Journal of Informetrics 8 (2014) 449–477

Contents lists available at ScienceDirect

Journal of Informetrics

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

An axiomatic approach to bibliometric rankings and indices☆

Denis Bouyssou[a,∗,1], Thierry Marchant[b,1]

[a] CNRS (LAMSADE, UMR 7243) & Université Paris Dauphine, Place du Maréchal de Lattre de Tassigny, F-75775 Paris Cedex 16, France
[b] Ghent University, Department of Data Analysis, H. Dunantlaan, 1, B-9000 Gent, Belgium

ARTICLE INFO

Article history:
Received 11 February 2014
Accepted 18 March 2014
Available online 18 April 2014

Keywords:
Bibliometric ranking
Bibliometric index
h-Index
g-Index
Axiomatization

ABSTRACT

This paper analyzes several well-known bibliometric indices using an axiomatic approach. We concentrate on indices aiming at capturing the global impact of a scientific output and do not investigate indices aiming at capturing an average impact. Hence, the indices that we study are designed to evaluate authors or groups of authors but not journals. The bibliometric indices that are studied include classic ones such as the number of highly cited papers as well as more recent ones such as the *h*-index and the *g*-index. We give conditions that characterize these indices, up to the multiplication by a positive constant. We also study the bibliometric rankings that are induced by these indices. Hence, we provide a general framework for the comparison of bibliometric rankings and indices.

1. Introduction

This paper studies, from an axiomatic point of view, several rankings and indices based on bibliometric data. It is a companion paper to Marchant (2009a). Compared to this paper, we have enlarged the list of rankings studied. It now includes the rankings based on: the number of highly cited papers, the number of papers, the number of citations received, the number of citations received by highly cited papers, the number of citations exceeding a threshold, the maximum number of citations, the *h*-index, and the *g*-index. Moreover, we will study both rankings *and* indices, while Marchant (2009a) only studied rankings. By studying several bibliometric rankings and indices simultaneously, we hope to provide a framework for understanding their similarities and their differences. Hence, we have tried hard to use conditions that can easily be interpreted and to make maximal use of conditions that are common to several rankings or indices.

We concentrate on indices aiming at capturing the global impact of a scientific output and do not investigate indices aiming at capturing an average impact. Hence, the indices that we study are designed to evaluate authors or groups of author x in terms of citations but not journals.

Among the huge literature (for review, see Alonso, Cabreziro, Herrera-Viedma, & Herrera, 2009; Egghe, 2010a; Norris & Oppenheim, 2010; Ruscio, Seaman, D'Oriano, Stremlo, & Mahalchik, 2012; Schreiber, Malesios, & Psarakis, 2011) on the *h*-index and its variants (Rousseau, García-Zoritad, & Sanz-Casadod, 2013, have seen this development as a "bubble"), there is already as sizeable literature on the axiomatic analysis of the *h*-index, most notably Woeginger (2008a, 2008b), Deineko

☆ We thank Antonio Quesada and Ludo Waltman for their comments on an earlier draft of this text. The usual caveat applies.
∗ Corresponding author. Tel.: +33 144054898; fax: +33 144054091.
E-mail addresses: bouyssou@lamsade.dauphine.fr (D. Bouyssou), thierry.marchant@UGent.be (T. Marchant).
[1] Both authors contributed equally to this paper.

http://dx.doi.org/10.1016/j.joi.2014.03.001
1751-1577/

and Woeginger (2009), Quesada (2009, 2010, 2011a, 2011b), Hwang (2013) and Miroiu (2013).[2] The axiomatic literature on the g-index is less abundant but nevertheless exists: Woeginger (2008c, 2009) and Quesada (2011a). Hence, the reader might wonder why we want to add to this literature. We do so because we feel that these previous axiomatizations have limitations. Let us discuss them using the example of the h-index. Similar remarks apply to the previous axiomatizations of the g-index.

The above papers on the axiomatization of the h-index use axioms that we do not find easy to interpret. Let us take the example of Theorem 4.1 in Woeginger (2008a). This result characterizes the h-index using three axioms called A1, B and D. Axiom A1 requires that a bibliometric index f should satisfy the following condition. Consider an author x. If the author y is identical to author x, except that she has published an additional paper with $f(x)$ citations, then $f(y) = f(x)$. Although this axiom is mathematically fine, we claim that this condition is quite hard to interpret. Indeed, an axiom is a condition imposed on an index f, where f is *any* index, not necessarily the h-index. Hence, when we read the above condition, we cannot suppose that f is the h-index. For instance, it could be the square of the number of papers or the logarithm of the total number of citations. Hence, we think that it makes little sense to say that "if we add a new paper with $f(x)$ citations, then . . .". Why would we find such a condition (normatively) appealing if we do not know what $f(x)$ represents? Let us develop this point further since it is important to motivate the present paper.

If $f(x)$ measures the scientific value of author x in terms citations per paper, then it is certainly right to say that adding to x a paper with $f(x)$ citations does not alter the index. But suppose now the president of your university asks you to help him choose an index. You ask him whether he wants a citation-monotonic index (more citations are always better). He probably says yes. You ask him whether he wants a paper-monotonic index (more papers are always better). He may say yes or no. You then ask him whether he wants that adding to x a paper with $f(x)$ citations does not alter the index. He will certainly answer "I don't know. It depends on what $f(x)$ represents". If you reply "it is the average number of citations per paper", then he will perhaps say yes. But, if the president is smart, he will ask you "Why do you question me about these axioms if you already decided that you want to use the average number of citations per paper?" Put differently, what is the point of axiomatically characterizing an index if the axioms cannot be understood without presuming what the index is?

Still in other words, in order to know whether an additional paper can increase the index of a scientist x, Woeginger's Axiom A1 compares $f(x)$ to the number of citations of the additional paper. If the number of citations of the new paper is not larger than $f(x)$, than the index cannot increase. This comparison is not really valid, because it compares pears and apples. We do not know whether $f(x)$ is a number of citations, a number of papers, the product of a number of citations and a number of papers, a squared number of citations or anything else. It does not make sense to compare them.

Axiom D in Woeginger (2008a) has the same problem. Finally, Woeginger (2008a) assumes that a bibliometric index must be a non-negative integer. Requiring that an index can only take integer values appears very restrictive and difficult to motivate.

A second limitation of the above mentioned axiomatizations of the h-index is that they concentrate on characterizing the index and do not study the ranking induced by the index. In many situations, a bibliometric index is only used via the ranking it induces on authors. Hence, we feel that this question should be studied too. Admittedly, there are situations in which one might want to use the index and not only the ranking it induces. This is the case, for instance, when research funds are allocated using a formula involving an index. Nevertheless, as stressed in Franceschini and Maisano (2010), the scale on which the h-index is measured is rather complex to analyze. Indeed, the effort involved for raising the h-index by 1 unit seems much smaller starting at 1 than starting at 100. A similar remark holds if one wishes to double the h-index.

The last limitation is that all papers proposing an axiomatic analysis of the h-index characterize only one index. We do not think that this is completely satisfactory. Indeed, consider an index h' defined as 100 times the h-index. Is it worse or better than the h-index? This question seems irrelevant, just like asking whether measuring distances in meters is "better" than measuring them in centimeters. Our axiomatic analysis will single out not a single index but a family of indices deduced from one another by the multiplication by a positive constant. One may object to this last argument that the fact that an author has an h-index equal to k has an immediate interpretation: the author has k papers having collected at least k citations. Such an interpretation does not hold if we multiply the h-index by a positive constant. However, because the scale on which the h-index is measured is difficult to ascertain (Franceschini & Maisano, 2010), we think that this ease of interpretation is largely illusory and can be misleading. Moreover, an advantage of focusing on the family of indices deduced from one another by the multiplication by a positive constant is that it allows a unified treatment of both rankings and indices. This will also allow us to use weaker axioms than the ones aiming at characterizing a unique index.

We are aware that the axiomatic analysis of the type proposed in this paper only considers one aspect of bibliometric rankings and indices. Our analysis does not appeal to an explicit model of production of publications and citations (for an introduction to such models, see Egghe, 2005). Similarly, we do not propose any empirical study that would allow to analyze the links between these rankings and indices on a given set of authors (for such analyses, see, e.g., Bornmann & Leydesdorff, 2013; Bornmann, Mutz, & Daniel, 2008; Bornmann, Mutz, Hug, & Daniel, 2011; van & Raan, 2006). Hence, we acknowledge

[2] While preparing the last draft of this text, we became aware of the work of Chambers and Miller (2014) who characterize a family of bibliometric rankings, called *step-based indices*, that includes as particular cases the number of highly cited papers, the number of papers, and the h-index. It does not provide a characterization of these particular cases.

the limitations of the axiomatic approach to bibliometrics put forward by Glänzel and Moed (2013). Nevertheless, we think such an axiomatic analysis can give useful elements about the similarities and differences between rankings or indices.

The paper is organized as follows. We introduce our setting and notation in Section 2. We then study the ranking and index based on the number of highly cited papers (Section 3), the number of papers (Section 4), the number of citations of highly cited papers (Section 5), the number of citations exceeding a threshold (Section 6), the number of citations (Section 7), the maximum number of citations (Section 8), the h-index (Section 9), and the g-index (Section 10). Section 11 contains additional remarks. A final section summarizes our findings and concludes.

2. Notation and setting

The set of non-negative integers is denoted by $\mathbb{N}$. We define $\mathbb{N}_+ = \{x \in \mathbb{N} : x > 0\}$. The set of real numbers is denoted by $\mathbb{R}$. We define $\mathbb{R}_0 = \{x \in \mathbb{R} : x \geq 0\}$ and $\mathbb{R}_+ = \{x \in \mathbb{R} : x > 0\}$

2.1. Setting

We use the general framework introduced in Marchant (2009a). We represent an author a as a mapping from $\mathbb{N}$ to $\mathbb{N}$. For $x \in \mathbb{N}$, we interpret $a(x)$ as the number of publications of author a with exactly x citations.

The number of publications of author a is given by $p_a = \sum_{x \in \mathbb{N}} a(x)$. The number of publications of author a having been cited at least once is given by $p'_a = \sum_{x \in \mathbb{N}_+} a(x)$. The total number of citations received by the publications of author a is given by $c_a = \sum_{x \in \mathbb{N}} x a(x)$.

The set of authors A is the set of all functions a from $\mathbb{N}$ to $\mathbb{N}$ such that p_a is finite.

Since authors in A are modelled as functions, it makes sense to speak of an author that is the addition of several authors or the multiplication of an author by a non-negative integer. Hence, if $a, b \in A$ and $n \in \mathbb{N}$, $[a+b] \in A$ and $[na] \in A$. We have, for all $x \in \mathbb{N}$, $[a+b](x) = a(x) + b(x)$ and $[na](x) = n \times a(x)$. When there is no risk of confusion, we omit the brackets around $[a+b]$ and $[na]$.

For all $x \in \mathbb{N}$, we denote by $\mathbf{1}_x$ the author in A with exactly one publication having x citations. We denote by $\mathbf{0}$ an author (the *null* author) without any publication. We say that an author a is *non-null* if she is not $\mathbf{0}$. We say that an author is *quasi-null* if she has only uncited papers, i.e., is equal to $x\mathbf{1}_0$, for some $x \in \mathbb{N}_+$. An author that is neither null nor quasi-null is said to be *strictly non-null*.

By construction, an author $a \in A$ can always be written as the sum of authors having a single publication:

$$a = \sum_{z \in \mathbb{N}} a(z)\mathbf{1}_z.$$

For a null author, all terms $a(z)$ are null. For quasi-null authors, all terms $a(z)$ are null, except $a(0)$. A strictly non-null author has at least one $a(z) > 0$, for some $z \in \mathbb{N}_+$.

Let $a \in A$. For $x \in \mathbb{N}$, let $a^+(x)$ be the number of papers published by a having received at least x citations, i.e., $a^+(x) = \sum_{i \geq x} a(i)$.

For $a, b \in A$, we write $a \trianglerighteq b$ if $a^+(x) \geq b^+(x)$, for all $x \in \mathbb{N}$ and we say that *a dominates b*. It is easy to check, since $a^+(0) = p_a$, that $a \trianglerighteq b$ implies $p_a \geq p_b$. The relation $\trianglerighteq$ is exactly equivalent to the dominance relation used by Woeginger (2008a, p. 225) to define his monotonicity condition (he uses the notation $\succeq$).

Remark 1. It is easy to check that if $a \trianglerighteq b$, then it possible to go from b to a using the following operations. We first add to b a number of uncited papers equal to $p_a - p_b$. This ensures that we have $a^+(0) = b^+(0)$. We then add a number of citations to the papers published by the modified author b to ensure that $a^+(x) = b^+(x)$, for all $x \in \mathbb{N}$.

The relation $\trianglerighteq$ is not identical to the relation $\sqsupseteq$ used in Marchant (2008a, p. 327). We have $a \sqsupseteq b$ when $a(x) \geq b(x)$, for all $x \in \mathbb{N}$. Clearly, if $a \sqsupseteq b$ holds, then $a \trianglerighteq b$ also holds. The converse is not true. Indeed, let the author b be identical to a except that b has received an additional citation. Between b and a, the relation $\sqsupseteq$ does not hold while $\trianglerighteq$ clearly does.

Any sensible bibliometric ranking (or index) should be minimally compatible with the above relation.[3] If a dominates b, we will require that a is not ranked below b (or has a lower value of the index). Moreover, the trivial ranking of authors declaring all authors in A to be equivalent has clearly little interest. These two requirements lead to the definitions of what we will call a bibliometric ranking and a bibliometric index.

2.2. Definitions

A ranking of authors is a complete and transitive binary relation on the set of authors A. When $\succsim$ is a ranking, the statement $a \succsim b$ is interpreted as meaning that author a has a performance that is at least as good as the performance of author b. The asymmetric part of $\succsim$ will be denoted by $\succ$ and is interpreted as "has a strictly better performance". The symmetric part of

[3] Chambers and Miller (2014) call the relation the "objective" information for comparing authors.

$\succsim$ will be denoted by $\sim$ and is interpreted as "has an equivalent performance". A similar convention will hold if subscripts or superscripts are added to the symbol $\succsim$.

Similarly, an index is a real-valued function f on the set of authors A.

In order to exclude obviously uninteresting rankings from the analysis, we will suppose throughout that any ranking satisfies the following two conditions.

A1 (Nontriviality) *There are* $a, b \in A$ *such that* $a \nsim b$.

The above condition excludes the trivial ranking always putting all authors in the same equivalence class. It seems quite innocuous since the trivial ranking has little interest.

A2 (Monotonicity) *For all* $a, b \in A$, $a \trianglerighteq b \Rightarrow a \succsim b$.

Remark 2. Observe that the above two conditions can easily be reformulated for an index f. The rules of this reformulation are quite simple, whenever the condition on rankings only uses $\succsim$ or some derived relations. Instead of saying that $a \sim b$ (resp. $a \nsim b$, $a \succsim b$, and $a \succ b$), we now say that $f(a) = f(b)$ (resp. $f(a) \neq f(b)$, $f(a) \geq f(b)$, and $f(a) > f(b)$). Hence, any condition on a ranking can be translated into a condition on an index. For instance, condition **A2**, stated for an index f, says that if $a \trianglerighteq b$, then we should have $f(a) \geq f(b)$.

