is defined by $f_1(x) := f_{\max}(x) + f_{nz}(x)$.
- The index $f_2 : X \to \mathbb{N}$ is defined by $f_2(x) := \max\{f_{\max}(x), f_{nz}(x)\}$.
- The index $f_3 : X \to \mathbb{N}$ is defined by $f_3(x) := \min\{f_{\max}(x), f_{nz}(x)\}$.

3. The symmetry axiom

Fig. 1 provides two geometric illustrations for the h-index. Publications are depicted on the horizontal axis, and the numbers $x_1 \geq x_2 \geq \cdots \geq x_n$ of citations per publication are on the vertical axis. The area under the curve gives the total

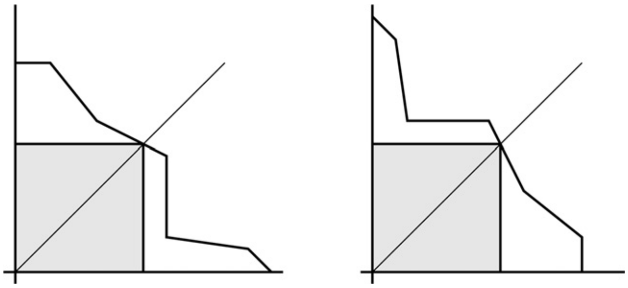

Fig. 1. Publications are on the horizontal axis, and the numbers $x_1 \geq x_2 \geq \cdots \geq x_n$ of citations per publication are on the vertical axis. The side length of the shaded square yields the h-index.

number of citations, and the side length of the shaded square yields the h-index. The publication curve on the right hand side results by reflecting the publication curve on the left hand side with respect to the 45° line.

We now provide a mathematical definition for such reflections of publication curves. Formally, we define for every vector $x=(x_1,\ldots,x_n)$ in X a corresponding reflected publication vector $R(x)=(x'_1,\ldots,x'_m)$, where

$$x'_\ell = |\{k : x_k \geq \ell\}|.$$

For instance, for $x=(6,3,3,2,1)$ we have $R(x)=(5,4,3,1,1,1)$. If vector x does not contain any zero-components, then $R(R(x))=x$. In the language of partitions of integers, x forms a partition of the integer $\sum_{k=1}^{n} x_k$ and $R(x)$ forms the corresponding conjugate partition; see Fig. 2 for an illustration. In a pictorial language, the publication curve corresponding to $R(x)$ is the reflection of the publication curve corresponding to x. The following axiom formulates a natural symmetry property for scientific impact indices.

S. For any $x \in X$ we have $f(x)=f(R(x))$

This symmetry property **S.** treats the data on the x-axis and the data on the y-axis in the same way and with the same scale. This conveys a sense of harmonious proportionality and balance, which pleases the human eye. Note that the reflection in Fig. 1 leaves the value of the h-index unchanged. This is in fact true for all publication vectors in X, and we think that at some psychological level this symmetry property **S.** might be one of the reasons for the tremendous success of the h-index.

Proposition 3.1. *The h-index and the w-index both satisfy axiom S.*

Now let us take a closer look at the impact indices introduced around Definition 2.3. The maximum-index $f_{\max}$ and the index f_{nz} behave in a kind of dual way with respect to reflections: They satisfy $f_{\max}(x)=f_{nz}(R(x))$ and $f_{nz}(x)=f_{\max}(R(x))$ for all $x \in X$. One immediate consequence of this observation is that $f_{\max}$ and f_{nz} both violate axiom **S**; for instance for $x=(3)$ with $R(x)=(1,1,1)$ we have $f_{\max}(x)=3 \neq 1=f_{\max}(R(x))$ and $f_{nz}(x)=1 \neq 3=f_{nz}(R(x))$. Another immediate consequence is the following proposition.

Proposition 3.2. *The scientific impact indices f_1, f_2, and f_3 satisfy axiom S.*

Also the zero-index f_0 (trivially) satisfies axiom **S**. Finally, we note that the g-index may violate axiom **S**: For $x=(9,0,0)$ with $R(x)=(1,1,1,1,1,1,1,1,1)$ we have $g(x)=3 \neq 1=g(R(x))$. This actually might be considered to be a point speaking against our symmetry axiom, since the g-index is generally perceived as a very natural impact index.

