

A new family of scientific impact measures: The generalized Kosmulski-indices

VLADIMIR G. DEINEKO,^a GERHARD J. WOEINGENGER^b

^a Warwick Business School, The University of Warwick, Coventry CV4 7AL, United Kingdom

^b Department of Mathematics and Computer Science, TU Eindhoven, Eindhoven, The Netherlands

This article introduces the generalized Kosmulski-indices as a new family of scientific impact measures for ranking the output of scientific researchers. As special cases, this family contains the well-known Hirsch-index h and the Kosmulski-index $h^{(2)}$. The main contribution is an axiomatic characterization that characterizes every generalized Kosmulski-index in terms of three axioms.

1. Introduction

In 2005, Jorge Hirsch proposed the h -index (or *Hirsch-index*) as a tool for quantifying the scientific productivity and the scientific impact of an individual researcher. This Hirsch-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.*" The Hirsch-index has many advantages: It is mathematically simple. It can be applied to any level of aggregation. It is a robust indicator; see [ROUSSEAU, 2007]. It is not strongly influenced by heavily cited publications. It is not influenced at all by unimportant (almost never cited) publications.

Although it is a fairly primitive indicator, the Hirsch-index has attracted a lot of attention among scientometricians and information scientists. Over the last few years it has been applied to a variety of areas; see for instance [BALL, 2005; BORNMAN & DANIEL, 2005; 2007; CRONIN & MEHO, 2006; GLÄNZEL, 2006; HIRSCH, 2007; LIU & ROUSSEAU, 2007; OPPENHEIM, 2007; VAN RAAN, 2006].

Recently KOSMULSKI [2006] has proposed another scientific impact index, that usually is called the $h^{(2)}$ -index or *Kosmulski-index*: "*A scientist has index k , if the top k of his or her n articles have at least k^2 citations each, whereas the remaining $n-k$ articles have at most $(k+1)^2-1$ citations each.*" KOSMULSKI [2006] states that in practice this index is easier to compute than the Hirsch-index: It reduces the amount of work that goes into the checking and verifying of author names, publications, and received

Received June 25, 2008; Published online April 16, 2009

Address for correspondence:

VLADIMIR G. DEINEKO

E-mail: Vladimir.Deineko@wbs.ac.uk

0138–9130/US \$ 20.00

Copyright © 2009 Akadémiai Kiadó, Budapest

All rights reserved

citations. On the other hand, the Kosmulski-index is still highly correlated with the total number of citations received. The Kosmulski-index has been studied for instance by LIU & ROUSSEAU [2007].

In this article, we will introduce a new family of scientific impact indices that form a common generalization of the Hirsch-index and of the Kosmulski-index. Our generalization is based on an infinite sequence $\langle s(1), s(2), s(3), s(4), \dots \rangle$ of critical thresholds values: “A scientist has index k , if the top k of his or her n articles have at least $s(k)$ citations each, whereas the remaining $n-k$ articles have at most $s(k+1)-1$ citations each.” It is not hard to see that for $s(k) = k$ this definition yields the Hirsch-index, and that for $s(k) = k^2$ this definition yields the Kosmulski-index. For every sequence $\langle s(1), s(2), s(3), \dots \rangle$ we provide an axiomatic characterization of the scientific impact index corresponding to that sequence. Our axiomatic characterizations follow and generalize the axiomatic characterization of the Hirsch-index provided by WOEGINGER [2008].

This article is organized as follows: Section 2 first provides some basic definitions around scientific impact indices, and then introduces the family of generalized Kosmulski-indices. Section 3 recalls the axiomatic characterization of the Hirsch-index in terms of three axioms as given in [WOEGINGER, 2008]. Then two of the underlying axioms are parameterized and appropriately generalized, so that they yield an axiomatic characterization of the generalized Kosmulski-indices. Sections 4 and 5 contain the proofs of our two main theorems on this characterization.

2. Scientific impact indices

The following set-up is essentially taken from WOEGINGER [2008]. A researcher with $n \geq 0$ publications is represented by a vector $x = (x_1, \dots, x_n)$ with non-negative integer components $x_1 \geq x_2 \geq \dots \geq x_n$; the k th component x_k of this vector states the total number of citations to this researcher’s k th-most important publication. For technical reasons, we will also assume that $x_k = 0$ holds for all indices $k \geq n+1$; this will simplify some of our arguments and definitions, and it will allow us to avoid tedious range checks for indices. Intuitively, these fictitious vector components correspond to fictitious publications without citations. Let X denote the set of all such vectors with non-increasing components. We say that a vector $x = (x_1, \dots, x_n) \in X$ is *dominated* by a vector $y = (y_1, \dots, y_n) \in X$, if $x_k \leq y_k$ holds for all $k \geq 1$.