In the sequel, we will use the same name for the condition on a ranking and the corresponding condition on an index. This clear abuse of notation should nevertheless not cause confusion.

The monotonicity condition says that publishing an additional paper and/or receiving an additional citation should not lower the position of an author. It is identical to the monotonicity condition used by Woeginger (2008a, p. 225). Although we find this condition quite compelling, it is not completely innocuous.[4] Suppose that we want to compare two authors a and b. They are identical in all respects, both having published 10 papers each of them having received 20 citations, except that a has published one more paper than b that has not received any citation. One may view the uncited paper of a as a signal that part of her research has not attracted any attention. Hence, one may want to penalize a for this additional uncited paper. Our monotonicity condition forbids this and implies that a should not be ranked lower than b. The monotonicity condition is tailored here for "total performance indices" and not for "average performance indices" in the sense of Waltman and van Eck (2009b) and Waltman, van Eck, van Leeuwen, Visser, and van Raan (2011). In other terms, one should interpret the bibliometric indices studied in this paper as "extensive" indices (see, e.g., Bouyssou and Marchant, 2010, p. 369), i.e., aiming at capturing the "global impact" of the scientific output of an author, contrary to "intensive" indices aiming at capturing the "average impact" of the scientific output of an author. We do not think that this is unreasonable when evaluating authors. Clearly, this would be much more objectionable for evaluating journals (see Bouyssou & Marchant, 2010, 2011a).

Additional remarks on these conditions **A1** and **A2** can be found in Section 11.

Definition 1. A *bibliometric ranking* $\succsim$ is a complete and transitive binary relation defined on the set of authors A that satisfies **A1** and **A2**.

Remark 3. The above definition differs from Marchant (2008, Def. 1, p. 327) in that it is supposed there that a bibliometric ranking always satisfies Nontriviality, a condition that is identical to **A1**, and a condition, weaker than Monotonicity, called CDNH (Citations Do Not Harm). The replacement of CDNH by Monotonicity is motivated by our wish to keep the number of conditions used to a minimum.

Definition 2. A *bibliometric index* f is a function associating with any author $a \in A$ a real number $f(a)$ satisfying **A1** and **A2**.

Clearly, any bibliometric index f induces a bibliometric ranking $\succsim_f$, letting, for all $a, b \in A$,

$$a \succsim_f b \Leftrightarrow f(a) \geq f(b).$$

Let φ be a strictly increasing function from $\mathbb{R}$ to $\mathbb{R}$. The bibliometric index $\varphi \circ f$ that associates with all $a \in A$ the value $\varphi(f(a))$ induces on A a bibliometric ranking that is identical to the one induced by f.

Remark 4. We could have defined bibliometric ranking and indices without requiring that they satisfy **A1** and **A2**. This would complicate the statement of many results and require many more examples to show that the conditions used in our results are independent. The choice made here is motivated by our desire to keep things simple and the fact that **A1** is quite innocuous while, as discussed above, **A2** is fairly reasonable for rankings and indices aiming at capturing the "global impact".

There are many possible bibliometric rankings (and, hence, indices). One possible way to compare and analyze them is to study whether or not they satisfy a number of easily interpretable properties (e.g., Rousseau, 2008, studies the properties satisfied by some variants of the h-index). Moreover, we may also try to find properties that are collectively satisfied by a unique ranking or index (or family of rankings or indices), i.e., characterizing properties. Table 1 gives a schematic view of the main conditions used below. These conditions will precisely be defined and discussed in the course of the text.

[4] We thank Antonio Quesada for having brought this point to our attention.

Table 1
Schematic view of conditions. The conditions labelled A and C apply to both rankings and indices. Conditions A will always be in force. Conditions ORI and ES are specific to indices.

Condition	Name	Expression
A1	Nontriviality	$\exists a, b \in A$ such that $a \not\sim b$
A2	Monotonicity	$a \trianglerighteq b \Rightarrow a \succsim b$
C1	Independence	$a \succsim b \Leftrightarrow a + c \succsim b + c$
C2	2-Gradedness modified	$[\mathbf{1}_y \succ \mathbf{0}$ and $x > y] \Rightarrow \mathbf{1}_x \sim \mathbf{1}_y$
C3	One is One	$\mathbf{1}_x \sim \mathbf{1}_y$
C4	Restricted Additivity	$\mathbf{1}_x \succ \mathbf{0}$ and $\mathbf{1}_y \succ \mathbf{0} \Rightarrow [\mathbf{1}_x + \mathbf{1}_y] \sim \mathbf{1}_{x+y}$
C5	Restricted Transfer	$\mathbf{1}_x \succ \mathbf{0}$ and $\mathbf{1}_y \succ \mathbf{0} \Rightarrow [\mathbf{1}_x + \mathbf{1}_y] \sim \mathbf{1}_{x+1} + \mathbf{1}_{y-1}$
C6	Additivity	$\mathbf{1}_x + \mathbf{1}_1 \sim \mathbf{1}_{x+1}$
C7	Weak Independence	$a \succsim b \Rightarrow a + c \succsim b + c$
C8	Strict Monotonicity	$x > y \Leftrightarrow \mathbf{1}_x \succ \mathbf{1}_y$
C9	Quasi-null authors	$\mathbf{1}_0 \sim \mathbf{0}$
C10	One Plus One Equals One	$\mathbf{1}_x + \mathbf{1}_x \sim \mathbf{1}_x$
C11	Strong quasi-null authors	$x\mathbf{1}_0 \sim \mathbf{0}$
C12	Tail independence	See text
C13	Square upwards	$y_i \geq x \Rightarrow x\mathbf{1}_x \succsim [\mathbf{1}_{y_1} + \mathbf{1}_{y_2} + \ldots \mathbf{1}_{y_x}]$
C14	Square rightwards	$y_i \leq x \Rightarrow x\mathbf{1}_x \succsim [x\mathbf{1}_x + \mathbf{1}_{y_1} + \mathbf{1}_{y_2} + \ldots \mathbf{1}_{y_j}]$
C15	Strong uniformity	$(x+1)\mathbf{1}_{x+1} \succ x\mathbf{1}_x \Leftrightarrow (y+1)\mathbf{1}_{y+1} \succ y\mathbf{1}_y$
C16	Lorenz monotonicity	$a \trianglerighteq^L b \Rightarrow a \succsim b$
C17	Single paper author	$x\mathbf{1}_x \succsim \mathbf{1}_{(x+1)^2-1}$
C18	Modified tail independence	See text
ORI	Condition Origin	$f(\mathbf{0}) = 0$
ES	Equal Spacing	See text

3. Number of highly cited papers

This is a classical bibliometric index (see, e.g., Chapron & Husté, 2006; van Eck & Waltman, 2008).

3.1. Setting

Let $\tau \in \mathbb{N}$. We are interested in the index f_τ associating with each author $a \in A$ the number of her papers that have received at least τ citations, i.e., we have:

$$f_\tau(a) = \sum_{x \geq \tau} a(x).$$

We will also be interested in the bibliometric ranking $\succsim_\tau$ induced by f_τ. We refer to van Eck and Waltman (2008) for a very insightful analysis of this index.

When $\tau = 0$, we obtain the index consisting in the number of papers. Higher values of τ may be justified by the fact that only papers with a sufficiently high number of citations should influence the performance of an author.

3.2. Conditions

Our first condition is independence. As discussed in Bouyssou and Marchant (2011b), Marchant (2009a, 2009b), Waltman and van Eck (2009b, 2012) this is an important property of some bibliometric rankings and indices.

C1 (Independence) *For all* $a, b, c \in A$, $a \succsim b \Leftrightarrow a + c \succsim b + c$.

This above condition is identical to condition A4 in Marchant (2008, p. 328). It is easy to check that, whatever $\tau \in \mathbb{N}$, it is satisfied by f_τ. The same is true for $\succsim_\tau$. Indeed, whenever a bibliometric index satisfies a condition expressed using a relation $\succsim$ (such conditions are labeled C in this paper) the corresponding bibliometric ranking clearly satisfies the same condition.

Independence says the following. Consider two authors $a, b \in A$ such that a is at least as good as b. Suppose that both a and b publish the same number of additional papers and that each of these additional papers receive the same number of citations. After the publication of these new papers, a becomes $a + c$ and b becomes $b + c$. Independence requires that $a + c$ is at least as good as $b + c$. This seems a sensible condition. We will nevertheless see that there are bibliometric rankings that violate it. In order to understand why this property might not be sensible, suppose that both a and b have published a small number of papers. Suppose that we consider that a is strictly better than b and that c contains a very large number of papers. Adding c to both a and b can dilute the advantage for a over b and we might want to conclude that $a + c$ and $b + c$ are equivalent. This also serves to motivate the weakening of this condition considered below in Section 8.2 (**C7**, Weak Independence).

For a detailed study of independence, we refer to Bouyssou and Marchant (2010), Marchant (2009b) and Waltman and van Eck (2012). Independence is the central condition used to characterize scoring rules in Marchant (2009b).

Remark 5. The independence condition can be viewed as the conjunction of the following two conditions stating that, for all $a, b, c \in A$,

$$ab \Rightarrow a+c \succsim b+c, \tag{1}$$

$$a \succ b \Rightarrow a+c \succ b+c. \tag{2}$$

Eq. (1) is the weakening of independence later defined as **C7** (Weak Independence). Since $\succsim$ is complete, (2) is equivalent to saying that $a+c \succsim b+c \Rightarrow a \succsim b$.

Remark 6. If a ranking $\succsim$ satisfies **C1**, it is easy to see that $a \succsim b$ and $c \sim d$ imply $a+c \succsim b+d$ (indeed, **C1** and $a \succsim b$ imply $a+c \succsim b+c$. Similarly, **C1** and $c \sim d$ imply $c+b \sim d+b$. Hence, we have $a+c \succsim b+d$). Similarly, $a \succ b$ and $c \sim d$ imply $a+c \succ b+d$.

Another useful consequence of **C1** is that $a \succsim b \Leftrightarrow na \succsim nb$ with $n \in \mathbb{N}_+$. Indeed, **C1** and $a \succsim b$ imply $a+a \succsim b+a$. Similarly, **C1** and $a \succsim b$ imply $a+b \succsim b+b$. Hence, we obtain $2a \succsim 2b$. Similarly, $a \succ b$ implies $2a \succ 2b$. Iterating the above reasoning leads to the desired conclusion.

We will often use such easy consequences of **C1**, without detailing them in the sequel.

The following lemma will be useful.

Lemma 1. *If a bibliometric ranking $\succsim$ satisfies **C1**, then we have $\mathbf{1}_x \succ \mathbf{0}$, for some $x \in \mathbb{N}$.*

Proof. Suppose that the thesis is violated, so that $\mathbf{0} \succsim \mathbf{1}_x$, for all $x \in \mathbb{N}$. Using **A2**, we must have $\mathbf{1}_x \sim \mathbf{0}$, for all $x \in \mathbb{N}$.

Any non-null $a \in A$ can be written as the sum of single paper authors. Since $\mathbf{1}_x \sim \mathbf{0}$, for all $x \in \mathbb{N}$, it is easy to see that repeated applications of **C1** imply that $a \sim \mathbf{0}$. This contradicts **A1**. □

The following is the (modified) 2-gradedness condition. It says that all authors having published a single paper are equivalent as soon as they are strictly better than the null author. This condition is strong since it amounts to be skeptical about the idea that citations are a sign of quality, Whatever $\tau \in \mathbb{N}$, this condition is satisfied by $\succsim_\tau$. Indeed, for this index we have $f_\tau(\mathbf{1}_x) = 0$, for all $x \in \mathbb{N}$ such that $x < \tau$, and $f_\tau(\mathbf{1}_x) = 1$, for all $x \in \mathbb{N}$ such that $x \geq \tau$. This condition is close but not identical to the 2-gradedness condition used in Marchant (2008a, p. 333), which explains its name.

C2 (2-Gradedness modified) *For all* $x, y \in \mathbb{N}$, $[\mathbf{1}_y \succ \mathbf{0}$ *and* $x > y] \Rightarrow \mathbf{1}_x \sim \mathbf{1}_y$.

The following lemmas will be useful.

Lemma 2. *If a bibliometric ranking $\succsim$ satisfies **C1** and **C2**, then there is $\tau \in \mathbb{N}$ such that, for all $x \in \mathbb{N}$, $x < \tau \Rightarrow \mathbf{1}_x \sim \mathbf{0}$ and $x \geq \tau \Rightarrow \mathbf{1}_x \sim \mathbf{1}_\tau \succ \mathbf{0}$.*

Proof. We know from **A2** that $\mathbf{1}_x \succsim \mathbf{0}$, for all $x \in \mathbb{N}$. Moreover, Lemma 1 implies that $\mathbf{1}_x \succ \mathbf{0}$, for some $x \in \mathbb{N}$. Define $\tau \in \mathbb{N}$ to be the smallest $x \in \mathbb{N}$ such that $\mathbf{1}_x \succ \mathbf{0}$. By construction, for all $x < \tau$, we have $\mathbf{1}_x \sim \mathbf{0}$. Using **C2**, we have $\mathbf{1}_x \sim \mathbf{1}_\tau \succ \mathbf{0}$, for all $x \geq \tau$.□

Lemma 3. *Let $\succsim$ be a bibliometric ranking satisfying **C1** and **C2**. Define τ as in Lemma 2. Let $a \in A$ be a non-null author. We have $a \sim a^+(\tau)\mathbf{1}_\tau$, where, as before, $a^+(\tau) = \sum_{i \geq \tau} a(i)$.*

Proof. We know that $x \geq \tau$ implies $\mathbf{1}_x \sim \mathbf{1}_\tau$ and that $x < \tau$ implies $\mathbf{1}_x \sim \mathbf{0}$. The thesis follows from repeated applications of **C1**. Since this is our first nontrivial proof, let us give details. Indeed, suppose that $a \in A$ is such that

$$a = \mathbf{1}_{x_1} + \mathbf{1}_{x_2} + \cdots + \mathbf{1}_{x_n} + \mathbf{1}_{y_1} + \mathbf{1}_{y_2} + \cdots + \mathbf{1}_{y_m},$$

with $x_1, x_2, \ldots, x_n \geq \tau$ and $y_1, y_2, \ldots, y_m < \tau$.

We have $\mathbf{1}_{y_i} \sim \mathbf{0}$, for $i = 1, 2, \ldots, m$ and $\mathbf{1}_{x_j} \sim \mathbf{1}_\tau$, for $j = 1, 2, \ldots, n$.