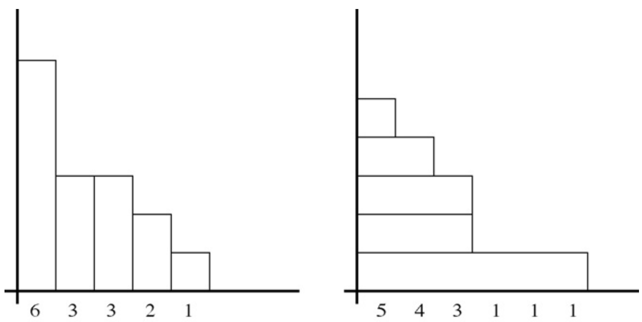

Fig. 2. The partition corresponding to $x=(6,3,3,2,1)$ to the left, and its conjugate partition corresponding to $R(x)=(5,4,3,1,1,1)$ to the right.

4. Axiomatic characterizations

We first review the axiomatic characterizations for the h-index and the w-index from Woeginger (2008). These characterizations are based on a number of axioms that capture certain desired elementary properties of a scientific impact index $f : X \to \mathbb{N}$. Axiom **S** states that if one adds a single publication to a publication list that is not substantially stronger than the current index, then this should not raise the index. Axiom A2 complements axiom A1: If one adds a single publication that is stronger than the current index, then the index should go up. Axioms B and C state that a minor change in the citation record should not lead to major changes in the index. Axiom D deals with the situation where both the number of publications and the number of citations go up.

A1. If the $(n+1)$-dimensional vector y results from the n-dimensional vector x by adding a new article with $f(x)$ citations, then $f(y) \leq f(x)$.
A2. If the $(n+1)$-dimensional vector y results from the n-dimensional vector x by adding a new article with $f(x)+1$ citations, then $f(y) > f(x)$.

B. If the n-dimensional vector y results from the n-dimensional vector x by increasing the number of citations of a single article, then $f(y) \leq f(x)+1$.
C. If the n-dimensional vector y results from the n-dimensional vector x by increasing the number of citations of every article by at most one, then $f(y) \leq f(x)+1$.
D. If the $(n+1)$-dimensional vector y results from the n-dimensional vector x by first adding an article with $f(x)$ citations and afterwards increasing the number of citations of every article by at least one, then $f(y) > f(x)$.

Woeginger (2008) showed that these axioms can be used to concisely characterize the h-index and the w-index.

Proposition 4.1. (Woeginger, 2008)
(i) *A scientific impact index $f : X \to \mathbb{N}$ satisfies the three axioms A1, B, and D, if and only if it is the h-index.*
(ii) *A scientific impact index $f : X \to \mathbb{N}$ satisfies the three axioms A2, B, and C, if and only if it is the w-index.*

Woeginger (2008) discusses these axioms, and states that A1, A2, C, and D are compelling, and precisely express one's intuition how a scientific impact index should behave. In contrast to this, axiom B is harder to justify. John Nash, for instance, has only published three influential papers in game theory before leaving the field; in 1994 he was awarded the Nobel prize in Economics for these three papers. Hence it is not at all obvious that a single very successful publication should never allow one's index to take off, as axiom B imposes.

In this article we will demonstrate that in Proposition 4.1 the somewhat unappealing axiom B may be replaced by the simple symmetry axiom S; see Theorems 4.2 and 4.3. Furthermore, we will show that our characterizations in these two theorems are tight in the following sense: We cannot drop any of the three characterizing axioms, without losing their uniqueness conclusion.

Theorem 4.2. (*Characterization of the h-index*).
A scientific impact index $f : X \to \mathbb{N}$ satisfies the three axioms A1, S, and D, if and only if it is the h-index.
Furthermore, there exist scientific impact indices that satisfy

(a) *the axioms D and S, but not A1*;
(b) *the axioms A1 and S, but not D*;
(c) *the axioms A1 and D, but not S*.

Theorem 4.3. (*Characterization of the w-index*).
A scientific impact index $f : X \to \mathbb{N}$ satisfies the three axioms A2, S, and C, if and only if it is the w-index.
Furthermore, there exist scientific impact indices that satisfy

(a) *the axioms C and S, but not A2*;
(b) *the axioms A2 and S, but not C*;
(c) *the axioms A2 and C, but not S*.

The h-index in fact satisfies the axioms A1, B, C, D, and S, but violates A2. The w-index satisfies the axioms A2, B, C, D, and S, but violates A1. The Theorems 4.2 and 4.3 will be proved in Sections 5 and 6, respectively.

5. The proofs for the *h*-index

In this section we prove Theorem 4.2. We first prove the characterization of the h-index in terms of the three axioms. One direction of the proof is straightforward, since the h-index clearly satisfies the three axioms A1, S, and D. For the other direction of the proof, we consider an arbitrary index f that satisfies axioms A1, S, and D. We start with the following technical lemma.

