Generalizations of Egghe's g-Index

Gerhard J. Woeginger
Department of Mathematics and Computer Science, Eindhoven Technical University, P.O. Box 513, 5600 MB Eindhoven, The Netherlands. E-mail: gwoegi@win.tue.nl

This paper introduces the generalized Egghe-indices as a new family of scientific impact measures for ranking the output of scientific researchers. The definition of this family is strongly inspired by Egghe's well-known g-index. The main contribution of the paper is a family of axiomatic characterizations that characterize every generalized Egghe-index in terms of four axioms.

Introduction

A Brief Summary of Scientific Impact Indices

Jorge Hirsch (2005) proposed the h-index (or Hirsch-index) as a tool for quantifying the scientific productivity and the scientific impact of an individual researcher: "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 main strength (and perhaps the main weakness) of the Hirsch-index is that it is a robust, very simple, and fairly primitive indicator. It has attracted a lot of attention, and it has been applied to a variety of areas; see for instance Ball (2005), Cronin and Meho (2006), Glänzel (2006), Liu and Rousseau (2007), Oppenheim (2007), and van Raan (2006).

Note that the h-index ignores the number of citations to each individual publication beyond what is needed to achieve a certain h-index. So what do we really know about a researcher with an h-index of 5? Perhaps this researcher is Professor X who has reached a meager total of 25 citations (5 papers with 5 citations). Or perhaps this researcher is the famous Mister Y whose research has attracted more than 1000 citations (5 papers with 200 citations plus another 50 papers each with 5 citations). Since the h-index fails to separate Professor X from Mister Y, Leo Egghe (2006a) proposed the so-called g-index, which assigns more weight to highly cited publications: "A scientist has index g if g is the largest integer such that his or her top g papers received together at least g^2 citations."

Note that the g-index of Professor X is a low 5, whereas the g-index of Mister Y equals a high 33. The g-index is constantly gaining popularity and visibility, and it is discussed by many authors; see for instance Egghe (2006b), Ruane and Tol (2008), Schreiber (2007, 2008), and Tol (2008).

Marek Kosmulski (2006) has proposed yet another scientific impact index, that usually is called the h^2-index or the 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 his index is significantly 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 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 and Rousseau (2007).

Deineko and Woeginger (in press) introduced a family of common generalizations of the Hirsch-index and the Kosmulski-index. These common generalizations are called the generalized Kosmulski-indices, and they are all built around certain infinite, nondecreasing sequences $\langle s(1), s(2), s(3), s(4), \dots\rangle$ of positive integers. "A scientist has index l if the top l of his or her n articles have at least $s(l)$ citations each, whereas the remaining $n-l$ articles have at most $s(l+1)-1$ citations each."

It is not hard to see that for the sequence $s(l) \equiv l$ this definition yields the Hirsch-index, and that for the sequence $s(l) \equiv l^2$ this definition yields the Kosmulski-index.

Wu (2008) discusses the Wu-index, which coincides with the generalized Kosmulski-index for the sequence $s(l) \equiv 10l$. Van Eck and Waltman (2008) introduce several generalizations of the Hirsch-index and of the g-index, and analyze their relationships with other indices. A distinguishing feature of these generalizations is that the indices are not restricted to integer values any more.

Received September 2, 2008; revised December 19, 2008; accepted January 19, 2009

 • Published online 20 February 2009 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.21061

Axiomatic Characterizations

Axiomatic analysis is one possible approach to analyzing and understanding the structure of mathematical decision rules. Which decision rules are the "best" rules to use? Which decision rules are able to meet our expectations? These are difficult questions that usually do not allow a clear answer. The central idea of axiomatic analysis is to describe a decision rule as a set of properties. If we understand which properties are satisfied and which properties are violated by a given decision rule, we will have a complete picture of its behavior. We refer the reader to Arrow (1951) and to Moulin (1988) for axiomatic characterizations of many concepts in mathematical decision making, and for an introduction into the underlying ideas.

With respect to scientific impact indices, Woeginger (2008b) characterized the h-index in terms of three axioms (called A1, B, and D), and Woeginger (2008a) characterized the g-index in terms of three other axioms (called E, T1, and T2). The axioms in these two characterizations are very simple and capture certain elementary, natural, desired properties of scientific impact indices. Marchant (in press) provides an axiomatic characterization in terms of six axioms for the ranking of scientists that results from the h-index, and for some other bibliometric rankings. For every sequence $\langle s(1), s(2), s(3) \dots \rangle$, Deineko and Woeginger (in press) provide an axiomatic characterization of the corresponding generalized Kosmulski-index. These axiomatic characterizations follow and naturally generalize the axiomatic characterization of the Hirsch-index in Woeginger (2008b).