Since $\mathbf{1}_{y_m} \sim \mathbf{0}$, **C1** implies that

$$\mathbf{1}_{y_m} + \mathbf{1}_{x_1} + \mathbf{1}_{x_2} + \cdots + \mathbf{1}_{x_n} + \mathbf{1}_{y_1} + \mathbf{1}_{y_2} + \cdots + \mathbf{1}_{y_{m-1}} \sim \\ \mathbf{0} + \mathbf{1}_{x_1} + \mathbf{1}_{x_2} + \cdots + \mathbf{1}_{x_n} + \mathbf{1}_{y_1} + \mathbf{1}_{y_2} + \cdots + \mathbf{1}_{y_{m-1}}.$$

Hence, we obtain

$$\mathbf{1}_{x_1} + \mathbf{1}_{x_2} + \cdots + \mathbf{1}_{x_n} + \mathbf{1}_{y_1} + \mathbf{1}_{y_2} + \cdots + \mathbf{1}_{y_{m-1}} + \mathbf{1}_{y_m} \sim \\ \mathbf{1}_{x_1} + \mathbf{1}_{x_2} + \cdots + \mathbf{1}_{x_n} + \mathbf{1}_{y_1} + \mathbf{1}_{y_2} + \cdots + \mathbf{1}_{y_{m-1}}.$$

Since $\mathbf{1}_{y_{m-1}} \sim \mathbf{0}$, **C1** implies that

$$\mathbf{1}_{y_{m-1}} + \mathbf{1}_{x_1} + \mathbf{1}_{x_2} + \cdots + \mathbf{1}_{x_n} + \mathbf{1}_{y_1} + \mathbf{1}_{y_2} + \cdots + \mathbf{1}_{y_{m-2}} \sim \\ \mathbf{0} + \mathbf{1}_{x_1} + \mathbf{1}_{x_2} + \cdots + \mathbf{1}_{x_n} + \mathbf{1}_{y_1} + \mathbf{1}_{y_2} + \cdots + \mathbf{1}_{y_{m-2}}.$$

Hence, we obtain

$$\mathbf{1}_{x_1} + \mathbf{1}_{x_2} + \cdots + \mathbf{1}_{x_n} + \mathbf{1}_{y_1} + \mathbf{1}_{y_2} + \cdots + \mathbf{1}_{y_{m-2}} + \mathbf{1}_{y_{m-1}} + \mathbf{1}_{y_m} \sim \\ \mathbf{1}_{x_1} + \mathbf{1}_{x_2} + \cdots + \mathbf{1}_{x_n} + \mathbf{1}_{y_1} + \mathbf{1}_{y_2} + \cdots + \mathbf{1}_{y_{m-2}} + \mathbf{1}_{y_{m-1}} \sim \\ \mathbf{1}_{x_1} + \mathbf{1}_{x_2} + \cdots + \mathbf{1}_{x_n} + \mathbf{1}_{y_1} + \mathbf{1}_{y_2} + \cdots + \mathbf{1}_{y_{m-2}}$$

Repeating the same reasoning leads to conclude that

$$\mathbf{1}_{x_1} + \mathbf{1}_{x_2} + \cdots + \mathbf{1}_{x_n} + \mathbf{1}_{y_1} + \mathbf{1}_{y_2} + \cdots + \mathbf{1}_{y_m} \sim \mathbf{1}_{x_1} + \mathbf{1}_{x_2} + \cdots + \mathbf{1}_{x_n}.$$

Since $\mathbf{1}_{x_n} \sim \mathbf{1}_\tau$, **C1** implies that

$$\mathbf{1}_{x_1} + \mathbf{1}_{x_2} + \cdots + \mathbf{1}_{x_{n-1}} + \mathbf{1}_{x_n} \sim \mathbf{1}_{x_1} + \mathbf{1}_{x_2} + \cdots + \mathbf{1}_{x_{n-1}} + \mathbf{1}_\tau.$$

Since $\mathbf{1}_{x_{n-1}} \sim \mathbf{1}_\tau$, **C1** implies that

$$\mathbf{1}_{x_1} + \mathbf{1}_{x_2} + \cdots + \mathbf{1}_{x_{n-2}} + \mathbf{1}_\tau + \mathbf{1}_{x_{n-1}} \sim \mathbf{1}_{x_1} + \mathbf{1}_{x_2} + \cdots + \mathbf{1}_{x_{n-2}} + \mathbf{1}_\tau + \mathbf{1}_\tau.$$

Hence, we obtain

$$\mathbf{1}_{x_1} + \mathbf{1}_{x_2} + \cdots + \mathbf{1}_{x_{n-1}} + \mathbf{1}_{x_n} \sim \mathbf{1}_{x_1} + \mathbf{1}_{x_2} + \cdots + \mathbf{1}_{x_{n-2}} + 2\mathbf{1}_\tau.$$

Repeating the same reasoning leads to conclude that

$$\mathbf{1}_{x_1} + \mathbf{1}_{x_2} + \cdots + \mathbf{1}_{x_{n-1}} + \mathbf{1}_{x_n} \sim a^+(\tau)\mathbf{1}_\tau.$$

□

All conditions introduced so far dealt with both rankings and indices. We now introduce conditions that are specific to indices.

In order to go from a ranking to an index, we have to specify the "origin" of the index. Using **A2**, we know that, for all $a \in A$, $a \succsim \mathbf{0}$. Our first condition for indices specifies the value of the index for the null author $\mathbf{0}$.

ORI (Condition Origin) $f(\mathbf{0}) = 0$.

Condition ORI is a very mild one. It is satisfied by f_τ. It will also hold for all indices studied in this paper. Together with **A2**, it implies that for all $a \in A$, $f(a) \geq 0$.

Once the origin of the index is settled, one also has to choose a "unit". We do so by specifying that some authors are equally spaced with respect to the index.

ES (Equal Spacing) *Let a, b, $c \in A$ be three authors such that $f(a) > f(b) > f(c)$. If, for all $d \in A \setminus \{b\}$, we have either $f(d) \geq f(a)$ or $f(d) \leq f(c)$, then $f(a) - f(b) = f(b) - f(c)$.*

Condition ES says that, if author b is "immediately above" c (meaning that there is no author having a value of the index strictly between $f(c)$ and $f(b)$) and a is "immediately above" b, then the differences $f(b) - f(c)$ and $f(a) - f(b)$ must be equal.[5] This implies that, going from an author to an author that is "immediately above" her, always raises the index by the same amount. It is not difficult to check that, whatever τ, the index f_τ satisfies ES.

Remark 7. The reader may have the impression that, starting with a bibliometric ranking, adding conditions ORI and ES always leads to define a bibliometric index. This is not so because requiring that a bibliometric index is real-valued imposes additional constraints. Let us illustrate this point with a simple example.

Consider the ranking that is obtained combining the total number of citations (i.e., the ranking $\succsim_c$ defined below) and the number of papers (i.e., the ranking $\succsim_0$ defined below) in a lexicographic way. For authors having a different total number of citations, the ranking is based on the number of citations. For authors having the same total number of citations, ties are broken according to the number of papers. Two authors are indifferent only when they have the same total number of citations and the same number of papers. It is simple to check that this ranking satisfies **A1** and **A2**, so that it is a bibliometric ranking. It leads to:

$$\mathbf{0} \prec 1\mathbf{1}_0 \prec 2\mathbf{1}_0 \prec 3\mathbf{1}_0 \prec \cdots \prec i\mathbf{1}_0 \prec \cdots$$

Moreover, it is simple to check that $1\mathbf{1}_0$ is immediately above $\mathbf{0}$ and, for all $i \in \mathbb{N}$, $(i+1)\mathbf{1}_0$ is immediately above $i\mathbf{1}_0$. Imposing conditions ORI and ES leads to conclude that, for the associated index, $f(i\mathbf{1}_0) = ik$, for all $i \in \mathbb{N}$, where $k = f(1\mathbf{1}_0) - f(\mathbf{0})$ is a strictly positive real number. But we also know that $\mathbf{1}_1 \succ i\mathbf{1}_0$, for all $i \in \mathbb{N}$. Hence, we should have $f(\mathbf{1}_1) > ik$, for all $i \in \mathbb{N}$. This is clearly impossible.

Let us finally observe that adding conditions ORI and ES to a bibliometric ranking leads to an index that is one among the many possible indices inducing the same bibliometric ranking. Consider, for instance, the index, studied below, that counts the number of citations of highly cited papers, i.e., given a threshold $\tau \in \mathbb{N}_+$, this index associates with an author $a \in A$ the number $f(a) = \sum_{x \geq \tau} xa(x)$. After having characterized the ranking induced by this index, we cannot simply add ORI and ES to characterize this index. This is because this index does not satisfy ES. Indeed, $\mathbf{1}_{\tau+1}$ is immediately above $\mathbf{1}_\tau$ that is, in turn,

[5] Strengthening this condition requiring that $f(b) - f(c) = f(a) - f(b) = 1$ would lead to characterize the index under study instead of the family of indices obtained from the index under study via the multiplication by a positive constant.

immediately above $\mathbf{0}$, whereas the value of the index counting the number of highly cited papers is respectively $\tau+1$, τ and 0.

3.3. Results

Our first result, in this section, characterizes $\succsim_\tau$.

Theorem 1. *A bibliometric ranking $\succsim$ satisfies conditions* **C1** *and* **C2** *if and only if (iff), for all $a, b \in A$, $a \succsim b \Leftrightarrow f_\tau(a) \geq f_\tau(b)$, for some $\tau \in \mathbb{N}$.*

Proof. Necessity is clear. We show sufficiency. Let $\tau \in \mathbb{N}$ be defined as in Lemma 2. In view of Lemma 3, for all non-null $a \in A$, we have $a \sim a^+(\tau)\mathbf{1}_\tau$. By construction, we know that $\mathbf{1}_\tau \succ \mathbf{0}$. Using **C1**, we obtain $(x+1)\mathbf{1}_\tau \succ x\mathbf{1}_\tau$, for all $x \in \mathbb{N}$. Hence, we have $a \succsim b \Leftrightarrow a^+(\tau) \geq b^+(\tau) \Leftrightarrow f_\tau(a) \geq f_\tau(b)$, for all non-null authors $a, b \in A$. The proof is complete observing that if $b \in A$ is such that $f_\tau(b)=0$, we must have $b \sim \mathbf{0}$ since we know from Lemma 2 that, for all $x \in \mathbb{N}$, $x<\tau \Rightarrow \mathbf{1}_x \sim \mathbf{0}$. □

Remark 8. Theorem 1 is essentially the same as Marchant (2009a, Th. 4, p. 334). We have used **A2** instead of CDNH and Lower Bound.

Observe however that Theorem 4 in Marchant (2009a) is stated for $\tau>0$. In fact, it is easy to check that it also holds for $\tau=0$. Moreover, our version of 2-gradedness differs from the one given in Marchant (2009a) saying that, for all $x, y, z \in \mathbb{N}$, $x>y>z \Rightarrow [\mathbf{1}_x \sim \mathbf{1}_y$ or $\mathbf{1}_y \sim \mathbf{1}_z]$. This does not forbid to have, for all $x, y \in \mathbb{N}_+$, $\mathbf{1}_x \sim \mathbf{1}_y \succ \mathbf{1}_0 \succ \mathbf{0}$. The ranking induced by the following index:

$$f(a) = a(0) + 2\sum_{x\geq 1} a(x),$$

satisfies all conditions in Marchant (2009a) but is not identical to $\succsim_\tau$. Our modified version of 2-gradedness corrects this point.

Adding the zero condition ORI and ES leads to a characterization of the index f_τ up to the multiplication by a positive constant.

We will use the following lemma.

Lemma 4. *If a bibliometric index f satisfies conditions* **C1**, **C2**, *ORI, and ES, then, for all $x, y \in \mathbb{N}$, $f(x\mathbf{1}_y)=xf(\mathbf{1}_y)$.*

Proof. Let $\tau \in \mathbb{N}$ be defined as in Lemma 2.

If $y<\tau$, Theorem 1 and ORI imply that $f(x\mathbf{1}_y)=f(\mathbf{1}_y)=f(\mathbf{0})=0$.

Let $y \geq \tau$. Using Theorem 1, we know that, for all $x \in \mathbb{N}$, $f(x\mathbf{1}_y)=f(x\mathbf{1}_\tau)$. Moreover, still using Theorem 1, we know that $\mathbf{1}_\tau$ is immediately above $\mathbf{0}$ and that, for all $x \in \mathbb{N}_+$, $(x+1)\mathbf{1}_\tau$ is immediately above $x\mathbf{1}_\tau$. Using ORI, we know that $f(\mathbf{0})=0$. Using ES we have $f(x\mathbf{1}_\tau)=xf(\mathbf{1}_\tau)$. Hence, we have $f(x\mathbf{1}_y)=f(x\mathbf{1}_\tau)=xf(\mathbf{1}_\tau)=xf(\mathbf{1}_y)$. □

Theorem 2. *A bibliometric index f satisfies conditions* **C1**, **C2**, *ORI, and ES iff, for all $a \in A$, $f(a)=\beta f_\tau(a)$, for some $\beta \in \mathbb{R}_+$ and some $\tau \in \mathbb{N}$.*

Proof. Necessity is clear. We show sufficiency. Let $\tau \in \mathbb{N}$ be defined as in Lemma 2. Let $f(\mathbf{1}_\tau)=\beta$. We know from Lemma 2 that $\beta>0$. If $a \in A$ is null, the thesis follows from ORI. In view of Lemma 3, for all non-null $a \in A$, we have $f(a)=f(a^+(\tau)\mathbf{1}_\tau)$. Using Theorem 1 and Lemma 4, we have $f(a)=f(a^+(\tau)\mathbf{1}_\tau)=a^+(\tau)f(\mathbf{1}_\tau)=a^+(\tau)\beta=\beta f_\tau(a)$. □

3.4. Independence of conditions

It is important in characterization results to use non-redundant conditions. Hence, we want to show that the conditions used in Theorem 1 (or in Theorem 2) are independent, i.e., that none of them is implied by the conjunction of the other ones. We do so by giving examples showing that it is possible to satisfy all but one of the conditions used in the above results.

When **C1** holds, condition **C3**, used in Theorem 3 below, implies condition **C2** (see Lemma 5). Hence, the examples used below to show that the conditions in Theorem 3 (resp. Theorem 4) are independent also show that the conditions used in Theorem 1 (resp. Theorem 2) are independent.

4. Number of papers

This is a standard bibliometric index (see, e.g., van & Raan, 2006).

4.1. Setting

We are interested in the index f_0 associating with each author $a \in A$ the total number of her papers, i.e., we have:

$$f_0(a) = p_a = \sum_{x\in\mathbb{N}} a(x).$$

We will also be interested in the bibliometric ranking $\succsim_0$ induced by f_0.

It is clear that this ranking (resp. index) is a particular instance of the ranking $\succsim_\tau$ (resp. index f_τ) with $\tau = 0$. Hence, all conditions that are satisfied by $\succsim_\tau$ (resp. f_τ) are also satisfied by $\succsim_0$ (resp. f_0).

4.2. Conditions

We will need the following condition. It says that, for authors having a single paper, the number of citations received by this paper is unimportant.

C3 (One is One) *For all* $x, y \in \mathbb{N}$, $\mathbf{1}_x \sim \mathbf{1}_y$.

Remark 9. The above condition is identical to condition A7 in Marchant (2009a, p. 329). It is clearly satisfied by $\succsim_0$. When $\tau > 0$, it is violated by $\succsim_\tau$.

As stated in Marchant (2009a), this condition says that two authors, each with exactly one publication, are equivalent irrespective of their number of citations. This condition is quite strong. It implies that, for authors with a single paper, citations are not viewed as a signal of impact.

Lemma 5. *If a bibliometric ranking satisfies conditions* **C1** *and* **C3**, *it also satisfies* **C2**.