Definition 2.1 A scientific impact index (or index, for short) 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 $x = (x_1, \dots, x_n)$ and $y = (x_1, \dots, x_n, 0)$, then $f(x) \leq f(y)$.
- Monotonicity: If x is dominated by y , then $f(x) \leq f(y)$.

These three conditions are fairly 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 Hirsch-index introduced by HIRSCH [2005] and the Kosmulski-index introduced by KOSMULSKI [2006].

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

The $h^{(2)}$ -index (or Kosmulski-index) is the scientific impact index $h^{(2)} : X \rightarrow \mathbb{N}$ that assigns to vector $x = (x_1, \dots, x_n)$ the value $h^{(2)}(x) := \max\{k : x_k \geq k^2\}$.

The next definition introduces a new family of scientific impact indices, which forms one of the main contributions of this article. We decided to call these new indices the *generalized Kosmulski-indices*, since their definition is heavily inspired by KOSMULSKI's definition [2006] of the Kosmulski-index $h^{(2)}$.

Definition 2.3 *A function $s : \mathbb{N} \rightarrow \mathbb{N}$ is well-behaved, if it satisfies $s(0) = 0$ and $s(k) \geq 1$ for all $k \geq 1$, and if it is non-decreasing on the set \mathbb{N} of non-negative integers.*

For any well-behaved function $s : \mathbb{N} \rightarrow \mathbb{N}$, the corresponding generalized Kosmulski-index $K[s]$ assigns to vector $x = (x_1, \dots, x_n)$ the value $K[s](x) := \max\{k : x_k \geq s(k)\}$.

We will also say that the function s is the underlying scaffold-function around which index $K[s]$ is built. Note that Definition 2.3 generalizes both indices in Definition 2.2: If the scaffold-function is the identity-function $\text{id}(k) = k$, then the corresponding index $K[\text{id}]$ is the Hirsch-index. If the scaffold-function is the square-function $\text{sqr}(k) = k^2$, then the corresponding index $K[\text{sqr}]$ is the Kosmulski-index. Another example for a generalized Kosmulski-index is the index that counts the number of publications that have been cited at least once: It corresponds to the scaffold-function s with $s(k) \equiv 1$ for all $k \geq 1$.

3. Characterizations around old and new axioms

In this section, we first recall three old axioms A1, B, and D from WOEGINGER [2008], and then present appropriately parameterized versions of A1 and D. Axiom A1 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 B states 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 $f(y) \leq f(x)$.

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 $f(y) \leq f(x)+1$.

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 three axioms concisely characterize the Hirsch-index.

Proposition 3.1 [WOEGINGER, 2008] *A scientific impact index $f: X \rightarrow \mathbb{N}$ satisfies the three axioms A1, B, and D, if and only if it is the Hirsch-index.*

Next, we will parameterize the above two axioms A1 and D in terms of a well-behaved scaffold-function $s: \mathbb{N} \rightarrow \mathbb{N}$.

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

D[s]. If the $(n+1)$ -dimensional vector y results from the n -dimensional vector x by first adding an article with $s(f(x))$ citations and afterwards increasing the number of citations of every article by at least $s(f(x)+1)-s(f(x))$, then $f(y) > f(x)$.

The special case where function s satisfies $s(k) = 1$ for all $k \geq 1$ trivializes the property in axiom A1[s]: It only concerns the addition of articles with at most 0 citations, and this situation is already fully covered by Definition 2.1. Hence, in this special case axiom A1[s] is empty. For all other well-behaved functions s , the property in axiom A1[s] truly restricts the behavior of a scientific impact index.

Furthermore, we note that if the scaffold-function is the identity-function $s(k) \equiv k$, then axioms A1[s] and D[s] boil down to the old axioms A1 and D, respectively. This observation gives some intuition and justification for the formulation of the two axioms A1[s] and D[s]. Since our goal is to get a separate axiomatic characterization for each generalized Kosmulski-index $K[s]$, it is natural to make also the axioms depend on the scaffold-function s .

The following theorem forms the main contribution of this article.

Theorem 3.2 *Let $s : \mathbb{N} \rightarrow \mathbb{N}$ be a well-behaved function. Then a scientific impact index $f : X \rightarrow \mathbb{N}$ satisfies the three axioms $A1[s]$, B , and $D[s]$, if and only if it is the generalized Kosmulski-index $K[s]$.*

Note that if the scaffold-function is the identity-function $s(k) \equiv k$, then Theorem 3.2 indeed boils down to Proposition 3.1. Theorem 3.3 shows that our characterization is tight: We cannot drop any of the three characterizing axioms, without losing the uniqueness conclusion.