Contribution of This Paper

We introduce a certain class of new generalizations of Egghe's g-index. These generalizations are built around the g-index in very much the same spirit as the generalized Kosmulski-indices of Deineko and Woeginger (in press) are built around the h-index and the Kosmulski-index. Every generalized Egghe-index is centered around an infinite nondecreasing sequence of positive integers. As a main result, we provide axiomatic characterizations of all these generalized Egghe-indices in terms of four axioms. Our axiomatic characterizations follow and generalize the axiomatic characterization of the g-index provided by Woeginger (2008a).

This article is organized as follows: The follwing section first provides several basic definitions and preliminaries around scientific impact indices. The next section then introduces the family of generalized Egghe-indices. The subsequent section recalls the axiomatic characterization of the g-index in terms of three axioms as given in Woeginger (2008a). Then one of the underlying axioms is parameterized and appropriately generalized, and another new technical axiom is introduced. The resulting four axioms yield the axiomatic characterization of the generalized Egghe-indices. The final sections contain the proofs of our two main results on this new characterization.

Definitions and Preliminaries

A researcher with $n \geq 0$ publications is represented by a vector $x = (x_1, \dots, x_n)$ with nonnegative integer components $x_1 \geq x_2 \geq \dots \geq x_n$; the kth component x_k of this vector states the total number of citations to the kth-most important publication. For technical reasons, we also assume that $x_k = 0$ holds for all components with indices $k \geq n+1$; this simplifies some of our arguments and definitions, and it allows us to avoid tedious range checks for indices. Intuitively, these fictitious zero-components correspond to fictitious publications without citations. Let X denote the set of all such vectors with nonincreasing components.

We say that a vector $x = (x_1, \dots, x_n) \in X$ is dominated by a vector $y = (y_1, \dots, y_m) \in X$, if $n \leq m$ and $x_k \leq y_k$ holds for all $k = 1, \dots, n$. We write $x \preceq y$ to denote this situation.

Definition 1. A scientific impact index (or index, for short) is a function f from the set X into the set N of nonnegative 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 $f(x) = f(y)$.
- Monotonicity: If $x \preceq y$, then $f(x) \leq f(y)$.

These three conditions are natural and easy to justify: A researcher without output has no impact. Publications without citations (and in particular the fictitious publications that correspond to the fictitious zero-components) 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.

Definition 2. Let $f: X \to \mathbb{N}$ be some scientific impact index, and let x be some element of X. Then the f-core of vector x consists of the first $f(x)$ components of x.

Note that because of the fictitious vector components with 0 citations, the f-core is indeed well-defined. Intuitively, the f-core of vector x contains the most important publications with respect to index f (that is, the author's core publications). One of our axioms (axiom T2 below) is centered around the f-core.

The Generalized Egghe-Indices

The following definition provides a formal mathematical description of the g-index introduced by Egghe (2006a).

Definition 3. The g-index is the scientific impact index $g: X \to \mathbb{N}$ that assigns to every vector $x = (x_1, \dots, x_n)$ the value $g(x) := \max\{k : \sum_{i=1}^{k} x_i \geq k^2\}$.

Another equivalent definition of the g-index is based on the function $A(k) = (\sum_{i=1}^{k} x_i)/k$, which specifies the average value of the k largest components of the vector $x = (x_1, \dots, x_n)$. This function $A(k)$ is a nonincreasing function in k, and the g-index is the maximum integer k satisfying $A(k) \geq k$.

DOI: 10.1002/asi

Definition 4. A function s: $\mathbb{N} \to \mathbb{N}$ is called gracious, if it satisfies the following three conditions:

1. $s(k) \geq 1$ for all $k \geq 1$;
2. s is nondecreasing, that is, $s(k-1) \leq s(k)$ for all $k \geq 1$;
3. s is convex, that is, $2s(k) \leq s(k-1) + s(k+1)$ for all $k \geq 1$.

In the following Definition 5, we finally introduce the main contribution of this article: A new and natural family of scientific impact indices. Since their definition is heavily inspired by Egghe's definition of the g-index, we decided to call these new indices the generalized Egghe-indices.

Definition 5. Let s: $\mathbb{N} \to \mathbb{N}$ be a gracious function. The corresponding generalized Egghe-index $G[s]$ assigns to every vector $x = (x_1, \ldots, x_n)$ in X the value $G[s](x) := \max\{k$: $\sum_{i=1}^{k} x_i \geq k \cdot s(k)]\}$.