Proof. From Lemma 1, we know that $\mathbf{1}_x \succ \mathbf{0}$, for some $x \in \mathbb{N}$. Using **C3**, we obtain that for all $x, y \in \mathbb{N}$, $\mathbf{1}_x \sim \mathbf{1}_y \succ \mathbf{0}$. This clearly implies **C2**. □

4.3. Results

Our first result, in this section, characterizes $\succsim_0$.

Theorem 3. *A bibliometric ranking* $\succsim$ *satisfies conditions* **C1** *and* **C3** *iff, for all* $a, b \in A$, $a \succsim b \Leftrightarrow f_0(a) \geq f_0(a)$.

Proof. Necessity is clear. We show sufficiency. From Lemma 5, we know that **C2** holds, so that Theorem 1 also holds. Using Lemma 1, we know that we have $\mathbf{1}_x \succ \mathbf{0}$, for some $x \in \mathbb{N}$. Then **C3** implies that $\mathbf{1}_x \sim \mathbf{1}_y \succ \mathbf{0}$, for all $x, y \in \mathbb{N}$. Hence, defining τ as in Lemma 2, we must have $\tau = 0$. □

Remark 10. Theorem 3 is identical to Marchant (2009a, Th. 1, p. 329) with **A2** replacing CDNH and Lower Bound. The proof given here is different however.

Adding conditions ORI and ES leads to a characterization of the index f_0 up to the multiplication by a positive constant.

Theorem 4. *A bibliometric index f satisfies conditions* **C1**, **C3**, *ORI, and ES iff, for all* $a \in A$, $f(a) = \beta f_0(a)$, *for some* $\beta \in \mathbb{R}_+$.

Proof. Necessity is clear. We show sufficiency. From Lemma 5, we know that **C2** holds, so that Theorem 2 also holds. Using Lemma 1, we know that we have $f(\mathbf{1}_x) > 0$, for some $x \in \mathbb{N}$. Then **C3** implies that $f(\mathbf{1}_x) = f(\mathbf{1}_y) > 0$, for all $x, y \in \mathbb{N}$. Hence, defining τ as in Lemma 2, we must have $\tau = 0$. □

4.4. Independence of conditions

We show below that the conditions used in Theorem 4 are independent (in order not to multiply examples, conditions **A1** and **A2** will always be assumed when discussing the independence of conditions).

Example 1 *(***C1** *Independence).* Consider the bibliometric index such that $f(a) = 1$, for all non-null authors and $f(\mathbf{0}) = 0$. It is clear that this index satisfies **A1** and **A2**.

This index clearly violates **C1** since, e.g., we have $f(\mathbf{1}_1) = 1 > f(\mathbf{0}) = 0$ but $f(2\mathbf{1}_1) = 1 = f(\mathbf{0} + \mathbf{1}_1)$. For all $x, y \in \mathbb{N}$, we have $f(\mathbf{1}_x) = f(\mathbf{1}_y) = 1$, so that **C3** holds. Condition ORI clearly holds. Condition ES trivially holds.

Remark 11. It is easy to check that the above example satisfies **C7**, a condition weakening **C1** that is defined below in Section 8.2. This shows that it is not possible to replace **C1** by **C7** in Theorems 1–4.

Example 2 *(***C3** *One is One).* Consider the bibliometric index defined by

$$f_c(a) = c_a = \sum_{x \in \mathbb{N}} x a(x),$$

i.e., the index giving the total number of citations received by the papers of an author. This index clearly satisfies **A1** and **A2**. It clearly violates **C3** since, e.g., $f(\mathbf{1}_2) = 2 > f(\mathbf{1}_1) = 1$. It is easy to check that all other conditions (**C1**, ORI, and ES) are satisfied (see Theorem 9 below).

Rephrasing the above examples in terms of the bibliometric ranking induced by the bibliometric index shows that the conditions used in Theorem 3 are independent.

Example 3 *(ORI).* Consider the bibliometric index f such that, for all $a \in A$, $f(a) = f_0(a) + 1$. This index clearly satisfies **A1** and **A2**. It clearly violates ORI. It is easy to check that all other conditions (**C1**, **C3**, and ES) are satisfied.

Example 4 *(ES)*. Consider the bibliometric index that assigns to all $a \in A$ the square of the value $f_0(a)$. It satisfies **A1** and **A2**. It clearly violates ES. It is easy to check that all other conditions (**C1**, **C3**, and ORI) are satisfied.

Remark 12. The above two examples showing that ORI (adding a positive constant to the value of the index) and ES (taking a strictly monotonic transformation of the value of the index) are independent will be used for all indices that we study. Instead of repeating them below, we will refer to them as the "standard examples".

4.5. Remarks

Theorem 3 characterizes $\succsim_0$ by replacing **C2** in Theorem 1 with **C3**. Clearly, an alternative characterization is obtained, keeping unchanged all conditions used in Theorem 1 and adding to them the requirement that $\mathbf{1}_0 \succ \mathbf{0}$. This alternative characterization uses three conditions that are independent (Example 1 satisfies this new condition and **C2** but violates **C1**. The index f_τ, with $\tau = 1$ satisfies **C1** and **C2** but violates the new condition. The index consisting of the sum of the number of papers and the number of citations satisfies **C1** and the new condition but violates **C2**).

Let us also observe that we may replace **C1** in Theorem 3 by the conjunction of (1) and (2) while keeping a set of independent conditions. We have already observed that Example 1 satisfies (1) as well as **C3**. It clearly violates (2) since we have $\mathbf{1}_1 \succ \mathbf{0}$ but $\mathbf{1}_1 + \mathbf{1}_1 \sim \mathbf{0} + \mathbf{1}_1$. The ranking in which all authors having at most one paper are tied with $\mathbf{0}$ and all other authors are ranked according to the number of papers (minus one if one wishes to satisfy ES) gives an example satisfying (2) and **C3** but violating (1) since $\mathbf{0} \succsim \mathbf{1}_1$ but $\mathbf{1}_1 + \mathbf{1}_1 \succ \mathbf{0} + \mathbf{1}_1$.

5. Number of citations of highly cited papers

To our knowledge, the literature has never made use of this index. It is a variant of the index based on the total number of citations, correcting it to take only "important papers" into account.

5.1. Setting

Let $\tau \in \mathbb{N}_+$. We are interested in the index f_{c_τ} associating with each author $a \in A$ the total number of citations received by her papers that have received at least τ citations, i.e., we have:

$$f_{c_\tau}(a) = c_\tau(a) = \sum_{x \geq \tau} x a(x).$$

We will also be interested in the bibliometric ranking $\succsim_{c_\tau}$ induced by f_{c_τ}.

5.2. Conditions

It is not difficult to check that f_{c_τ} satisfies condition **C1**.

We will need the following condition.

C4 (Restricted Additivity) *For all* $x, y \in \mathbb{N}$, $\mathbf{1}_x \succ \mathbf{0}$ *and* $\mathbf{1}_y \succ \mathbf{0}$ *imply* $[\mathbf{1}_x + \mathbf{1}_y] \sim \mathbf{1}_{x+y}$.

The above condition says that, for authors having a single publication that are better than the null author, getting y more citations for the single paper is equivalent to publishing an additional paper with y citations, as long as the author having a single paper with y citations is not considered equivalent to the null author. It is easy to check that this condition is violated by $\succsim_0$. When $\tau > 0$, this condition is also violated by $\succsim_\tau$.

The following lemmas will be useful.

Lemma 6. *If a bibliometric ranking* $\succsim$ *satisfies* **C1** *and* **C4**, *then* $\mathbf{1}_0 \sim \mathbf{0}$.

Proof. **A2** implies $\mathbf{1}_0 \succsim \mathbf{0}$. Suppose that $\mathbf{1}_0 \succ \mathbf{0}$. **A2** implies $\mathbf{1}_1 \succ \mathbf{0}$. Using **C1**, $\mathbf{1}_0 \succ \mathbf{0}$ implies $\mathbf{1}_0 + \mathbf{1}_1 \succ \mathbf{0} + \mathbf{1}_1 = \mathbf{1}_1$. Since $\mathbf{1}_0 \succ \mathbf{0}$ and $\mathbf{1}_1 \succ \mathbf{0}$, we use **C4** to obtain $\mathbf{1}_0 + \mathbf{1}_1 \sim \mathbf{1}_{0+1} = \mathbf{1}_1$, a contradiction. □

Lemma 7. *If a bibliometric ranking* $\succsim$ *satisfies* **C1** *and* **C4**, *then there is* $\tau \in \mathbb{N}_+$ *such that, for all* $x \in \mathbb{N}$, $x < \tau \Rightarrow \mathbf{1}_x \sim \mathbf{0}$.

Proof. We know from **A2** that $\mathbf{1}_x \succsim \mathbf{0}$, for all $x \in \mathbb{N}$. Moreover, Lemma 1 implies that $\mathbf{1}_x \succ \mathbf{0}$, for some $x \in \mathbb{N}$. Define $\tau \in \mathbb{N}$ to be the smallest $x \in \mathbb{N}$ such that $\mathbf{1}_x \succ \mathbf{0}$. By construction, for all $x < \tau$, we have $\mathbf{1}_x \sim \mathbf{0}$. Lemma 6 implies that we have $\tau > 0$. □

Lemma 8. *If a bibliometric ranking* $\succsim$ *satisfies* **C1** *and* **C4**, *then, for all* $a \in A$, $a \sim a + \mathbf{1}_x$ *for all* $x \in \mathbb{N}$ *with* $x < \tau$ *and* τ *defined as in Lemma 7.*

Proof. We know from Lemma 7 that $x < \tau$ implies that $\mathbf{1}_x \sim \mathbf{0}$. The conclusion follows from **C1**. □

Lemma 9. *If a bibliometric ranking* $\succsim$ *satisfies* **C1** *and* **C4**, *then, for all* $x, y \in \mathbb{N}$, *such that* $x > y \geq \tau$, *with* τ *defined as in Lemma 7*, $\mathbf{1}_x \succ \mathbf{1}_y$.

Proof. Let $x, y \in \mathbb{N}$, such that $x > y \geq \tau$. Since $y \geq \tau$, we know that $\mathbf{1}_y \succ \mathbf{0}$. Using **A2**, we have $\mathbf{1}_x \succsim \mathbf{1}_y \succ \mathbf{0}$. Suppose, in contradiction with the thesis, that $\mathbf{1}_x \sim \mathbf{1}_y$. Using **C1**, this implies $y\mathbf{1}_x \sim y\mathbf{1}_y$. Using **C4** repeatedly, we obtain $\mathbf{1}_{xy} \sim \mathbf{1}_{yy}$. Since $x > y$ and $\mathbf{1}_y \succ \mathbf{0}$, **C1** implies $x\mathbf{1}_y \succ y\mathbf{1}_y$. Using **C4** repeatedly, we obtain $\mathbf{1}_{xy} \succ \mathbf{1}_{yy}$, a contradiction. □

Lemma 10. *Let $\succsim$ be a bibliometric ranking satisfying* **C1** *and* **C4**. *Define τ as in Lemma 7. Let $a \in A$ and let $c_\tau(a) = \sum_{x \geq \tau} x a(x)$. Then $a \sim \mathbf{1}_{c_\tau(a)}$.*

Proof. Suppose first that $a \in A$ has no paper having been cited at least τ times, so that $c_\tau(a) = 0$. Using Lemmas 6 and 7 and **C1**, we obtain that $a \sim \mathbf{1}_0 \sim \mathbf{0}$.

Suppose now that $a \in A$ is such that

$$a = \mathbf{1}_{x_1} + \mathbf{1}_{x_2} + \cdots + \mathbf{1}_{x_n} + \mathbf{1}_{y_1} + \mathbf{1}_{y_2} + \cdots + \mathbf{1}_{y_m},$$

with $x_1, x_2, \ldots, x_n \geq \tau$ and $y_1, y_2, \ldots, y_m < \tau$.

Using Lemma 7 and **C1**, it is easy to see that $a \sim [\mathbf{1}_{x_1} + \mathbf{1}_{x_2} + \cdots + \mathbf{1}_{x_n}]$. Using **C4** repeatedly, we obtain $a \sim \mathbf{1}_{x_1 + x_2 + \cdots + x_n} \sim \mathbf{1}_{c_\tau(a)}$. □

5.3. Results

Our first result, in this section, characterizes $\succsim_{c_\tau}$.

Theorem 5. *A bibliometric ranking $\succsim$ satisfies conditions* **C1** *and* **C4**, *iff, for all $a, b \in A$, $a \succsim b \Leftrightarrow c_\tau(a) \geq c_\tau(b)$, for some $\tau \in \mathbb{N}_+$.*

Proof. Necessity is clear. We show sufficiency. Let $\tau \in \mathbb{N}_+$ be defined as in Lemma 7. Lemma 10 implies that, for all $a \in A$, we have $a \sim \mathbf{1}_{c_\tau(a)}$. We have $a \succsim b \Leftrightarrow \mathbf{1}_{c_\tau(a)} \succsim \mathbf{1}_{c_\tau(b)}$. In view of Lemma 9, this holds iff $c_\tau(a) \geq c_\tau(b)$. □

The index f_{c_τ} satisfies ORI but we already observed that it violates ES (unless $\tau = 1$). Instead of characterizing f_{c_τ} we will characterize the index $f^*_{c_\tau}$ that is such that

$$f^*_{c_\tau}(a) = \begin{cases} 0 & \text{if} \quad f_{c_\tau}(a) = 0, \\ f_{c_\tau}(a) + 1 - \tau & \text{otherwise.} \end{cases}$$

Clearly the indices f_{c_τ} and $f^*_{c_\tau}$ induce the same ranking on the set of authors. The difference between f_{c_τ} and $f^*_{c_\tau}$ is that $f^*_{c_\tau}$ has been rescaled so as to satisfy ES.

Adding conditions ORI and ES to the conditions used in Theorem 5 leads to a characterization of the index $f^*_{c_\tau}$ up to the multiplication by a positive constant.

Theorem 6. *A bibliometric index f satisfies conditions* **C1**, **C4**, *ORI, and ES iff, for all $a \in A$, $f(a) = \beta f^*_{c_\tau}(a)$, for some $\beta \in \mathbb{R}_+$ and some $\tau \in \mathbb{N}_+$.*

Proof. Necessity is clear. We show sufficiency. Let $\tau \in \mathbb{N}_+$ be defined as in Lemma 7. Let $f(\mathbf{1}_\tau) = \beta > 0$. If $c_\tau(a) = 0$, we know from Lemma 10 that $a \sim \mathbf{1}_0 \sim \mathbf{0}$. Using ORI, we have $f(a) = 0$, while, by definition, we have $f^*_{c_\tau}(a) = 0$. Hence, we have $f(a) = \beta f^*_{c_\tau}(a)$. Suppose now that $c_\tau(a) > 0$, which, by construction, implies that $c_\tau(a) \geq \tau$. Lemma 10 implies that, for all $a \in A$, we have $f(a) = f(\mathbf{1}_{c_\tau(a)})$. Using ORI and ES implies that $f(a) = f(\mathbf{1}_{c_\tau(a)}) = (c_\tau(a) + 1 - \tau) f(\mathbf{1}_\tau) = \beta(c_\tau(a) + 1 - \tau) = \beta f^*_{c_\tau}(a)$. □

5.4. Independence of conditions

In presence of **C1**, condition **C6**, used in Theorem 9 below, implies condition **C4** (see Lemma 16). Hence, the examples used below to show that the conditions in Theorem 9 (resp. Theorem 10) are independent also show that the conditions used in Theorem 5 (resp. Theorem 6) are independent.