Lemma 5.1. *For $k \geq 0$, let $u^{[k]}$ denote the k-dimensional vector that consists of exactly k components of value exactly k. Then $f(u^{[k]}) = k$.*

Proof. First, we will argue that $f(u^{[k]}) \leq k$ holds for every $k \geq 0$. Suppose for the sake of contradiction that there exists a counter-example that satisfies $\ell := f(u^{[k]}) \geq k+1$.

- Let y denote the vector that consists of k components of value $k+1$ and of one component of value $\ell+1$. Axiom D yields $f(y) > f(u^{[k]}) = \ell$, and hence $f(y) \geq \ell+1$. The reflected vector $R(y)$ consists of $k+1$ components of value $k+1$ and of $\ell-k$ components of value 1. Axiom S yields $f(R(y)) = f(y) \geq \ell+1$.
- Let$z$ denote the vector that results from $u^{[k]}$ by adding $\ell+1$ components of value ℓ. By ℓ applications of axiom A1 we get $f(z) \leq f(u^{[k]}) = \ell$. Since$z$ also dominates $u^{[k]}$, we conclude $f(z) = \ell$.

Let us summarize: Vector $R(y)$ consists of $\ell+1$ components that are all less than or equal to $k+1$. Vector z contains $\ell+1$ components of value ℓ. Since $\ell \geq k+1$, vector$z$ dominates vector $R(y)$. Monotonicity yields $f(z) \geq f(y)$, which contradicts $f(z) = \ell$ or $f(R(y)) \geq \ell+1$. Hence $f(u^{[k]}) \leq k$ holds indeed for every $k \geq 0$.

Next, we prove by induction on $k \geq 0$ that $f(u^{[k]}) = k$. The statement for $k = 0$ follows from Definition 2.1. In the inductive step, we derive from the inductive assumption and from axiom D that $f(u^{[k+1]}) > f(u^{[k]}) = k$. Hence $f(u^{[k+1]}) \geq k+1$. The above upper bound $f(u^{[k+1]}) \leq k+1$ implies the desired $f(u^{[k+1]}) = k+1$. □

Now let us show that f coincides with the h-index. Consider an arbitrary vector $x = (x_1, \ldots, x_n)$, and let $k := h(x)$.

- The vector y results from $u^{[k]}$ by adding $x_1 - k$ components of value k. Axiom A1 and monotonicity together yield $f(y) = f(u^{[k]}) = k$.
- The vector $R(y)$ consists of k components of value x_1. Axiom S yields $f(R(y)) = k$.
- The vector$z$ results from $R(y)$ by adding $n-k$ components of value k. Axiom A1 and monotonicity together yield $f(z) = f(R(y)) = k$.

Vector z consists of k components of value x_1 and of $n-k$ components of value k. Hence$z$ dominates x, and monotonicity yields $k = f(z) \geq f(x)$. Since x dominates $u^{[k]}$, we also have $f(x) \geq k$. Altogether, this yields $f(x) = k = h(x)$. This completes the proof of the characterization part of Theorem 4.3.

To see that the characterization is tight, we use the following three indices. For (a) we use the w-index. For (b) we use the zero-index f_0 from Definition 2.3; the zero-index satisfies axioms A1, B, C, S, but violates A2 and D. For (c) we use the maximum-index $f_{\max}$ from Definition 2.3; the maximum-index satisfies axioms A1, A2, C, D, but violates B and S. This completes the proof of Theorem 4.3.

6. The proofs for the *w*-index

In this section we will establish Theorem 4.3. We first prove the characterization of the w-index in terms of the three axioms A2, S, and C. One direction of the proof is straightforward, since the w-index clearly satisfies the three axioms A2, S, and C. For the other direction of the proof, we consider an arbitrary index f that satisfies axioms A2, S, and C. We start with some technical lemmas.

Lemma 6.1. *If all components of a vector x are less than or equal to m, then $f(x) \leq m$. If a vector x has at most m non-zero components, then $f(x) \leq m$.*

Proof. The first statement follows by an easy inductive argument: We start from an all-zero vector, apply Definition 2.1, and then repeatedly apply axiom C. The second statement follows by applying the symmetry axiom S to the first statement. □

Lemma 6.2. *For $k \geq 0$, let $v^{[k]}$ denote the $(k+1)$-dimensional vector $(k, k-1, \ldots, 2, 1, 0)$. Then $f(v^{[k]}) = k$.*

Proof. The proof is done by induction. The case $k = 0$ is trivial. In the inductive step, the inductive assumption and axiom A2 together yield that $f(v^{[k+1]}) > f(v^{[k]}) = k$, and Lemma 6.1 yields $f(v^{[k+1]}) \leq k+1$. This gives $f(v^{[k+1]}) = k+1$, as desired. □

Lemma 6.3. *Let $x = (x_1, \ldots, x_n) \in X$ be a vector that satisfies $x_m \leq k-m+1$ for some $k \geq 0$ and $m \geq 1$. Then $f(x) \leq k$.*

Proof. The proof is done by induction on k. For $k = 0$, the only relevant case is $m = 1$. Then $x_1 \leq 0$ and $x = (0, 0, \ldots, 0)$, and the statement is implied by Definition 2.1.