It is easily seen that the generalized Egghe-index $G[s]$ as introduced in Definition 5 is well-defined: The average function $A(k)$ is nonincreasing and tends to 0, as k tends to infinity. The values $s(k)$ are nondecreasing and positive for $k \geq 1$. Hence there exists a unique largest integer k for which $A(k) \geq s(k)$, and this unique value coincides with $G[s](x)$. For all k smaller than this value we have $A(k) \geq s(k)$, and for all k larger than this value we have $A(k) \leq s(k)$.

Furthermore, we will say that the function s is the underlying scaffold-function on which the generalized Egghe-index $G[s]$ is built. Note that Definition 5 generalizes the index in Definition 3: If the scaffold-function is the gracious identity-function $\mathrm{id}(k) = k$, then the corresponding index $G[\mathrm{id}]$ is the g-index. Another example for a generalized Egghe-index is the index that counts the overall number of citations to all publications. It corresponds to the gracious scaffold-function s with $(k) \equiv 1$.

We conclude this section with a simple but useful observation.

Observation 1. Let s be a gracious function, and let $x = (x_1, \ldots, x_n)$ be some element of X with $G[s](x) = k$. Then any component of vector x outside the $G[s]$-core has value at most $s(k+1) - 1$.

Proof. Suppose otherwise. Then x contains $k+1$ components that are greater or equal to $s(k+1)$, and the sum of its first $k+1$ components would be at least $(k+1) \cdot s(k+1)$, which is a contradiction. □

Some Old and Some New Axioms

In this section, we will first recall the three old axioms E, T1, and T2 from Woeginger (2008a), and afterwards introduce two appropriately parameterized axioms. Axiom E states that by adding a strong new publication and consistently improving the citations to one's old publications, one should also raise one's index. Axioms T1 and T2 concern the transfer of citations between publications.

- E: If the $(n+1)$-dimensional vector y results from the n-dimensional vector x by first adding an article with $f(x)$ or with $f(x) + 1$ citations and afterwards increasing the number of citations of each article by one, then $f(y) = f(x) + 1$.
- T1: Let $x = (x_1, \ldots, x_n) \in X$, and let $1 \leq i < j \leq n$. If vector $y \in X$ results from x by setting $x_i := x_i + 1$ and $x_j := x_j - 1$, then $f(y) \geq f(x)$.
- T2: Let $x = (x_1, \ldots, x_n) \in X$, and let $1 \leq i < j \leq f(x)$ with $x_i \geq x_j + 2$. If vector $y \in X$ results from x by setting $x_i := x_i \pm 1$ and $x_j := x_j + 1$, then $f(y) = f(x)$.

The main result of Woeginger (2008a) shows that these three axioms together concisely characterize Egghe's g-index.

Proposition 1 (Woeginger, 2008a)**.** A scientific impact index f: $X \to \mathbb{N}$ satisfies the three axioms E, T1, and T2, if and only if it is the g-index.

Next, we will parameterize axiom E in terms of a gracious scaffold-function s: $\mathbb{N} \to \mathbb{N}$. Since our global goal is to get a separate axiomatic characterization for every generalized Egghe-index $G[s]$ for every gracious scaffold-function s, it is only natural to make the axioms depend on the function s.

- E[s]: Let the $(n+1)$-dimensional vector y result from the n-dimensional vector x by first adding an article with c citations, where c satisfies $s(f(x)) \leq c \leq s(f(x)+1)$, and by afterwards increasing the number of citations of every article by $s(f(x)+1) - s(f(x))$. Then $f(y) = f(x) + 1$ holds.

Furthermore, we introduce the following purely technical axiom Z[s]. Note that axiom Z[s] becomes vacuous in the cases where $s(1) = 1$ holds, since then its statement is already covered by Definition 1.

$Z[s]$: $f(s(1) = 1, \ldots s(1) - 1) = 0$.

Consider the special case where the scaffold-function is the identity-function $s(k) \equiv k$: Then axiom E[s] exactly boils down to the old axiom E (and this observation provides some intuition and justification for the formulation of E[s]). Furthermore, in this case the technical axiom Z[s] boils down to the condition $f(0, \ldots, 0) = 0$, which is implied by Definition 1.

The following theorem forms the main contribution of this article.

Theorem 1. Let s: $\mathbb{N} \to \mathbb{N}$ be a gracious function. Then a scientific impact index f: $X \to \mathbb{N}$ satisfies the four axioms E[s], T1, T2, and Z[s] if and only if it is the generalized Egghe-index $G[s]$.