6. Number of citations exceeding a threshold

To our knowledge, the literature has never made use of this index (it should be noted that Albarrán, Ortuño, & Ruiz-Castillo, 2011a, 2011b have used a similar idea to propose and analyze indices for groups of authors that are independent of the size of the group[6]). It is a variant of the index based on the total number of citations, correcting it to take only citations exceeding a threshold into account.

[6] We thank Lude Waltman for having brought these papers to our attention.

6.1. Setting

Let $\tau \in \mathbb{N}_+$. We are interested in the index f_{t_τ} associating with each author $a \in A$ the total number of citations above τ received by her papers, i.e., we have:

$$f_{t_\tau}(a) = t_\tau(a) = \sum_{x \geq \tau} (x + 1 - \tau) a(x).$$

We will also be interested in the bibliometric ranking $\succsim_{t_\tau}$ induced by f_{t_τ}.

Observe that this index is different from f_{c_τ} and $f^*_{c_\tau}$. For instance, with $\tau = 10$, we have: $f_{t_\tau}(\mathbf{1}_{12}) = f^*_{c_\tau}(\mathbf{1}_{12}) = 3$, while $f_{t_\tau}(2\mathbf{1}_{10}) = 2$ and $f^*_{c_\tau}(2\mathbf{1}_{10}) = 11$.

It is easy to check that f_{t_τ} satisfies **C1**, as well as ORI and ES. It violates **C4**. Indeed, taking $\tau = 10$, we have $f_{t_\tau}(\mathbf{1}_{10}) = 1 > f_{t_\tau}(\mathbf{0}) = 0$, while $f_{t_\tau}(\mathbf{1}_{10} + \mathbf{1}_{10}) = 2 \neq f_{t_\tau}(\mathbf{1}_{10+10}) = 11$.

6.2. Conditions

We already observed that f_{t_τ} satisfies conditions **C1**, ORI, and ES, whatever $\tau \geq 1$. Clearly it also satisfies condition **C9**, introduced later in Section 8.2 (saying that $\mathbf{1}_0 \sim \mathbf{0}$). We will need the following condition.

C5 (Restricted Transfer) *For all $x, y \in \mathbb{N}$, with $y > 0$, if $\mathbf{1}_x \succ \mathbf{0}$ and $\mathbf{1}_y \succ \mathbf{0}$, then $[\mathbf{1}_x + \mathbf{1}_y] \sim \mathbf{1}_{x+1} + \mathbf{1}_{y-1}$.*

It is easy to check that the above condition is satisfied by $\succsim_{t_\tau}$, whatever $\tau \geq 1$. It is easy to check that condition **C5** is violated by $\succsim_\tau$ (with $\tau > 0$) and $\succsim_{c_\tau}$ (with $\tau > 1$). It is satisfied by $\succsim_0$ and $\succsim_c$. The following lemmas will be useful.

Lemma 11. *If a bibliometric ranking $\succsim$ satisfies* **C1**, *then $\mathbf{1}_x \succ \mathbf{0}$, for some $x \in \mathbb{N}$. There is $\tau \in \mathbb{N}$ such that, for all $x \in \mathbb{N}$, $x < \tau$ implies $\mathbf{1}_x \sim \mathbf{0}$. Moreover, if* **C9** *holds, we have $\tau \geq 1$.*

Proof. The first part results from Lemma 1. We know from **A2** that $\mathbf{1}_x \succsim \mathbf{0}$, for all $x \in \mathbb{N}$. Define $\tau \in \mathbb{N}$ to be the smallest $x \in \mathbb{N}$ such that $\mathbf{1}_x \succ \mathbf{0}$. By construction, for all $x < \tau$, we have $\mathbf{1}_x \sim \mathbf{0}$. The last part is a direct consequence of **C9**. □

Lemma 12. *Let $\succsim$ be a bibliometric ranking satisfying* **C1** *and* **C5**. *For all $y \in \mathbb{N}$, $\mathbf{1}_y \succ \mathbf{0}$ implies $\mathbf{1}_{y+1} \succ \mathbf{1}_y$.*

Proof. Define τ as in Lemma 11. By construction, $\mathbf{1}_y \succ \mathbf{0}$ implies $y \geq \tau$. We know that $\mathbf{1}_\tau \succ \mathbf{0}$ and $\mathbf{1}_{\tau-1} \sim \mathbf{0}$.

Let us first show that $\mathbf{1}_{\tau+1} \succ \mathbf{1}_\tau$. Suppose, contrary to the thesis, in view of **A2**, that $\mathbf{1}_{\tau+1} \sim \mathbf{1}_\tau$. Using **C5**, we have $\mathbf{1}_\tau + \mathbf{1}_\tau \sim \mathbf{1}_{\tau+1} + \mathbf{1}_{\tau-1}$. Using **C1**, this is contradictory since $\mathbf{1}_{\tau+1} \sim \mathbf{1}_\tau \succ \mathbf{1}_{\tau-1} \sim \mathbf{0}$.

Let us now show that $\mathbf{1}_{\tau+2} \succ \mathbf{1}_{\tau+1}$. Suppose, contrary to the thesis, in view of **A2**, that $\mathbf{1}_{\tau+2} \sim \mathbf{1}_{\tau+1}$. We know that $\mathbf{1}_{\tau+1} \succ \mathbf{1}_\tau \succ \mathbf{0}$. Using **C5**, we have $\mathbf{1}_{\tau+1} + \mathbf{1}_{\tau+1} \sim \mathbf{1}_{\tau+2} + \mathbf{1}_\tau$. Using **C1**, this is contradictory since $\mathbf{1}_{\tau+2} \sim \mathbf{1}_{\tau+1} \succ \mathbf{1}_\tau$. Repeating the above reasoning leads to the desired conclusion. □

Lemma 13. *Let $\succsim$ be a bibliometric ranking satisfying* **C1**, **C5**, *and* **C9**. *Define τ as in Lemma 11. Let $a \in A$ and let $t_\tau(a) = \sum_{x \geq \tau}(x + 1 - \tau)a(x)$. If $t_\tau(a) = 0$, then $a \sim \mathbf{0}$. If $t_\tau(a) > 0$, then $a \sim \mathbf{1}_{\tau-1+t_\tau(a)}$.*

Proof. Suppose first that $a \in A$ has no papers having been cited at least τ times, so that $t_\tau(a) = 0$. Using Lemma 11 and **C1** implies that $a \sim \mathbf{0}$.

Suppose now that $a \in A$ is such that, for some $n \in \mathbb{N}_+$ and some $m \in \mathbb{N}$,

$$a = \mathbf{1}_{x_1} + \mathbf{1}_{x_2} + \cdots + \mathbf{1}_{x_n} + \mathbf{1}_{y_1} + \mathbf{1}_{y_2} + \cdots + \mathbf{1}_{y_m},$$

with $x_1, x_2, \ldots, x_n \geq \tau$ and $y_1, y_2, \ldots, y_m < \tau$, so that $t_\tau(a) > 0$.

Using Lemma 11 and **C1**, it is easy to see that $a \sim [\mathbf{1}_{x_1} + \mathbf{1}_{x_2} + \cdots + \mathbf{1}_{x_n}]$. By construction, we know that $\mathbf{1}_{x_i} \succ \mathbf{0}$, for $i = 1, 2, \ldots, n$. Let $x_i = \tau + \alpha_i$, for $i = 1, 2, \ldots, n$, so that we have $a \sim [\mathbf{1}_{\tau+\alpha_1} + \mathbf{1}_{\tau+\alpha_2} + \cdots + \mathbf{1}_{\tau+\alpha_n}]$. By construction, we know that $t_\tau(a) = \sum_{i=1}^n (\alpha_i + 1)$. Using **C5**, we have $a \sim [\mathbf{1}_{\tau+\alpha_1+1} + \mathbf{1}_{\tau+\alpha_2} + \cdots + \mathbf{1}_{\tau+\alpha_{n-1}} + \mathbf{1}_{\tau+\alpha_n-1}]$. Repeating the process leads to $a \sim [\mathbf{1}_{\tau+\alpha_1+\alpha_n+1} + \mathbf{1}_{\tau+\alpha_2} + \cdots + \mathbf{1}_{\tau+\alpha_{n-1}} + \mathbf{1}_{\tau-1}]$. Since we know from Lemma 11 that $\mathbf{1}_{\tau-1} \sim \mathbf{0}$, using **C1**, we obtain $a \sim [\mathbf{1}_{\tau+\alpha_1+\alpha_n+1} + \mathbf{1}_{\tau+\alpha_2} + \cdots + \mathbf{1}_{\tau+\alpha_{n-1}}]$. Repeating the process, we obtain $a \sim [\mathbf{1}_{\tau+\alpha_1+\alpha_2+\cdots\alpha_n+(n-1)}]$. We know that $t_\tau(a) = \sum_{i=1}^n (\alpha_i + 1)$, so that we have $a \sim \mathbf{1}_{\tau-1+t_\tau(a)}$. □

6.3. Results

Our first result, in this section, characterizes $\succsim_{t_\tau}$.

Theorem 7. *A bibliometric ranking $\succsim$ satisfies conditions* **C1**, **C5**, *and* **C9** *iff, for all $a, b \in A$, $a \succsim b \Leftrightarrow t_\tau(a) \geq t_\tau(b)$, for some $\tau \in \mathbb{N}_+$.*

Proof. Necessity is clear. We show sufficiency. Let $\tau \in \mathbb{N}_+$ be defined as in Lemma 11. If $t_\tau(a) = 0$, Lemma 13 implies that $a \sim \mathbf{0}$. If $t_\tau(a) \geq 1$, Lemma 13 implies that $a \sim \mathbf{1}_{\tau-1+t_\tau(a)}$. By construction, we know that $\mathbf{1}_{\tau-1+t_\tau(a)} \succ \mathbf{0}$. The conclusion therefore follows using Lemma 12. □

Adding conditions ORI and ES to the conditions used in Theorem 7 leads to a characterization of the index f_{t_τ} up to the multiplication by a positive constant.

Theorem 8. *A bibliometric index f satisfies conditions* **C1**, **C5**, **C9**, *ORI, and ES iff, for all* $a \in A$, $f(a) = \beta f_{t_\tau}(a)$, *for some* $\beta \in \mathbb{R}_+$ *and some* $\tau \in \mathbb{N}_+$.

Proof. Necessity is clear. We show sufficiency. Let $\tau \in \mathbb{N}_+$ be defined as in Lemma 11. Let $f(\mathbf{1}_\tau) = \beta > 0$.

If $t_\tau(a) = 0$, we know, using Theorem 7, that $f(a) = f(\mathbf{0})$. Using ORI, we know that $f(\mathbf{0}) = 0$, so that the conclusion holds in this case. If $t_\tau(a) \geq 1$, we know, using Theorem 7, that $f(a) = f(\mathbf{1}_{\tau-1+t_\tau(a)})$. Using ES, it is clear that $f(\mathbf{1}_{\tau-1+t_\tau(a)}) = t_\tau(a) f(\mathbf{1}_\tau) = \beta f_{t_\tau}(a)$. □

Remark 13. The rankings $\succsim_{c_\tau}$ and $\succsim_{t_\tau}$ are both based on the citation count of highly cited papers. Although they aim at capturing the same basic idea, they do so in different ways. Examining their characterizing properties do not give completely convincing arguments in favor of one of them. The index f_{c_τ} violates ES. This has motivated the introduction of the rescaled index $f^*_{c_\tau}$. This seems to favor f_{t_τ} over $f^*_{c_\tau}$. This is all the more true since the index $f^*_{c_\tau}$ has jumps when going from authors with one paper to authors with more than one paper that are not completely intuitive.

6.4. Independence of conditions

We show below that the conditions used in Theorem 8 are independent.

Example 5 *(***C9** *Quasi-null authors).* Define f as the bibliometric index such that $f_a = f_0(a) + f_c(a)$ (i.e., the sum of the numbers of papers and the number of citations). It is clear that this index satisfies **A1** and **A2**. Conditions ORI and ES are satisfied. **C9** is violated since $f(\mathbf{1}_0) = 1$. Condition **C1** and **C5** clearly holds.

Example 6 *(***C5** *Restricted Transfer).* The bibliometric index $f^*_{c_\tau}$ clearly satisfies **C1** and **C9**. It is easy to check that it violates **C5**. For instance, with $\tau = 3$, we have $\mathbf{1}_4 + \mathbf{1}_3 \succ_{c_\tau} \mathbf{1}_5 + \mathbf{1}_2$. Conditions ORI and ES clearly hold.

Example 7 *(***C1** *Independence).* Consider the bibliometric index such that $f(a) = 0$ for all authors with at most 1 paper and $f(a) = 1$ otherwise.

It is clear that this index satisfies **A1** and **A2**. Condition **C9** clearly holds. Condition **C5** holds since all authors having two papers are tied. Condition **C1** is violated since $\mathbf{1}_1 \sim \mathbf{0}$ but $2\mathbf{1}_1 \succ \mathbf{1}_1$. Conditions ORI and ES trivially hold.

Rephrasing the above examples in terms of the bibliometric ranking induced by the bibliometric index shows that the conditions used in Theorem 7 are independent. The standard examples show that ORI and ES cannot be omitted in Theorem 8.

7. Number of citations

This is a standard bibliometric index (see, e.g., van & Raan, 2006).

7.1. Setting

We are interested in the index f_c associating with each author $a \in A$ the total number of citations received by her papers, i.e., we have:

$$f_c(a) = c_a = \sum_{x \in \mathbb{N}} x a(x).$$

We will also be interested in the bibliometric ranking $\succsim_c$ induced by f_c.

It is clear that this ranking (resp. index) is a particular instance of the ranking $\succsim_{c_\tau}$ (resp. index f_{c_τ}) with $\tau = 1$. The same is true with $\succsim_{t_\tau}$ and f_{t_τ} with $\tau = 1$. Hence, all conditions that are satisfied by $\succsim_{c_\tau}$ (resp. f_{c_τ}) are also satisfied by $\succsim_c$ (resp. f_c). The same is true with $\succsim_{t_\tau}$ and f_{t_τ}.

7.2. Conditions

We only need one new condition that is clearly satisfied by $\succsim_c$.

C6 (Additivity) *For all* $x \in \mathbb{N}$, $\mathbf{1}_x + \mathbf{1}_1 \sim \mathbf{1}_{x+1}$.

When $\tau > 1$, the above condition is violated by $\succsim_{c_\tau}$ and $\succsim_{t_\tau}$. It is clearly violated by $\succsim_0$ and by $\succsim_\tau$ with $\tau > 0$.

Remark 14. The above condition is identical to Condition A8 in Marchant (2000a, p. 330). It is clearly satisfied by $\succsim_c$. It says that for authors having a single publication, obtaining one more citation for that paper or publishing one additional paper with one citation has the same impact.

Our first lemma is identical to Lemma 6 with **C6** replacing **C4**.

Lemma 14. *If a bibliometric ranking* $\succsim$ *satisfies* **C1** *and* **C6**, *then* $\mathbf{1}_0 \sim \mathbf{0}$.