In the inductive step, we consider a vector x that satisfies $x_m \leq k-m+2$ for some $m \geq 1$. We branch into two cases. The first case deals with $x_m = 0$. Then $m \leq k+2$, and vector x contains at most $k+1$ non-zero components. Then Lemma 6.1 yields $f(x) \leq k+1$, as desired. The second case deals with $x_m \geq 1$. We decrease all non-zero components in x by 1, while leaving the zero-components untouched. The resulting vector is denoted $y = (y_1, \ldots, y_n)$. Then $y_m = x_m - 1 \leq k - m + 1$, and the inductive assumption yields $f(y) \leq k$. Axiom C yields $f(x) \leq f(y) + 1 \leq k + 1$. $\square$

Now let us finally show that f coincides with the w-index. Consider an arbitrary vector $x = (x_1, \ldots, x_n)$, and let $k := w(x)$. According to Definition 2.2 then $x_m \geq k - m + 1$ must hold for all $m \leq k$.

- If $n = k$, then Lemma 6.1 yields $f(x) \leq k$. Since x dominates $v^{[k]}$, we also have $f(x) \geq k$. Hence $f(x) = k$.
- If $n > k$, then at least one of the inequalities $x_m \geq k - m + 1$ must hold as equality. This equality together with Lemma 6.3 yields $f(x) \leq k$. Since x dominates $v^{[k]}$, we also have $f(x) \geq k$. Hence $f(x) = k$.

In both cases, we end up with $f(x) = k = w(x)$. This completes the proof of the characterization part of Theorem 4.3.

To see that the characterization is tight, we use the following three indices. For (a) we use the h-index or the zero-index. For (b) we use the index f_1 from Definition 2.3. This index f_1 satisfies axioms A2, D, S, but violates A1, B, C. In particular axiom C is violated by $x = (2, 0, 0)$ and $y = (3, 1, 1)$ with $f_1(x) = 3$ and $f_1(y) = 6$. For (c) we use the maximum-index $f_{\max}$ from Definition 2.3; the maximum-index satisfies axioms A1, A2, C, D, but violates B and S. This completes the proof of Theorem 4.3.

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*(3), 391–392.

Bornmann, L., & Daniel, H. D. (2007). What do we know about the *h*-index? *Journal of the American Society for Information Science and Technology*, *58*(9), 1381–1385.

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

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

Egghe, L. (2006b). Theory and practice of the g-index. *Scientometrics*, *69*, 131–152.

Glänzel, W. (2006). On the *h*-index—A mathematical approach to a new measure of publication activity and citation impact. *Scientometrics*, *67*(2), 315–321.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, *102*(46), 16569–16572.

Hirsch, J. E. (2007). Does the *h*-index have predictive power? *Proceedings of the National Academy of Sciences*, *104*(49), 19193–19198.

Liu, Y. X., & Rousseau, R. (2007). Hirsch-type indices and library management: The case of Tongji University Library. In *Proceedings of the 11th International Conference of the International Society for Scientometrics and Informetrics (ISSI'2007)* (pp. 514–522).

Marchant, T. (in press). An axiomatic characterization of the ranking based on the *h*-index and some other bibliometric rankings of authors. *Scientometrics*.

Moulin, H. (1988). *Axioms of cooperative decision making*. Cambridge: Cambridge University Press.

Oppenheim, C. (2007). Using the *h*-index to rank influential British researchers in information science and librarianship. *Journal of the American Society for Information Science and Technology*, *58*(2), 297–301.

Rousseau, R. (2007). The influence of missing publications on the Hirsch index. *Journal of Informetrics*, *1*, 2–7.

Tol, R. S. J. (2008). A rational, successive *g*-index applied to economics departments in Ireland. *Journal of Informetrics*, *2*, 149–155.

van Raan, A. F. J. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgement for 147 chemistry research groups. *Scientometrics*, *67*(3), 491–502.

Woeginger G.J. (2008). An axiomatic characterization of the Hirsch-index. *Mathematical Social Sciences, 56*, 224–232.