Journal of Informetrics 8 (2014) 659–666

Contents lists available at ScienceDirect

Journal of Informetrics

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

Investigations on the step-based research indices of Chambers and Miller☆

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 5 March 2014
Received in revised form 27 May 2014
Accepted 3 June 2014

Keywords:
Scientific impact measure
Citation analysis
Axiomatization

ABSTRACT

In a recent paper, Chambers and Miller introduced two fundamental axioms for scientific research indices. We perform a detailed analysis of these two axioms, thereby providing clean combinatorial characterizations of the research indices that satisfy these axioms and of the so-called step-based indices. We single out the staircase indices as a particularly simple subfamily of the step-based indices, and we provide a simple axiomatic characterization for them.

1. Introduction

Hirsch (2005) introduced the *h*-index (or Hirsch-index) as a simple and efficient tool for quantifying the scientific productivity and the scientific impact of an individual researcher. The *h*-index is based on the researcher's articles and on the number of citations that they have received: "*A scientist has index h, if h of his n articles have at least h citations each, whereas the other n–h articles have at most h citations each.*" This *h*-index has become extremely popular over the years, and it has attracted considerable attention among scientometricians and information scientists. There is a huge body of literature on the *h*-index and its variations; for a survey of the area we refer the reader to Alonso, Cabreziro, Herrera-Viedma, and Herrera (2009), Egghe (2010), Norris and Oppenheim (2010), and Schreiber, Malesios, and Psarakis (2011).

One branch of the literature has turned to the axiomatic analysis of scientific impact indices. Every axiom captures a certain property of an impact index, for instance how the index should change if the underlying data is slightly perturbed. The objective is to describe an impact index in terms of a small number of simple axioms. This branch has been started by Marchant (2009a) and Woeginger (2008a) who provided axiomatic characterizations for the ranking of scientists resulting from the *h*-index and for the *h*-index itself, respectively. For other representative results in this research branch, we point the reader to Bouyssou and Marchant (2010, 2014), Deineko and Woeginger (2009), Marchant (2009b), Miroiu (2013), Quesada (2011), and Woeginger (2008b, 2008c).

In a recent paper, Chambers and Miller (2014) introduced two fascinating new axioms that are simple and appealing, and that have a nice mathematical motivation. They also may be interpreted algebraically by means of certain homomorphisms between certain semi-groups. Most importantly, these axioms are satisfied by the *h*-index and also by some other natural indices. They proved that these two axioms characterize a family of scientific research indices that they called *step-based*

☆ Gerhard Woeginger acknowledges support by DIAMANT (a mathematics cluster of the Netherlands Organization for Scientific Research NWO), and by the Alexander von Humboldt Foundation, Bonn, Germany.

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

http://dx.doi.org/10.1016/j.joi.2014.06.001
1751-1577/

indices and that contain the h-index, the maximum-index, the i10-index and the publication count as special cases; see Section 2 for the definitions of these indices.

In this paper, we perform a deeper and more detailed analysis of the Chambers–Miller axioms. We combinatorially characterize the indices that satisfy the first axiom, we characterize the indices that satisfy the second axiom, and we characterize the indices that satisfy both axioms together. The two axioms are kind of symmetric to each other, and this symmetry is clearly reflected in our characterizations. As a side-result, we also produce a new proof for the main result of Chambers and Miller (2014). In the second part of the paper, we investigate the so-called *staircase indices* which on the one hand form a natural generalization of the h-index and on the other hand form a neat sub-class of the step-based indices. We provide an axiomatic characterization of the staircase indices in terms of four axioms; our axioms are a generalization of the axioms in the characterization of the h-index in Woeginger (2008a).

The paper is organized as follows. Section 2 states the basic definitions on research indices, and Section 3 summarizes the axioms of Chambers and Miller (2014). Section 4 presents our three combinatorial characterizations for the indices satisfying the Chambers–Miller axioms. Section 5 introduces step-based indices and staircase indices, and Section 6 finally provides the axiomatic characterization of the staircase indices.

2. Definitions and preliminaries

We denote by $\mathbb{N}$ the set of non-negative integers. The symbol ∞ (infinity) is used to denote an abstract upper bound on $\mathbb{N}$ that is larger than all the elements of $\mathbb{N}$. Next, let us introduce two sets of integer sequences that will be crucial throughout the paper.

- The set X^∞ consists of all infinite, non-increasing sequences with elements from $\mathbb{N} \cup \{\infty\}$.
- The set X contains all infinite, non-increasing sequences with elements from $\mathbb{N}$ that contain only finitely many non-zero elements.

Note that X is a subset of X^∞. Note furthermore that by stripping off the infinitely long tail of zeros from some sequence in X, we produce a positive integer vector of finite length. We will sometimes specify sequences in X by listing only some finite initial piece of the sequence, with the understanding that all the non-listed elements are zeros.

For two sequences $x = (x_k)_{k \ge 1}$ and $y = (y_k)_{k \ge 1}$ in X^∞, we say that x is *dominated* by y, if $x_k \le y_k$ holds for all $k \ge 1$; we write $x \preceq y$ to denote this situation. Clearly, this dominance relation is reflexive, anti-symmetric and transitive, and hence yields a partial order on X^∞. This dominance relation $\preceq$induces a dominance relation on the set X in the natural way. A *chain* is a sequence of sequences in X^∞ that is totally ordered with respect to the dominance relation; every sequence in such a chain dominates all its predecessor sequences and is dominated by all its successor sequences.

Next, consider a researcher who has published $n \ge 0$ publications that have received a positive number of citations. (We will ignore publications without citations, since they have not generated any impact.) If $n \ge 1$, this researcher may be represented by a vector $x = (x_1, \ldots, x_n)$ with n positive integer components $x_1 \ge x_2 \ge \cdots \ge x_n$ that are arranged in non-increasing order; in other words, the kth component x_k of this vector states the total number of citations to this researcher's kth-most cited publication. Equivalently, this researcher is represented by the sequence in X that results from x by appending an infinitely long tail of zeros. In the special case $n = 0$, the researcher has not received a single citation to his publications; the corresponding vector x is empty and the corresponding sequence in X entirely consists of zeros.

Definition 2.1. (Woeginger, 2008a) A scientific impact index is a function $f : X \to \mathbb{N}$ that satisfies the following conditions:

- If x entirely consists of zeros, then $f(x) = 0$.
- Monotonicity: If $x \preceq y$, then $f(x) \le f(y)$.

These conditions seem to be fundamental: A researcher without citations apparently has no impact. 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 list summarizes some of the standard research indices that we are going to use throughout for illustrating purposes.

- The Hirsch index assigns to $x = (x_k)_{k \ge 1}$ the value $h(x) := \max\{k : x_k \ge k\}$.
- The maximum-index assigns the value $f_{\max}(x) := x_1$.
- The publication count assigns the number $\mathrm{pc}(x) := \max\{k : x_k \ge 1\}$ of publications that have been cited at least once.
- The i10-index assigns the number $i10(x) := \max\{k : x_k \ge 10\}$ of publications that have been cited at least ten times.
- The total citation number index assigns the number $f_{\mathrm{tot}}(x) := \sum_{k \ge 1} x_k$ of citations.
- The Egghe-index assigns the value $g(x) := \max\{k : \sum_{i=1}^{k} x_i \ge k^2\}$.
- The Kosmulski-index assigns the value $h^{(2)}(x) := \max\{k : x_k \ge k^2\}$.
- The w-index assigns $w(x) := \max\{k : x_i \ge k - i + 1 \quad \text{forall} \quad i \le k\}$.