Note that for the identity-function $s(k) \equiv k$, our characterization Theorem 1 indeed boils down to the characterization of Egghe's g-index in Proposition 1.

The following Theorem 2 demonstrates that our characterization in Theorem 1 is tight: We cannot drop any of the four characterizing axioms, without losing the uniqueness conclusion. Another interesting consequence of Theorem 2 is that none of these four axioms is implied by the other axioms.

DOI: 10.1002/asi

Theorem 2. Let s: $\mathbb{N} \to \mathbb{N}$ be a gracious function. Then there exist scientific impact indices that satisfy

1. the axioms T1, T2, and Z[s], but not E[s];
2. the axioms E[s], T2, and Z[s], but not T1;
3. the axioms E[s], T1, and Z[s], but not T2.
   Furthermore, if the gracious function s satisfies $s(1) \geq 2$, then there exists a scientific impact index that satisfies
4. the axioms E[s], T1, and T2, but not Z[s].

The proofs of Theorems 1 and 2 will be given below.

Proof of the Characterization Theorem

In this section we prove Theorem 1. Hence, fix an arbitrary gracious scaffold-function s: $\mathbb{N} \to \mathbb{N}$ and consider the corresponding generalized Egghe-index $G[s]$. We begin with a simple observation that will be applied many times in this section and also in the next section.

Observation 2. Let x and y be two vectors as defined in axiom E[s], and let $k = f(x)$. Then the newly added component in y has a value d that satisfies $s(k+1) \leq d \leq s(k+2)$.

Proof. Since c satisfies $s(k) \leq c \leq s(k+1)$ and since $d = c + s(k+1) - s(k)$, we get $s(k+1) \leq d \leq 2s(k+1) - s(k)$. The convexity condition 3 in Definition 4 yields $2s(k+1) - s(k) \leq s(k+2)$. □

For proving the if-part of Theorem 1, we will check that $G[s]$ indeed satisfies the four axioms E[s], T1, T2, and Z[s]. Consider an n-dimensional vector $x = (x_1, \ldots, x_n)$ with $G[s](x) = k$. This means that

$$\sum_{i=1}^{k} x_i \geq k \cdot s(k), \tag{1}$$

and that simultaneously

$$\sum_{i=1}^{k+1} x_i < (k+1) \cdot s(k+1). \tag{2}$$

We first consider axiom E[s]. Let the $(n+1)$-dimensional vector y result from x as described in the statement of axiom E[s]. Then vector y contains the k components $x_1 + s(k+1) - s(k), \ldots x_k + s(k+1) - s(k)$ and a newly added component of value d that satisfies $s(k+1) \leq d \leq s(k+2)$; see Observation 2. By using Inequality 1, we get that

$$\begin{aligned}\sum_{i=1}^{k+1} y_i &\geq d + \sum_{i=1}^{k}(x_i + s(k+1) - s(k)) \\ &\geq s(k+1) + k \cdot s(k) + k \cdot (s(k+1) - s(k)) \\ &= (k+1) \cdot s(k+1).\end{aligned} \tag{3}$$

Furthermore, Observation 1 yields $x_{k+2} < s(k+1)$. Then the corresponding component in vector y has value at most $2s(k+1) - s(k) \leq s(k+2)$, and also the newly added component satisfies $d \leq s(k+2)$; see Observation 2. Therefore the $k+2$ largest components in y are the $k+1$ values $x_1 + s(k+1) - s(k), \ldots, x_{k+1} + s(k+1) - s(k)$ and some component of value at most $s(k+2)$. By using Inequality 2 and the convexity condition 3 in Definition 3, we get

$$\begin{aligned}\sum_{i=1}^{k+2} y_i &\leq \sum_{i=1}^{k+1}(x_i + s(k+1) - s(k)) + s(k+2) \\ &< (k+1) \cdot s(k+1) + (k+1)(s(k+1) - s(k)) \\ &\quad + s(k+2) \\ &= (k+1) \cdot (2s(k+1) - s(k)) + s(k+2) \\ &\leq (k+2) \cdot s(k+2).\end{aligned} \tag{4}$$

The two Inequalities 3 and 4 together imply $G[s](y) = k+1$, and this establishes axiom E[s] for the generalized Egghe-index $G[s]$.

Also, axioms T1 and T2 are satisfied by index $G[s]$: Moving citations to the higher-value components in vector x can never lead to a violation of Inequality 1. And moving citations to smaller-value components within the $G[s]$-core of vector x will leave both crucial Inequalities 1 and 2 untouched. Finally, index $G[s]$ satisfies axiom Z[s] by definition.