Proof. Using **A2**, we know that $\mathbf{1}_0 \succsim \mathbf{0}$. Suppose that $\mathbf{1}_0 \succ \mathbf{0}$. Using **C1**, we obtain $\mathbf{1}_0 + \mathbf{1}_1 \succ \mathbf{0} + \mathbf{1}_1 = \mathbf{1}_1$. Using **C6**, we know that $\mathbf{1}_0 + \mathbf{1}_1 \sim \mathbf{1}_1$. Hence, we obtain $\mathbf{1}_1 \succ \mathbf{1}_1$, a contradiction. □

The following lemmas will be useful.

Lemma 15. *If a bibliometric ranking* $\succsim$ *satisfies* **C1** *and* **C6***, then* $\mathbf{1}_1 \succ \mathbf{0}$.

Proof. Suppose, in contradiction with the thesis that $\mathbf{0} \succsim \mathbf{1}_1$. Using **A2**, we have $\mathbf{1}_1 \sim \mathbf{0}$. Using **C1**, we obtain $\mathbf{1}_1 + \mathbf{1}_1 \sim \mathbf{0} + \mathbf{1}_1 = \mathbf{1}_1$. Using **C6**, we know that $\mathbf{1}_1 + \mathbf{1}_1 \sim \mathbf{1}_2$, so that we have $\mathbf{1}_2 \sim \mathbf{1}_1 \sim \mathbf{0}$. Repeating the same reasoning shows that we have $\mathbf{1}_x \sim \mathbf{0}$, for all $x \in \mathbb{N}_+$. Lemma 14 has shown that $\mathbf{1}_0 \sim \mathbf{0}$. Hence, we obtain a violation of Lemma 1. □

Lemma 16. *If a bibliometric ranking* $\succsim$ *satisfies* **C1** *and* **C6***, then it satisfies* **C4**.

Proof. We have to show that, for all $x, y \in \mathbb{N}$, $\mathbf{1}_x \succ \mathbf{0}$ and $\mathbf{1}_y \succ \mathbf{0}$ imply $\mathbf{1}_x + \mathbf{1}_y \sim \mathbf{1}_{x+y}$. The claim is trivial, using **C6**, if $y = 1$. Let us show that it holds for $y = 2$. Using **C6**, we know that $\mathbf{1}_2 \sim \mathbf{1}_1 + \mathbf{1}_1$. Using **C1**, $\mathbf{1}_x + \mathbf{1}_2 \sim \mathbf{1}_x + \mathbf{1}_1 + \mathbf{1}_1$. **C6** implies that $\mathbf{1}_{x+1} \sim \mathbf{1}_x + \mathbf{1}_1$. Hence, using **C1**, $\mathbf{1}_x + \mathbf{1}_1 + \mathbf{1}_1 \sim \mathbf{1}_{x+1} + \mathbf{1}_1$. Using **C6** again, we obtain $\mathbf{1}_x + \mathbf{1}_1 + \mathbf{1}_1 \sim \mathbf{1}_{x+2}$. Hence, we have $\mathbf{1}_x + \mathbf{1}_2 \sim \mathbf{1}_{x+2}$. Repeating the above reasoning proves the claim. □

Lemma 17. *If a bibliometric ranking satisfies* **C1** *and* **C6***, then it satisfies* **C5**.

Proof. Using **C6**, we know that $\mathbf{1}_x + \mathbf{1}_1 \sim \mathbf{1}_{x+1}$ and $\mathbf{1}_{y-1} + \mathbf{1}_1 \sim \mathbf{1}_y$. Using **C1**, we have $\mathbf{1}_{x+1} + \mathbf{1}_{y-1} \sim \mathbf{1}_x + \mathbf{1}_1 + \mathbf{1}_{y-1} \sim \mathbf{1}_x + \mathbf{1}_y$. This clearly implies **C5**. □

7.3. Results

Our first result, in this section, characterizes $\succsim_c$.

Theorem 9. *A bibliometric ranking* $\succsim$ *satisfies* **C1** *and* **C6** *iff, for all* $a, b \in A$, $a \succsim b \Leftrightarrow f_c(a) \geq f_c(b)$.

Proof. Necessity is clear. We show sufficiency. From Lemma 16, we know that **C4** holds, so that Theorem 5 also holds. Using Lemmas 14 and 15, we have $\mathbf{1}_0 \sim \mathbf{0}$ and $\mathbf{1}_1 \succ \mathbf{0}$. Hence, defining τ as in Lemma 7, we must have $\tau = 1$. □

Remark 15. Theorem 9 is almost identical to Marchant (2009a). We have used here **A2** instead of Lower Bound and CDNH. The proof given here is different however.

Adding conditions ORI and ES leads to a characterization of the index f_c up to the multiplication by a positive constant.

Theorem 10. *A bibliometric index f satisfies* **C1**, **C6**, *ORI, and ES iff, for all* $a \in A$, $f(a) = \beta f_c(a)$, *for some* $\beta \in \mathbb{R}_+$.

Proof. Necessity is clear. We show sufficiency. From Lemma 16, we know that **C4** holds, so that Theorem 6 also holds. Using Lemmas 14 and 15, we have $f(\mathbf{1}_0) = f(\mathbf{0})$ and $f(\mathbf{1}_1) > f(\mathbf{0})$. Hence, defining τ as in Lemma 7, we must have $\tau = 1$. Using Lemma 10, we know that $f(a) = f(\mathbf{1}_{c_a})$. Let $\beta = f(\mathbf{1}_1) > f(\mathbf{0})$. ORI implies that $\beta > 0$. Using Theorem 9, ORI and ES, we obtain $f(a) = f(\mathbf{1}_{c_a}) = \beta c_a = \beta f_c(a)$. □

7.4. Independence of conditions

We show below that the conditions used in Theorem 10 are independent.

Example 8 *(***C1** *Independence).* Consider the bibliometric index such that $f(a) = 1$, for all non-null authors and $f(\mathbf{0}) = 0$. It is clear that this index satisfies **A1** and **A2**.

This index clearly violates **C1** since, e.g., we have $f(\mathbf{1}_1) = 1 > f(\mathbf{0}) = 0$ but $f(2\mathbf{1}_1) = 1 = f(\mathbf{0} + \mathbf{1}_1)$. For all $x \in \mathbb{N}$, we have $f(\mathbf{1}_x + \mathbf{1}_1) = f(\mathbf{1}_{x+1}) = 1$, so that **C6** holds. Condition ORI clearly holds. Condition ES trivially holds.

Remark 16. It is easy to check that the above example satisfies **C7**, a condition weakening **C1** that is defined below in Section 8.2. This shows that it is not possible to replace **C1** by **C7** in Theorems 5–10.

Example 9 *(***C6** *Additivity).* The index f_0 clearly violates **C6**. Theorem 4 has shown that it satisfies all other conditions.

Rephrasing the above examples in terms of the bibliometric ranking induced by the bibliometric index shows that the conditions used in Theorem 9 are independent. The standard examples show that ORI and ES cannot be omitted in Theorem 10.

7.5. Remarks

Theorem 9 characterizes $\succsim_c$ by replacing **C4** in Theorem 5 with **C6**. Clearly, an alternative characterization is obtained, keeping unchanged all conditions used in Theorem 5 and adding to them the requirement that $\mathbf{1}_1 \succ \mathbf{1}_0$. This alternative characterization uses three conditions that are independent. Example 8 satisfies this new condition and **C4** but violates **C1**. The index $f^*_{c_\tau}$, with $\tau = 2$ satisfies **C1** and **C4** but violates the new condition. The index consisting of the sum of the number of papers and the number of citations satisfies **C1** and the new condition but violates **C4** since we have $\mathbf{1}_2 \succ \mathbf{0}$ and $\mathbf{1}_3 \succ \mathbf{0}$ but $\mathbf{1}_2 + \mathbf{1}_3 \succ \mathbf{1}_5$.

Let us also observe that we may replace **C1** in Theorem 9 by the conjunction of (1) and (2) while keeping a set of independent conditions. We have already observed that Example 8 satisfies (1) as well as **C6**. It clearly violates (2) since we have $\mathbf{1}_1 \succ \mathbf{0}$ but $\mathbf{1}_1 + \mathbf{1}_1 \sim \mathbf{0} + \mathbf{1}_1$. The ranking in which all authors having at most one citation are tied and all other authors are ranked according to the number of citations (minus one if one wishes to satisfy ES) gives an example satisfying (2) and **C3** but violating (1) since $\mathbf{0} \succsim \mathbf{1}_1$ but $\mathbf{1}_1 + \mathbf{1}_1 \succ \mathbf{0} + \mathbf{1}_1$.

8. Maximum number of citations

This is a bibliometric index that is less frequently used than the ones analyzed so far. It is nevertheless sometimes used in the literature (see Eto, 2003).

8.1. Setting

We are interested in the index f_M associating with each author $a \in A$ the number of citations received by her most cited paper, i.e., we have:

$$f_M(a) = \max\{x \in \mathbb{N} : a(x) > 0\},$$

where it is understood that taking the maximum over an empty set leads to the value 0.

We will also be interested in the bibliometric ranking $\succsim_M$ induced by f_M.

8.2. Conditions

The ranking $\succsim_M$ violates **C1** (simple examples show that it also violates **C2**, **C3**, **C4**, **C5**, and **C6**). It is easy to check that it satisfies the weakened version of **C1** that is introduced below (it is identical to (1) introduced above).

C7 (Weak Independence) **For all** $a, b, c \in A$, $a \succsim b \Rightarrow a + c \succsim b + c$.

Remark 17. This above condition is identical to condition A5 in Marchant (2009a, p. 328). The interest of this weakening of **C1** has already been discussed. Since $\succsim_\tau$ and $\succsim_{c_\tau}$ satisfy **C1**, they also satisfy **C7**, whatever τ.

The following lemma extends Lemma 1 to the case of weak independence.

Lemma 18. *If a bibliometric ranking $\succsim$ satisfies* **C7**, *then we have $\mathbf{1}_x \succ \mathbf{0}$, for some $x \in \mathbb{N}$.*

Proof. Suppose that the thesis is violated, so that $\mathbf{0} \succsim \mathbf{1}_x$, for all $x \in \mathbb{N}$. Using **A2**, we must have $\mathbf{1}_x \sim \mathbf{0}$, for all $x \in \mathbb{N}$.

Any non-null $a \in A$ can be written as the sum of single paper authors. Since $\mathbf{1}_x \sim \mathbf{0}$, for all $x \in \mathbb{N}$, it is easy to see that repeated applications of **C7** imply that $a \sim \mathbf{0}$. This contradicts **A1**. □

The following condition says that for authors having only one paper, citations always have a positive effect. It is clearly satisfied by $\succsim_M$. It seems innocuous as soon as the idea that citations are a signal of quality is accepted. It is violated by $\succsim_0$ and $\succsim_\tau$, whatever τ. When $\tau > 1$, it is also violated by $\succsim_{c_\tau}$ and $\succsim_{t_\tau}$. It is satisfied by $\succsim_c$.

C8 (Strict Monotonicity) *For all $x, y \in \mathbb{N}$ with $x > y$, $\mathbf{1}_x \succ \mathbf{1}_y$.*

The next condition says that a quasi-null author with a single paper should not be distinguished from the null author. It is clearly satisfied by $\succsim_M$. It seems rather innocuous. It is violated by $\succsim_0$. It is satisfied by $\succsim_\tau$ when $\tau > 0$. It is satisfied by $\succsim_{c_\tau}$, whatever τ, and, hence, by $\succsim_c$. We already know that it is satisfied by $\succsim_{t_\tau}$, whatever τ.

C9 (Quasi-null authors) $\mathbf{1}_0 \sim \mathbf{0}$.

The following condition says that having two papers with the same number of citations does not have a positive impact on the ranking. It is clearly satisfied by $\succsim_M$.

C10 (One Plus One Equals One) *For all $x \in \mathbb{N}$, $\mathbf{1}_x + \mathbf{1}_x \sim \mathbf{1}_x$.*

Remark 18. The above condition is identical to condition A9 in Marchant (2009a, p. 332). It is clearly satisfied by $\succsim_M$. As discussed in Marchant (2009a), this condition says that for an author having a single publication, publishing another paper with exactly the same number of citations has no impact. It favors quality (in terms of citations) over quantity (in terms of papers). It does so in a very strong way. It is easy to check that it is violated by $\succsim_0$ and $\succsim_c$. It is also violated by $\succsim_\tau$, $\succsim_{c_\tau}$, and $\succsim_{t_\tau}$, whatever $\tau > 0$.

The following lemmas will be useful.

Lemma 19. *If a bibliometric ranking $\succsim$ satisfies* **C7**, **C8**, **C9**, *and* **C10**, *then for all $x, y \in \mathbb{N}$ such that $x \geq y$, $\mathbf{1}_x \sim \mathbf{1}_x + \mathbf{1}_y$.*

Proof. If $y = 0$, **C9** implies that $\mathbf{1}_y \sim \mathbf{0}$ and the conclusion follows using **C7**. Suppose henceforth that $y > 0$. Using **C8** and **C9**, we know that $\mathbf{1}_y \succ \mathbf{1}_0 \sim \mathbf{0}$, so that $\mathbf{1}_y \succ \mathbf{0}$. Using **C7**, we obtain $\mathbf{1}_y + \mathbf{1}_x \succsim \mathbf{0} + \mathbf{1}_x = \mathbf{1}_x$. Conversely, using **A2**, we know that $\mathbf{1}_x \succsim \mathbf{1}_y$. Using **C7**, we obtain $\mathbf{1}_x + \mathbf{1}_x \succsim \mathbf{1}_y + \mathbf{1}_x$. Using **C10**, we know that $\mathbf{1}_x + \mathbf{1}_x \sim \mathbf{1}_x$, so that $\mathbf{1}_x \succsim \mathbf{1}_x + \mathbf{1}_y$. Hence, we have $\mathbf{1}_x \sim \mathbf{1}_x + \mathbf{1}_y$. □

Lemma 20. *If a bibliometric ranking $\succsim$ satisfies* **C7**, **C8**, **C9**, *and* **C10**, *then, for all $n \in \mathbb{N}$ and all $x, y_1, y_2, \ldots, y_n \in \mathbb{N}$ such that $x \geq y_i$, for $i = 1, 2, \ldots, n$, $\mathbf{1}_x \sim \mathbf{1}_x + \mathbf{1}_{y_1} + \mathbf{1}_{y_2} + \cdots + \mathbf{1}_{y_n}$.*

Proof. Using Lemma 19, we know that $\mathbf{1}_x \sim \mathbf{1}_x + \mathbf{1}_{y_1}$. Using **C7**, we obtain $\mathbf{1}_x + \mathbf{1}_{y_2} \sim \mathbf{1}_x + \mathbf{1}_{y_1} + \mathbf{1}_{y_2}$. Using Lemma 19 again, we know that $\mathbf{1}_x \sim \mathbf{1}_x + \mathbf{1}_{y_2}$, so that $\mathbf{1}_x \sim \mathbf{1}_x + \mathbf{1}_{y_1} + \mathbf{1}_{y_2}$. Repeating the reasoning leads to the desired conclusion. □

Lemma 21. *If a bibliometric ranking $\succsim$ satisfies* **C7** *and* **C10**, *then, for all $x \in \mathbb{N}$ and all $y \in \mathbb{N}_+$, $y\mathbf{1}_x \sim \mathbf{1}_x$.*

Proof. Using **C10**, we know that, $\mathbf{1}_x + \mathbf{1}_x = 2\mathbf{1}_x \sim \mathbf{1}_x$. Using **C7**, we have $2\mathbf{1}_x + \mathbf{1}_x \sim \mathbf{1}_x + \mathbf{1}_x$. Hence, we obtain $3\mathbf{1}_x \sim 2\mathbf{1}_x \sim \mathbf{1}_x$. Repeating the above reasoning leads to the desired conclusion. □

8.3. Results

Our first result, in this section, characterizes $\succsim_M$.