Theorem 3.3 *Let $s : \mathbb{N} \rightarrow \mathbb{N}$ be a well-behaved function. Then there exist scientific impact indices that satisfy*

- (a) *the axioms $A1[s]$ and $D[s]$, but not B ;*
- (b) *the axioms $A1[s]$ and B , but not $D[s]$.*

Furthermore, if the well-behaved function s is not of the form $s(k) = 1$ for all $k \geq 1$, then there exists a scientific impact index that satisfies

- (c) *the axioms B and $D[s]$, but not $A1[s]$.*

Theorems 3.2 and 3.3 will be proved in Sections 4 and 5, respectively.

4. Proof of the characterization theorem

In this section we prove Theorem 3.2. Fix an arbitrary well-behaved scaffold-function $s : \mathbb{N} \rightarrow \mathbb{N}$, and consider the corresponding generalized Kosmulski-index $K[s]$. The if-part of the proof is fairly straightforward: Consider an n -dimensional vector $x = (x_1, \dots, x_n)$ with $K[s](x) = k$. This means that $x_k \geq s(k)$ and that $x_{k+1} < s(k+1)$.

- If the $(n+1)$ -dimensional vector y results from x by adding a new article with $s(k+1)-1$ citations, then $y_{k+1} < s(k+1)$ must hold. Hence $K[s](y) \leq k = K[s](x)$, and index $K[s]$ satisfies axiom $A1[s]$.

- If the n -dimensional vector y results from x by increasing the number of citations of a single article, then $y_{k+2} \leq x_{k+1} < s(k+1) \leq s(k+2)$. Hence $K[s](y) \leq k+1 = K[s](x)+1$, and index $K[s]$ satisfies axiom B .

- If the $(n+1)$ -dimensional vector y results from the n -dimensional vector x by first adding an article with $s(k)$ citations and afterwards increasing the number of citations of every article by at least $s(k+1)-s(k)$, then $y_{k+1} \geq s(k+1)$. Hence $K[s](y) \geq k+1 = K[s](x)$, and index $K[s]$ satisfies axiom $D[s]$.

For the only-if-part of the proof, we consider an arbitrary scientific impact index f that satisfies the three axioms $A1[s]$, B , and $D[s]$. We will show that index f coincides with the generalized Kosmulski-index $K[s]$.

Lemma 4.1 Any vector x with at most ℓ non-zero components satisfies $f(x) \leq \ell$.

Proof. Definition 2.1 yields that the vector y that consists of ℓ zeroes satisfies $f(y) = 0$. Next we repeatedly apply axiom B to y , and transform its first ℓ components step by step into the first ℓ components x_1, \dots, x_ℓ of vector x ; the resulting vector z agrees with x in all non-zero components. Axiom B implies $f(z) \leq f(y) + \ell$, and Definition 2.1 yields $f(x) \leq f(z)$. \square

Lemma 4.2 For $k \geq 0$, let $u^{[k]}$ denote the k -dimensional vector that consists of k components of value $s(k)$. Then $f(u^{[k]}) = k$.

Proof. The statement is proved by induction on $k \geq 0$. The statement for $k = 0$ follows from Definition 2.1. In the inductive step, axiom D with $x := u^{[k]}$ and $y := u^{[k+1]}$ yields $f(u^{[k+1]}) > f(u^{[k]}) = k$. Since Lemma 4.1 yields $f(u^{[k+1]}) \leq k+1$, we get the desired $f(u^{[k+1]}) = k+1$. \square

Now let us wrap up things. Consider an arbitrary vector $x = (x_1, \dots, x_n) \in X$, and let $k := K[s](x)$. Let $y = (x_1, \dots, x_k)$ denote the vector that consists of the first k components of x . Since these components all are at least $s(k)$, we get that $u^{[k]}$ is dominated by y . The monotonicity condition in Definition 2.1 implies $f(u^{[k]}) \leq f(y)$. Then Lemma 4.1 yields $f(y) \leq k$ and Lemma 4.2 yields $f(y) \geq k$; hence $f(y) = k$. Since vector x results from vector y by adding components of values at most $s(k+1) - 1$ to it, repeated application of axiom A1 together with monotonicity gives $f(x) = f(y) = k$. Therefore $f(x) = K[s](x)$ for all x .

This completes the proof of Theorem 3.2.