For the proof of the only-if-part of Theorem 1, we consider an arbitrary scientific impact index f that satisfies the four axioms E[s], T1, T2, and Z[s]. We will show that index f coincides with the generalized Egghe-index $G[s]$. Our argument is based on the following four technical lemmas.

Lemma 1. For $k \geq 0$, let the k-dimensional vector $u^{[k]}$ consist of exactly k components of value exactly $s(k)$. Then $f(u^{[k]}) = k$ holds.

Proof. The proof is done by induction on $k \geq 0$. The statement for $k = 0$ follows from Definition 1. In the inductive step, we consider the two vectors $x = u^{[k]}$ and $y = u^{[k+1]}$. The inductive assumption yields $f(x) = k$. Note that vector y results from x by first adding an article with $c = s(k)$ citations and afterwards increasing the number of citations of every article by $s(k=1) + s(k)$. Now axiom E[s] yields $f(y) = f(x) + 1$, and hence $f(u^{[k+1]}) = k + 1$. □

Lemma 2. Every vector $x = (x_1, \ldots, x_n) \in X$ satisfies $f(x) \geq G[s](x)$.

Proof. Since the statement is trivial if $G[s](x) = 0$, we will throughout assume that $k := G[s](x) \geq 1$. This implies $\sum_{i=1}^{k} x_i \geq k \cdot s(k)$. We define t as the smallest index that satisfies $\sum_{i=1}^{t} x_i \geq k \cdot s(k)$; note that $t \leq k$. We define a k-dimensional vector $y = (y_1, \ldots, y_k)$ by setting $y_i = x_i$ for $1 \leq i \leq t-1$, by setting the t-th component $y_t = k \cdot s(k) - \sum_{i=1}^{t-1} y_i$, and by setting $y_i = 0$ for $t+1 \leq i \leq k$. Note that $\sum_{i=1}^{t} y_i = \sum_{i=1}^{k} y_i = k \cdot s(k)$.

First, observe that all components in y are bounded by the corresponding components in vector x. Hence $y \preceq x$, and $f(y) \leq f(x)$. Secondly, we claim that $f(u^{[k]}) \leq f(y)$ (with vector $u^{[k]}$ as introduced in Lemma 1). Indeed, define r as the largest index for which $y_r \geq s(k)$. Then one can produce

DOI: 10.1002/asi

vector y by starting from vector $u^{[k]}$ and by repeatedly moving citations from the components with index greater than r to the components with index at most r. Then axiom T1 yields $f(u^{[k]}) \leq f(y)$.

Summarizing, we have the two inequalities $f(y) \leq f(x)$ and $f(u^{[k]}) \leq f(y)$, which together with Lemma 1 imply $k \leq f(x)$. This completes the proof of the lemma. □

Lemma 3. For $k \geq 1$ and $l \geq 0$, let the $(k+l-1)$-dimensional vector $v^{[k,l]}$ consist of $k-1$ components of value $s(k)$ together with l components of value $s(k)-1$. Then $f(v^{[k,l]}) = k-1$ holds.

Proof. The proof is done by induction on $k \geq 1$. The statement for $k = 1$ concerns vectors of the form $(s(1)-1, \ldots, s(1)-1)$, and follows immediately from axiom Z[s].

In the inductive step, we consider the vectors $x = v^{[k,l]}$ and $y = v^{[k+1,l]}$. The inductive assumption yields $f(x) = k-1$. Vector y results from x by first adding an article with $c = s(k)$ citations, and afterwards increasing the number of citations of every article by $s(k+1) - s(k)$; note that c satisfies $s(f(x)) \leq c \leq s(f(x)+1)$. Now axiom E[s] yields $f(y) = f(x)+1$, and hence $f(v^{[k+1,l]}) = k$. □

Lemma 4. Every vector $x = (x_1, \ldots, x_n) \in X$ satisfies $f(x) \leq G[s](x)$.

Proof. Let $k = f(x)$. We modify vector x step by step in the following way: As long as the f-core contains two components x_i and x_j with $i < j$ and $x_i \geq x_j + 2$, we update $x_i := x_i - 1$ and $x_j := x_j + 1$. By axiom T2, this cannot change the value that index f takes for this vector. Eventually, this process must terminate with a vector y of the following shape: For some index t with $1 \leq t \leq k$, the first t components of y all have the same value m. The components y_i with $t+1 \leq i \leq k$ all have the same value, $m-1$. The components y_i with $i \geq k+1$ all take values at most y_k. Since f and $G[s]$ both satisfy axiom T2, we get $f(y) = f(x) = k$ and also $G[s](y) = G[s](x)$.