Theorem 11. *A bibliometric ranking $\succsim$ satisfies* **C7**, **C8**, **C9**, *and* **C10** *iff, for all $a, b \in A$, $a \succsim b \Leftrightarrow f_M(a) \geq f_M(b)$.*

Proof. Necessity is clear. We show sufficiency. Let us first deal with the case of authors that have a null value for the index f_M. Such authors are either null or quasi-null. Using **C9** and Lemma 21, we know that, for all $y \in \mathbb{N}_+$, $y\mathbf{1}_0 \sim \mathbf{1}_0 \sim \mathbf{0}$. Hence all authors having a null value for the index f_M are tied in the ranking. Consider now an author $a \in A$ such that $f_M(a) = k > 0$. Using Lemmas 20 and 21, we know that $a \sim \mathbf{1}_k$. Using **C8** and **C9**, we know that $\mathbf{1}_k \succ \mathbf{1}_0 \sim \mathbf{0}$. Hence all authors having a strictly positive value for the index f_M are ranked above all authors having a null value for the index f_M.

Consider now two authors $a, b \in A$ such that $f_M(a) = k$ and $f_M(b) = k'$ with $k, k' > 0$. Using Lemmas 20 and 21, we know that $a \sim \mathbf{1}_k$ and $b \sim \mathbf{1}_{k'}$. If $k = k'$, this implies $a \sim b$. Suppose that $k > k'$. Using **C8**, we know that $\mathbf{1}_k \succ \mathbf{1}_{k'}$, so that $a \succ b$, which completes the proof. □

Remark 19. The above theorem is similar to Marchant (2009a) with **A2** instead of Lower Bound and CDNH. Observe however that Marchant (2009a, Th. 3) is not entirely correct as it is stated. A counter-example is the ranking with two equivalence classes with **0** being the only element of the last equivalence class. The problem comes from the formulation of Uniformity in Marchant (2009a) (saying that, for all $x \in \mathbb{N}_+$, $\mathbf{1}_x \succ \mathbf{1}_{x-1}$ iff $\mathbf{1}_{x+1} \succ \mathbf{1}_x$). In the proof of Theorem 3 in Marchant (2009a) it is said that Lemma 2 in Marchant (2009a) (saying that there is $x \in \mathbb{N}$ such that $\mathbf{1}_x \succ \mathbf{0}$) and Uniformity imply that $x > y \Leftrightarrow \mathbf{1}_x \succ \mathbf{1}_y$. This is not correct (a counterexample is given by a ranking in which all $\mathbf{1}_x$ are indifferent and strictly above **0**).

We have corrected the above problem replacing the Uniformity condition in Marchant (2009a) by our condition **C8**.

Bringing condition ORI and ES into the picture leads to a characterization of the index f_M up to the multiplication by a positive constant. We start with a simple lemma.

Lemma 22. *If a bibliometric index f satisfies conditions* **C7**, **C8**, **C9**, **C10**, *ORI, and ES, then, for all $x \in \mathbb{N}$, $f(\mathbf{1}_x) = xf(\mathbf{1}_1)$.*

Proof. Using Theorem 11 and ORI, we know that $f(\mathbf{1}_1) > f(\mathbf{1}_0) = f(\mathbf{0}) = 0$. Using Theorem 11, it is clear that $\mathbf{1}_1$ is immediately above **0** and that, for all $y \in \mathbb{N}_+$, $\mathbf{1}_{y+1}$ is immediately above $\mathbf{1}_y$. Given that ORI holds, repeated applications of ES imply, for all $x \in \mathbb{N}_+$, $f(\mathbf{1}_x) = xf(\mathbf{1}_1)$. □

Theorem 12. *A bibliometric index f satisfies* **C7**, **C8**, **C9**, **C10**, *ORI, and ES iff, for all $a \in A$, $f(a) = \beta f_M(a)$, for some $\beta \in \mathbb{R}_+$.*

Proof. Necessity is clear. Using Theorem 11 and ORI, we know that $f(y\mathbf{1}_0) = f(\mathbf{1}_0) = f(\mathbf{0}) = 0$. The claim therefore holds for null and quasi-null authors.

Let $f(\mathbf{1}_1) = \beta$. Using Theorem 11 and ORI, we know that $\beta > 0$. Consider now an author $a \in A$ that is strictly non-null. Suppose that $f_M(a) = k > 0$. Using Lemmas 20 and 21, we know that $f(a) = f(\mathbf{1}_k)$. Using Lemma 22, we obtain $f(a) = f(\mathbf{1}_k) = kf(\mathbf{1}_1) = \beta k = \beta f_M(a)$. □

8.4. Independence of conditions

We show below that the conditions used in Theorem 12 are independent.

Example 10 *(***C7** *Weak Independence).* Let $a \in A$ be a strictly non-null author and let $f_M(a) = k_a$. Clearly we can always find $x \in \mathbb{N}_+$ such that $x\mathbf{1}_{k_a} \trianglerighteq a$. Let x_a be the smallest integer such that $x_a\mathbf{1}_{k_a} \trianglerighteq a$.

Let f be the index such that:

$$f(a) = \begin{cases} 0 & \text{if} \quad a \,\text{is null or quasi} - \text{null}, \\ k_a & \text{if} \quad x_a \leq 2, \\ k_a + 1 & \text{if} \quad x_a > 2. \end{cases}$$

It is simple to check that **A1** and **A2** hold so that this index is a bibliometric index.

Conditions **C9** and ORI trivially hold. It is not difficult to check that ES holds. For all $x \in \mathbb{N}$, we have $f(\mathbf{1}_x) = x$. This shows that **C8** holds. For all $x \in \mathbb{N}$, we have $f(2\mathbf{1}_x) = f(\mathbf{1}_x)$, which shows that **C10** holds.

Condition **C7** is violated. For instance, we have $f(2\mathbf{1}_3) = 3 = f(\mathbf{1}_3)$ and $f(3\mathbf{1}_3) = 4 > f(2\mathbf{1}_3)$.

Example 11 *(***C8** *Strict Monotonicity).* Consider the bibliometric index f that is equal to 0 for all authors having at most one paper, and equal to f_M otherwise.

It is simple to check that **A1** and **A2** hold so that this index is a bibliometric index. This index clearly violates **C8**. It is easy to check that all other conditions are satisfied.

Example 12 *(**C9** Quasi-null authors).* Consider the index f that is equal to $f_M + 1$ whenever an author is strictly non-null. We let $f(a) = 1$, for all quasi-null authors $a \in A$ and $f(\mathbf{0}) = 0$.

It is simple to check that **A1** and **A2** hold so that this index is a bibliometric index. Conditions **C7**, **C8**, **C10**, ORI and ES clearly hold. Condition **C9** is violated.

Example 13 *(**C10** OPOEO).* The index f_c clearly violates **C10**. All other conditions (**A1**, **A2**, **C7**, **C8**, and **C9**) clearly hold.

Rephrasing the above examples in terms of the bibliometric ranking induced by the bibliometric index shows that the conditions used in Theorem 11 are independent. The standard examples show that ORI and ES cannot be omitted in Theorem 12.

9. Hirsch index (*h*-index)

This index was proposed by Hirsch (2005) and has been at the forefront of research on bibliometric indices ever since.

9.1. Additional notation

Consider a non-null author $a \in A$, so that $p_a \geq 1$. We index the p_a papers of author a as $1, 2, \ldots, p_a$. Let x_i^a be the number of citations received by the ith paper of author a.

Using the above notation, we can view the non-null author $a \in A$ as the following sum of single paper authors as follows:

$$a = \mathbf{1}_{x_1^a} + \mathbf{1}_{x_2^a} + \cdots + \mathbf{1}_{x_{p_a}^a}.$$

Note that in the above expression, we can always choose to order the integers $x_1^a, x_2^a, \ldots, x_{p_a}^a$ in decreasing order. When we want to emphasize this ordering, we will write an author $a \in A$ as

$$a = \mathbf{1}_{x_{(1)}^a} + \mathbf{1}_{x_{(2)}^a} + \cdots + \mathbf{1}_{x_{(p_a)}^a},$$

with the convention that:

$$x_{(1)}^a \geq x_{(2)}^a \geq \ldots \geq x_{(p_a)}^a.$$

Hence, for a non-null author $a = \mathbf{1}_{x_{(1)}^a} + \mathbf{1}_{x_{(2)}^a} + \cdots + \mathbf{1}_{x_{(p_a)}^a} \in A$, $x_{(1)}^a$ is the number of citations of her most cited paper and $x_{(p_a)}^a$ is the number of citations of her least cited paper. It will be convenient to suppose that, for a null author $a \in A$, $x_{(1)}^a = x_{(p_a)}^a = 0$.

Let $x \in \mathbb{N}_+$. An author with x publications each of them having been cited x times is written as $x\mathbf{1}_x$. We also say that this author is a *square author* of size x. Whereas square authors played no particular rôle for the analysis of the rankings and indices studied so far, they will play a central one for the analysis of the h-index and of the g-index in the next section.

Observe that an author that is strictly non-null always dominates the square author of size 1.

9.2. Setting

Our definition of the h-index is different from the one usually found in the literature. It is however easy to see that our definition is equivalent to the usual one.

Consider an author $a \in A$. If a is null or quasi-null she has a Hirsch (2005) index (henceforth h-index) of 0. Otherwise $a \in A$ is strictly non-null, so that $p_a > 0$ and $c_a > 0$. We know that a dominates the square author of size 1. The h-index of a strictly non-null author a is equal to $k \in \mathbb{N}_+$ if $a \trianglerighteq k\mathbf{1}_k$ and $a \ntrianglerighteq (k+1)\mathbf{1}_{k+1}$. Summarizing, we have:

$$f_h(a) = \begin{cases} 0 & \text{if} \quad a \text{ is null of quasi null,} \\ k & \text{if} \quad a \trianglerighteq k\mathbf{1}_k \text{ and } a \ntrianglerighteq (k+1)\mathbf{1}_{k+1} \text{ otherwise.} \end{cases}$$

When $f_h(a) = k > 0$, any set of k papers having received at least k citations is called the *h-core* of $a \in A$.

Hence, for a strictly non-null author $a \in A$, an h-index of k means that she has k of her publications that have been cited at least k times (i.e., have a number of citations that is greater than or equal to k) and $n - k$ of her publications have been cited at most k times (i.e., have a number of citations that is less than or equal to k). If author $a \in A$ is null or quasi-null, her h-index is 0. The bibliometric index associating with each author her h-index is denoted by f_h. We will also be interested in the bibliometric ranking $\succsim_h$ induced by f_h.

It is possible to give a nice geometric interpretation of the h-index (see, e.g., Gagolewski & Grzegorzewski, 2009; Liu, Zuo, Goa, & Qian, 2013; Woeginger, 2008a). Let $a = \mathbf{1}_{x_{(1)}^a} + \mathbf{1}_{x_{(2)}^a} + \cdots + \mathbf{1}_{x_{(p_a)}^a}$. We associate with each of the papers of a, starting with the most cited one, a vertical bar the height of which is equal to the number of citations received (see Fig. 1). Hence, each $a \in A$ can be viewed as a collection of bars in a two-dimensional plane. With this representation in mind, an author $a \in A$ is such that $f_h(a) \geq k$ if the collection of bars associated with a in this representation lies “above” a square of size k.

It is easy to check that $\succsim_h$ satisfies **A1** and **A2**, so that it is a bibliometric ranking.

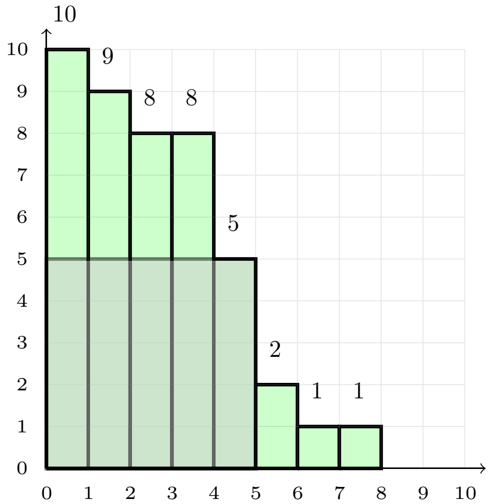

Fig. 1. Geometrical interpretation of the h-index. The author $a = \mathbf{1}_{10} + \mathbf{1}_9 + \mathbf{1}_8 + \mathbf{1}_8 + \mathbf{1}_5 + \mathbf{1}_2 + \mathbf{1}_1 + \mathbf{1}_1$ has $f_h(a) = 5$.

With our conventions regarding $x^a_{(1)}$ and $x^a_{(p_a)}$ for null authors (they are both equal to 0), it can easily be checked, as observed by Quesada (2009), that for any $a \in A$, we have

$$\min(p\prime_a, x^a_{(p_a)}) \le f_h(a) \le \min(p_a, x^a_{(1)}).$$

Suppose that an author $a \in A$ is such that $f_h(a) = k \ge 1$. It is easy to check that adding to a a single paper with at most k citations cannot improve her h-index. It may also be instructive to observe that, for all $a \in A$ and all $x \in \mathbb{N}_+$, $f_h(a) + 1 \ge f_h(a + \mathbf{1}_x)$, i.e., the addition of a single paper can at most increase the h-index by 1.

It is well-known (Bouyssou & Marchant, 2011b; Marchant, 2009a; Waltman, Costas, & van Eck, 2012; Waltman & van Eck, 2009a, 2009b, 2012 that the ranking $\succsim_h$ violates independence (**C1**) and even weak independence (**C7**). For instance, we have $f_h(5\mathbf{1}_6) = f_h(5\mathbf{1}_5) = 5$, while $f_h(5\mathbf{1}_6 + \mathbf{1}_7) = 6 > f_h(5\mathbf{1}_5 + \mathbf{1}_7) = 5$. This explains why the analysis of this ranking will require conditions that are rather different from the ones used so far.

It is easy to check that f_h satisfies **C2**, **C6**, and **C9**. It violates **C3** (since $f_h(\mathbf{1}_1) = 1 > f_h(\mathbf{1}_0) = 0$), **C4** (since $f_h(\mathbf{1}_2) = 1 > 0$, while $f_h(2\mathbf{1}_2) = 2 > f_h(\mathbf{1}_4) = 1$), **C8** (since $f_h(\mathbf{1}_2) = f_h(\mathbf{1}_1) = 1$), and **C10** (since $f_h(\mathbf{1}_2) = 1$, while $f_h(2\mathbf{1}_2) = 2$).

9.3. Conditions

Our analysis is inspired by Quesada (2009). Because $\succsim_h$ does not satisfy **C7** (and, hence, **C1**), many new conditions will be needed.