5. Proof of the tightness theorem

In this section we will prove Theorem 3.3. Fix an arbitrary well-behaved scaffold-function $s : \mathbb{N} \rightarrow \mathbb{N}$. In the proof of statement (a), we will distinguish two subcases. In the first subcase we assume that the function s is unbounded. We consider the index $f_1 : X \rightarrow \mathbb{N}$ that assigns to every vector $x = (x_1, \dots, x_n)$ the value

$$f_1(x) = \max \{k : x_1 \geq s(k)\}$$

Since the function s is unbounded, the index $f_1(x)$ is well-defined and does only take finite values. If $f_1(x) = k$ holds for some vector x , then adding a component of value $s(k+1) - 1$ will not increase the index, whereas adding a component of value $s(k+1)$ does increase it. Consequently, index f_1 satisfies axioms A1[s] and D[s]. Next, let p be an integer that satisfies $s(p-1) < s(p) \leq s(p+1)$. Then the two vectors $x = (s(p-1))$ and $y = (s(p))$ with $f(x) = p-1$ and $f(y) \geq p+1$ violate axiom B.

In the second subcase we assume that function s is bounded from above. Then $s(k)$ eventually becomes a constant function, and we let p denote the threshold where this occurs. In other words, we define $p \geq 1$ as the smallest integer for which $s(p) = s(p+k)$

holds for all $k \geq 1$. We define the index $f_2 : X \rightarrow \mathbb{N}$ that assigns to every vector $x = (x_1, \dots, x_n)$ the value

$$f_2(x) = \begin{cases} K[s](x) & \text{if } K[s](x) < p \\ 2 \cdot K[s](x) & \text{if } K[s](x) \geq p. \end{cases}$$

Let us quickly verify that f_2 satisfies axiom A1[s]: Consider a vector y with $K[s](x) = k$. If $k \leq p$, then $f_2(x) = k$; hence adding a component of value $s(k+1)-1$ will neither increase $K[s](x)$ nor $f_2(x)$. If $k > p$, then $f_2(x) = 2k$; hence adding a component of value $s(2k+1)-1 = s(p)-1 = s(k+1)-1$ will neither increase $K[s](x)$ nor $f_2(x)$. It is also easily verified that f_2 satisfies axiom D[s]. On the other hand index f_2 violates axiom B: Consider the p -dimensional vector x whose first $p-1$ components have value $s(p)$, whereas its last component has value 0. Then $K[s](x) = f_2(x) = p-1$. If we increase the last component to $s(p)$, then the resulting vector y has $K[s](y) = p$ and $f_2(y) = 2p$. Hence, $f_2(y) > f_2(x)+1$ violates B.

In the proof of statement (b), we consider the zero-index $f_3 : X \rightarrow \mathbb{N}$ that assigns to every vector x the value $f_3(x) = 0$. This trivial index satisfies the two axioms A1[s] and B, but violates axiom D[s].

In the proof of statement (c), we consider the index $f_4 : X \rightarrow \mathbb{N}$ that assigns to every vector $x = (x_1, \dots, x_n)$ the number of its non-zero components, that is, $f_4(x) = |\{k : x_k \geq 1\}|$. This index f_4 clearly satisfies axioms B and D[s]. Since the well-behaved function s is not of the form $s(k) = 1$ for all $k \geq 1$, there exists some $p \geq 1$ with $s(p) \geq 2$. Then the p -dimensional vector $x = (1, \dots, 1)$ has $f_4(x) = p$. But if we add a component of value $s(p+1)-1 \geq 1$ to it, then the resulting vector y will have $f_4(x) = p+1 > f_4(x)$. Hence, index f_4 violates axiom A1[s].

This completes the proof of Theorem 3.3.

*

This research has been supported by Warwick University's Centre for Discrete Mathematics and its Applications (DIMAP), by the Netherlands Organisation for Scientific Research (NWO), grant 639.033.403, and by BSIK grant 03018 (BRICKS: Basic Research in Informatics for Creating the Knowledge Society).

References

BALL, P. (2005), Index aims for fair ranking of scientists. *Nature*, 436 : 900.
 BORNMAN, L., H. D. DANIEL (2005), Does the h-index for ranking of scientists really work? *Scientometrics*, 65(3) : 391–392.
 BORNMAN, L., H. D. DANIEL (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., L. T. MEHO (2006), Using the h-index to rank influential information scientists. *Journal of the American Society for Information Science and Technology*, 57(9) : 1275–1278.

- 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.
- KOSMULSKI, M. (2006), A new Hirsch-type index saves time and works equally well as the original h-index. *ISSI Newsletter*, 2 : 4–6.
- LIU, Y., R. ROUSSEAU (2007), Hirsch-type indices and library management: The case of Tongji University Library. *Proceedings of the 11th International Conference of the International Society for Scientometrics and Informetrics (ISSI'2007)*, pp. 514–522.
- 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.
- 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.