Now let us investigate the f-core of vector y. Suppose for the sake of contradiction that either (a) $m \leq s(k) - 1$ holds, or that (b) $m = s(k)$ and $t \leq k-1$ holds. In either case, vector y is dominated by the vector $v^{[k,n-k+1]}$ (which has been introduced and discussed in Lemma 3). This domination implies $f(y) \leq f(v^{[k,n-k+1]}) = k-1$, and contradicts $f(y) = k$. This contradiction establishes that either (c) $m = s(k)$ and $t = k$, or (d) $m \geq s(k)+1$ must hold. In either case we derive $\sum_{i=1}^{k} y_i \geq k \cdot s(k)$. This implies $G[s](y) \geq k$, which yields the desired inequality $G[s](y) \geq k = f(x)$. □

The statements in Lemma 2 and Lemma 4 together yield $f(x) = G[s](x)$ for every vector $x = (x_1, \ldots, x_n)$ in X. This completes the proof of Theorem 1.

Proof of the Tightness Theorem

In this section we will prove Theorem 2. Throughout this section, we consider some fixed gracious scaffold-function $s: \mathbb{N} \to \mathbb{N}$. The following straightforward lemma covers Part 1 of Theorem 2.

Lemma 5. Consider the zero-index f_a: $X \to \mathbb{N}$ that assigns to every vector the value 0. This index f_a satisfies the axioms T1, T2, and Z[s], but violates E[s].

Proof. The index f_a trivially satisfies the axioms T1, T2, and Z[s]. Furthermore the vectors $x = (0)$ and $y = (s(1), s(1))$ demonstrate that f_a violates axiom E[s]. □

The following Lemma 6 settles Part 2 of Theorem 2. It is centered around an index f_b that belongs to the generalized Kosmulski-indices as studied by Deineko and Woeginger (in press).

Lemma 6. Consider the index f_b: $X \to \mathbb{N}$ that assigns to every vector $x = (x_1, \ldots, x_n)$ the value $f_b(x) := \max\{l: x_l \geq s(l)\}$. This index f_b satisfies the three axioms E[s], T2, and Z[s], but violates T1.

Proof. The index f_b obviously satisfies Z[s]. Consider a vector $x = (x_1, \ldots, x_n)$ with $f_b(x) = k$. This is equivalent to the two inequalities $x_k \geq s(k)$ and $x_{k+1} < s(k+1)$. Clearly, neither of these two inequalities can be violated by moving citations to the smaller-value components within the first k components of x. Therefore, index f_b satisfies T2.

Next, let us consider axiom E[s]. Let the $(n+1)$-dimensional vector y result from x as described in the statement of axiom E[s]. Then vector y contains the k components $x_1 + s(k+1) - s(k), \ldots, x_k + s(k+1) - s(k)$ and a newly added component of value at least $s(k+1)$; see Observation 2. Since $x_k + s(k+1) - s(k) \geq s(k+1)$, we conclude $y_{k+1} \geq s(k+1)$. The convexity condition 3 in Definition 4 implies

$$y_{k+2} \leq x_{k+1} + s(k+1) - s(k) < 2s(k+1) - s(k) \leq s(k+2).$$

The two inequalities $y_{k+1} \geq s(k+1)$ and $y_{k+2} \geq s(k+2)$ together yield $f_b(y) = k+1$. Hence, index f_b satisfies E[s].

Finally, consider the two vectors $x = (s(2), s(2))$ and $y = (s(2)+1, s(2)-1)$. Since $f_b(x) = 2$ and $f_b(y) < 2$, index f_b violates axiom T1. □

Now let us turn to Part 3 of Theorem 2. This is somewhat tedious, since we have to distinguish two cases on function s: In the first case we assume that the function s is bounded from above. Since s also satisfies the convexity condition 3 in Definition 4, it then must be a constant function with $s(l) \equiv p$ for all $l \geq 1$. This first case is handled in Lemma 7. In the second case we assume that the function s is unbounded. This second case is handled in Lemma 8.

Lemma 7. Assume that the gracious function s is bounded, and satisfies $s(l) \equiv p$ for all $l \geq 0$. Consider the index f_c: $X \to \mathbb{N}$ that assigns to every vector $x = (x_1, \ldots, x_n)$ the value.

- $f_c(x) = G[s](x)$, if $x_1 \leq p$;
- $f_c(x) = G[s](x) + 1$, if $x_1 \geq p + 1$

DOI: 10.1002/asi

This index f_c satisfies the three axioms E[s], T1, and Z[s], but violates T2.

Proof. The index f_c violates axiom T2, since for $x=(p+1,\ p-1,\ 0)$ and $y=(p,\ p,\ 0)$, we have $f_c(x)=3$ but $f_c(y)=2$. The index f_c satisfies axiom Z[s], since $G[s](p-1,\ldots,p-1)=0$ implies $y_c(p-1,\ldots,p-1)=0$. We claim that index f_c also satisfies axiom T1: Indeed, consider a vector y that results from vector x by moving citations to the higher-value components. Since $G[s]$ satisfies axiom T1, we have $G[s](x)\leq G[s](y)$. Furthermore we have $x_1\leq y_1$. This implies the desired $f_c(x)\leq f_c(y)$.

It remains to show that index f_c satisfies axiom E[s]. Consider an arbitrary n-dimensional nonzero vector $x=(x_1,\ldots,x_n)\in X$ with $f'_c(x)=k$. Then the $(n+1)$-dimensional vector y in the statement of axiom E[s] results by adding one new component of value p to vector x. This implies $G[s](y)=G[s](x)+1$, and $y_1\geq p+1$ if and only if $x_1\geq p+1$. Hence the desired $f_c(y)=f_c(y)+1$ holds. □

Lemma 8. Assume that the gracious function s is unbounded. The index f'_c: $X\to\mathbb{N}$ assigns to every vector $x=(x_1,\ldots,x_n)$ with $x_1<s(1)$ the value $f'_c(x)=0$, and to every vector $x=(x_1,\ldots,x_n)$ with $x_1\geq s(1)$ the largest integer k that satisfies one of the following three conditions:

- $x_1=s(k)$ and $x_2=x_3=\ldots=x_k=s(k)$;
- $s(k)<x_1<s(k+1)$;
- $x_1=s(k+1)$ and $x_{k+1}<s(k+1)$.

The index f'_c satisfies the three axioms E[s], T1, and Z[s], but violates T2.

Proof. Note that the index f'_c is well-defined, as the function s is unbounded. Let us argue that index f'_c violates axiom T2: Since s is unbounded, there exists an integer $m\geq 2$ with $3\leq s(m)<s(m+1)$. Then $f'_c(s(m)+1,\ 0)=m$, whereas $f'_c(s(m),1)<m$; this contradicts with axiom T2. Furthermore, index f'_c satisfies axiom Z[s] by definition, and it is easy to see that f'_c satisfies axiom T1.

It remains to show that index f'_c satisfies axiom E[s]. Consider an arbitrary n-dimensional vector $x=(x_1,\ldots,x_n)\in X$ with $f'_c(x)=k$. Let the $(n+1)$-dimensional vector $y=(y_1,\ldots,y_{n+1})$ result from x as described in the statement of axiom E[s]. By Observation 2, the newly added component of value d satisfies $s(k+1)\leq d\leq s(k+2)$. First we discuss the case where $f(x)=k=0$. Then $s(1)\leq d\leq s(2)$ implies $y_1\geq s(1)$, and hence $f(y)\geq 1$. Furthermore $x_1<s(1)$ implies $y_2<2s(1)-s(0)\geq s(2)$, and hence $f(y)\leq 1$. This yields the desired $f'_c(y)=f'_c(x)+1$.

From now on we assume $f(x)=k\geq 1$. We branch into three cases that correspond to the three cases in the definition of f'_c.

- If $x_1=s(k)$ and $x_2=x_3=\ldots=x_k=s(k)$, then y contains k components of value $s(k+1)$ plus the new component d with $s(k+1)\leq d\leq s(k+2)$. Then $f'_c(y)=k+1$.
- If $s(k)<x_1<s(k+1)$ holds, then $s(k+1)<x_1+s(k+1)-s(k)<s(k+2)$. The newly added component d in y satisfies $s(k+1)\leq d\leq s(k+2)$. This implies that the largest component y_1 in y satisfies $s(k+1)<y_1\leq s(k+2)$, and that the second-largest component satisfies $y_2<s(k+2)$. Hence $f'_c(y)=k+1$.
- Finally assume $x_1=s(k+1)$ and $x_{k+1}<s(k+1)$. Then $s(k+1)\leq x_1+s(k+1)-s(k)\leq s(k+2)$, and $x_{k+1}+s(k+1)-s(k)<s(k+2)$. The newly added component d in y satisfies $s(k+1)\leq d\leq s(k+2)$. This implies $s(k+1)\leq y_1\leq s(k+2)$, and $y_{k+2}<s(k+2)$. Hence $f'_c(y)=k+1$.

Since all three cases led to $f'_c(y)=f'_c(y)+1$, index f'_c indeed satisfies axiom E[s]. □

Finally, Lemma 9 settles Part 4 of Theorem 2.

Lemma 9. Assume that the gracious scaffold function s satisfies $s(1)\geq 2$. Consider the impact index f_d: $X\to\mathbb{N}$ that assigns to every vector $x=(x_1,\ldots,x_n)$ the value $f_d(x)=\max\{l\colon \sum_{i=1}^{l}x_i\geq l\cdot s(l)-1\}$. This index f_d satisfies the three axioms E[s], T1, and T2, but violates Z[s].

Proof. The index f_d clearly violates axiom Z[s], since $f_d(s(1)-1,\ldots,s(1)-1)=1$. For axioms E[s], T1, and T2, consider a vector x with $f_d(x)=k$. This means that

$$\sum_{i=1}^{k}x_i\geq k\cdot s(k)-1, \tag{5}$$

and that

$$\sum_{i=1}^{k+1}x_i<(k+1)\cdot s(k+1)-1. \tag{6}$$

The index f_d satisfies axiom T1, since moving citations to the higher-value components in vector x can never violate Inequality 5. The index f_d satisfies axiom T2, since moving citations to smaller-value components within the f_d-core of vector x will leave both Inequalities 5 and 6 untouched.

It remains to show that index f_d satisfies axiom E[s]. Let the $(n+1)$ -dimensional vector y result from x as described in the statement of axiom E[s]. Then y contains the k components $x_1+s(k+1)-s(k),\ldots,x_k+s(k+1)-s(k)$ and a newly added component of value d, where $s(k+1)\leq d\leq s(k+2)$; see Observation 2. By using Inequality 1, we get that

$$\sum_{i=1}^{k+1}y_i\geq d+\sum_{i=1}^{k}(x_i+s(k+1)-s(k))\geq(k+1)\cdot s(k+1)-1. \tag{7}$$

Furthermore, it can be seen that $x_{k+2}<s(k+1)$. We conclude that the $k+2$ largest components in y are the $k+1$ values $x_1+s(k+1)-s(k),\ldots,x_{k+1}+s(k+1)-s(k)$ and some component of value at most $2s(k+1)-s(k)\leq s(k+2)$. We derive

$$\begin{aligned}\sum_{i=1}^{k+2}y_i&\leq\sum_{i=1}^{k+1}(x_i+s(k+1)-s(k))+s(k+2)\\&<(k+2)\cdot s(k+2)-1.\end{aligned} \tag{8}$$

The two displayed Inequalities 7 and 8 together imply $f_d(y)=k+1$, which establishes axiom E[s]. □

This completes the proof of Theorem 2.

DOI: 10.1002/asi

Acknowledgments

This research has been supported 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

Arrow, K.J. (1951). Social choice and individual values. New York: Wiley.

Ball, P. (2005). Index aims for fair ranking of scientists. Nature, 436, 900.

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, 1275–1278.

Deineko, V.G., & Woeginger, G.J. (in press). A new family of scientific impact measures: The generalized Kosmulski-indices. Scientometrics.

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, 315–321.

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

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., & Rousseau, R. (2007). Hirsch-type indices and library management: The case of Tongji University Library. In D. Torres-Salinas & H.F. Moed (Eds.), Proceedings of the 11th International Conference of the International Society for Scientometrics and Informetrics (ISSI 2007), 514–522. Madrid, Spain: Centre for Scientific Information and Documentation of the Spanish Research Council.

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, England: 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, 297–301.

Ruane, F., & Tol, R.S.J. (2008). Rational (successive) h-indices: An application to economics in the Republic of Ireland. Scientometrics, 75, 395–405.

Schreiber, M. (2007). The influence of self-citation corrections on Egghe's g-index. Scientometrics, 76, 187–200.

Schreiber, M. (2008). An empirical investigation of the g-index for 26 physicists in comparison with the h-index, the A-index, and the R-index. Journal of the American Society for Information Science and Technology, 59, 1513–1522.

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

van Eck, N.J., & Waltman, L. (2008). Generalizing the h- and g-indices. Journal of Informetrics, 2, 263–271.

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, 491–502.

Woeginger, G.J. (2008a). An axiomatic analysis of Egghe's g-index. Journal of Informetrics, 2, 364–368.

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

Wu, Q. (2008). The w-index: A significant improvement of the h-index. Retrieved February 10, 2009, from http://arxiv.org/ftp/arxiv/papers/0805/0805.4650.pdf

DOI: 10.1002/asi