C11 (Strong quasi-null authors) *For all* $x \in \mathbb{N}_+$, $x\mathbf{1}_0 \sim \mathbf{0}$.

The above condition is implied by condition A1 in Quesada (2009). It is a strengthening of **C9**. It asserts that all quasi-null authors (i.e., authors of the type $x\mathbf{1}_0$ with $x \in \mathbb{N}_+$), should not be distinguished from $\mathbf{0}$. This condition is clearly satisfied by f_h. Indeed, $f_h(a) = 0$ implies that a is null or quasi-null. It is easy to check that this condition is satisfied by $\succsim_{c_\tau}$ and $\succsim_{t_\tau}$, whatever τ, and by $\succsim_M$. It is clearly violated by $\succsim_0$. It is satisfied by $\succsim_\tau$, when $\tau > 0$.

C12 (Tail independence) *Let* $a, b \in A$ *be strictly non-null authors. Suppose that* $p_a = p_b = n$ *and that* $a = \mathbf{1}_{x^a_{(1)}} + \mathbf{1}_{x^a_{(2)}} + \cdots + \mathbf{1}_{x^a_{(n)}}$ *and* $b = \mathbf{1}_{x^b_{(1)}} + \mathbf{1}_{x^b_{(2)}} + \cdots + \mathbf{1}_{x^b_{(n)}}$. *Let* $z \in \mathbb{N}$ *be such that* $z \le x^a_{(n)}$ and $z \le x^b_{(n)}$. *Then*

$$a \sim b \Rightarrow [a + \mathbf{1}_z] \sim [b + \mathbf{1}_z].$$

The above condition is a variation on condition A3 in Quesada (2009). It has the flavor of an independence condition and is clearly implied by **C7** and, hence, **C1**. Suppose that two authors a and b have the same number of papers. Suppose furthermore that these two authors have at least one paper with at least one citation. Suppose finally that these two authors are equally ranked. The condition relates the performance of a and b after they both publish an additional paper with the same number of citations. If this number of citations of this additional paper is at most equal to the minimum of the number of citations of the least cited paper of a and the number of citations of the least cited paper of b, the modified authors are still equally ranked.

Let us show that this condition is satisfied by $\succsim_h$. Suppose that a and b have at least one paper with at least one citation. Suppose furthermore that $p_a = p_b = n$ and $f_h(a) = f_h(b) = k$. We clearly have $n \ge k \ge 1$. Suppose first that $n > k$. Because $f_h(a) = f_h(b) = k < n$, we know that $x^a_{(n)} \le k$ and $x^b_{(n)} \le k$. Adding to both a and b a paper $\mathbf{1}_z$ with $z \le k$ cannot modify their h-index. The condition therefore holds in this case and we have $a \sim_h b \sim_h a + \mathbf{1}_z \sim_h b + \mathbf{1}_z$. Suppose now that $n = k$. We have

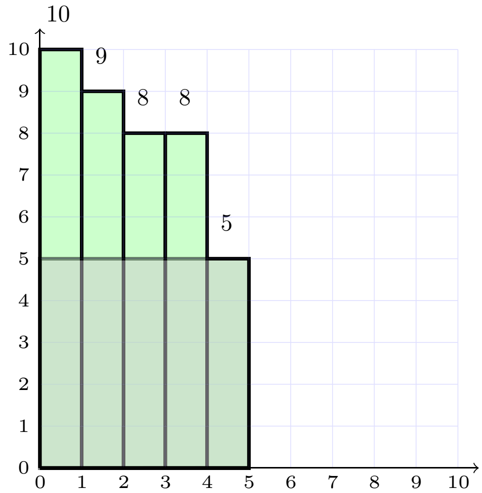

Fig. 2. Geometrical interpretation of **C13** Square upwards. We have $5\mathbf{1}_5 \succsim \mathbf{1}_{10} + \mathbf{1}_9 + \mathbf{1}_9 + \mathbf{1}_8 + \mathbf{1}_8 + \mathbf{1}_5$.

$x^a_{(n)} \geq k$ and $x^b_{(n)} \geq k$. Let $\ell = \min(x^a_{(n)}, x^b_{(n)}) \geq k$. Let us analyze the impact of the addition to both a and b of a paper $\mathbf{1}_z$ with $z \leq \ell$. We distinguish two cases. If $z \leq k$, this addition cannot improve the h-index of a and b. The condition therefore holds and we have $a \sim_h b \sim_h a + \mathbf{1}_z \sim_h b + \mathbf{1}_z$. If $\ell \geq z > k$, after the addition of $\mathbf{1}_z$, the h-index of both a and b increases by 1. The condition therefore holds and we have $f_h(a) = f_h(b)$, $f_h(a + \mathbf{1}_z) = f_h(a) + 1$ and $f_h(b + \mathbf{1}_z) = f_h(b) + 1$, so that $a + \mathbf{1}_z \sim_h b + \mathbf{1}_z$.

It is easy to check that **C12** is satisfied by $\succsim_\tau$ (and, hence, $\succsim_0$), $\succsim_{c_\tau}$, $\succsim_{t_\tau}$ (and, hence, $\succsim_c$) and $\succsim_M$ (since **C7**, and, hence, **C1**, implies **C12**).

C13 (Square upwards) *Let* $x \in \mathbb{N}_+$. *Let* $y_1, y_2, \ldots, y_x \in \mathbb{N}$ *such that* $y_i \geq x$, for $i = 1, 2, \ldots, x$. *Then*

$$x\mathbf{1}_x \succsim [\mathbf{1}_{y_1} + \mathbf{1}_{y_2} + \cdots \mathbf{1}_{y_x}].$$

The above condition is satisfied by f_h since we have $f_h(x\mathbf{1}_x) = x$ and $f_h(\mathbf{1}_{y_1} + \mathbf{1}_{y_2} + \cdots + \mathbf{1}_{y_x}) = x$. It is a variation on condition A1 in Quesada (2009). It says that a square author cannot improve her position by just collecting more citations. This is a strong condition. It is illustrated in Fig. 2.

Although **C13** may seem to be quite specific to $\succsim_h$, it is also satisfied by $\succsim_0$. It is easy to check that it is violated by $\succsim_c$ and by $\succsim_M$. It is also violated by $\succsim_\tau$ (whatever τ), $\succsim_{c_\tau}$, and $\succsim_{t_\tau}$ (with $\tau > 1$).

C14 (Square rightwards) *Let* $x, j \in \mathbb{N}_+$. *Let* $y_1, y_2, \ldots, y_j \in \mathbb{N}$ *be such that* $y_i \leq x$, for $i = 1, 2, \ldots, j$. *Then*

$$x\mathbf{1}_x \succsim [x\mathbf{1}_x + \mathbf{1}_{y_1} + \mathbf{1}_{y_2} + \cdots + \mathbf{1}_{y_j}].$$

This condition is satisfied by $\succsim_h$ since we have $f_h(x\mathbf{1}_x) = x$ and $f_h(x\mathbf{1}_x + \mathbf{1}_{y_1} + \mathbf{1}_{y_2} + \cdots + \mathbf{1}_{y_j}) = x$. This condition is also a variation on condition A1 in Quesada (2009). It says that adding publications to a square author of size x cannot improve the index when each of these publications are cited at most x times. This is also a strong condition. It is illustrated in Fig. 3. A consequence of this condition is that a square author cannot improve her position by just publishing new papers that receive a low number of citations.

Although **C14** may seem to be quite specific to $\succsim_h$, it is also clearly also satisfied by $\succsim_M$. It is violated by $\succsim_0$ and $\succsim_c$. It is also violated by $\succsim_\tau$ (with $\tau > 0$) and by $\succsim_{c_\tau}$ and $\succsim_{t_\tau}$ (with $\tau > 1$).

C15 (Strong Uniformity) *For all* $x, y \in \mathbb{N}$,

$$(x+1)\mathbf{1}_{x+1} \succ x\mathbf{1}_x \Leftrightarrow (y+1)\mathbf{1}_{y+1} \succ y\mathbf{1}_y.$$

Together with **A2**, the above condition implies that, if $1\mathbf{1}_1 = \mathbf{1}_1 \sim 0\mathbf{1}_0 = \mathbf{0}$, then all square authors will belong to the same equivalence class of the bibliometric ranking. Conversely, if $1\mathbf{1}_1 = \mathbf{1}_1 \succ 0\mathbf{1}_0 = \mathbf{0}$, then we have, for all $x \in \mathbb{N}_+$ $(x+1)\mathbf{1}_{x+1} \succ x\mathbf{1}_x$. Condition **C15** is clearly satisfied by f_h since we have $f_h(x\mathbf{1}_x) = x$.

It is easy to check that **C15** is satisfied by $\succsim_0$, $\succsim_c$ and $\succsim_M$. It is easy to check that it is violated by $\succsim_\tau$ (with $\tau > 0$) and by $\succsim_{c_\tau}$ and $\succsim_{t_\tau}$ (with $\tau > 1$).

The following lemmas will be useful.

Lemma 23. *If a bibliometric ranking* $\succsim$ *satisfies* **C11** *and* **C15**, *then* $\mathbf{1}_1 \succ \mathbf{1}_0 \sim \mathbf{0}$.

Proof. Using **C11**, we know that $\mathbf{1}_0 \sim \mathbf{0}$. Suppose now that $\mathbf{1}_0 \succsim \mathbf{1}_1$. Using **A2**, we must have $\mathbf{1}_1 \sim \mathbf{1}_0$. Using **A2** and **C15**, this implies that, for all $x \in \mathbb{N}_+$, all square authors $x\mathbf{1}_x$ belong to the same indifference class as $\mathbf{0}$. Take any $a \in A$. Take any $z \in \mathbb{N}_+$ such that $z\mathbf{1}_z \trianglerighteq a$. Using **A2**, we know that $z\mathbf{1}_z \succsim a$. Since we know that $\mathbf{1}_1 \sim \mathbf{0}$, **A2** and **C15** imply that $(x+1)\mathbf{1}_{x+1} \sim x\mathbf{1}_x$, for all $x \in \mathbb{N}_+$. Hence, we have $z\mathbf{1}_z \succsim a$ and $z\mathbf{1}_z \sim \mathbf{0}$, so that $\mathbf{0} \succsim a$. Using **A2**, this implies $a \sim \mathbf{0}$, for all $a \in A$. This violates **A1**.□

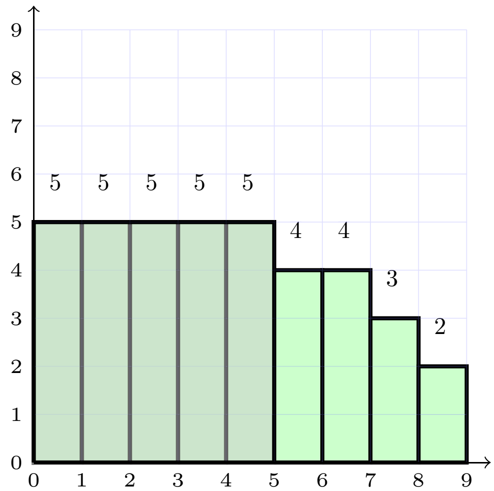

Fig. 3. Geometrical interpretation of **C14** Square rightwards. We have $5\mathbf{1}_5 \succsim 5\mathbf{1}_5 + \mathbf{1}_4 + \mathbf{1}_4 + \mathbf{1}_3 + \mathbf{1}_2$.

Lemma 24. *If a bibliometric ranking $\succsim$ satisfies* **C11**, **C12**, **C13**, *and* **C14**, *then, for all $a \in A$, $f_h(a) = k > 0$ implies $a \sim k\mathbf{1}_k$.*

Proof. Let $a \in A$ be such that $f_h(a) = k > 0$. This implies that $p_a \geq k$. Suppose that

$$a = \mathbf{1}_{x^a_{(1)}} + \mathbf{1}_{x^a_{(2)}} + \cdots + \mathbf{1}_{x^a_{(p_a)}}.$$

Since $f_h(a) = k$, we know that $x^a_{(i)} \geq k$, for $i = 1, 2, \ldots, k$ and that $x^a_{(k+j)} \leq k$ for $j = 1, 2, \ldots, p_a - k$.

On the basis of a we define several other authors. For $i = k, k + 1, \ldots, p_a$, let a^{*i} be an author with i publications such that, if $i = k$,

$$a^{*i} = i\mathbf{1}_k = k\mathbf{1}_k,$$

and, if $i > k$:

$$a^{*i} = k\mathbf{1}_k + \mathbf{1}_{x^a_{(k+1)}} + \mathbf{1}_{x^a_{(k+2)}} + \cdots + \mathbf{1}_{x^a_{(i)}}.$$

Similarly, for $i = k, k + 1, \ldots, p_a$, let a^i be an author with i publications such that

$$a^i = \mathbf{1}_{x^a_{(1)}} + \mathbf{1}_{x^a_{(2)}} + \cdots + \mathbf{1}_{x^a_{(i)}}.$$

For $i = k, k + 1, \ldots, p_a$ consider the authors a^i and a^{*i}. They both have the same number of papers (i.e., $p_{a^i} = p_{a^{*i}}$). Both have at least one paper with at least one citation. Moreover, we have $a^i \trianglerighteq a^{*i}$.

Using **A2**, we know that $a^k \succsim a^{*k}$.

Using **C13**, we know that $a^{*k} \succsim a^k$.

Hence, we know that $a^k \sim a^{*k}$. If $p_a = k$, we know that $a = a^k$, so that $a^{*k} \sim a^k \sim a$.

Suppose henceforth that $p_a > k$. Notice that, the combination of **A2** and **C14** implies that $a^{*k+j} \sim a^{*k}$, for $j = 1, 2, \ldots, p_a - k$.

We first show that $a^{k+1} \sim a^{*k+1}$. Using **A2**, we know that $a^{k+1} \succsim a^{*k+1}$. By construction, we have:

$$a^{k+1} = a^k + \mathbf{1}_{x^a_{(k+1)}},$$
$$a^{*k+1} = a^{*k} + \mathbf{1}_{x^a_{(k+1)}}.$$

Since $a^k \sim a^{*k}$, **C12** implies that $a^{k+1} \sim a^{*k+1}$. We know that $a^{*k+1} \sim a^{*k}$. Hence, we obtain $a^{k+1} \sim a^{*k+1} \sim a^{*k}$.

Suppose now that for $i = 1, 2, \ldots, j$ with $j < p_a - k$, we have $a^{k+i} \sim a^{*k+i}$. Let us show that we must also have $a^{k+j+1} \sim a^{*k+j+1}$. By construction, we have:

$$a^{k+j+1} = a^{k+j} + \mathbf{1}_{x^a_{(k+j+1)}},$$
$$a^{*k+j+1} = a^{*k+j} + \mathbf{1}_{x^a_{(k+j+1)}}.$$

Since $a^{k+j} \sim a^{*k+j}$, using **C12** implies that $a^{k+j+1} \sim a^{*k+j+1}$. We know that $a^{*k+j+1} \sim a^{*k}$. Hence, we obtain $a^{k+j+1} \sim a^{*k+j+1}$.

The above argument shows that, for all $a \in A$ such that $f_h(a) = k > 0$, we have $a \sim k\mathbf{1}_k$. □

9.4. Results

Our first result, in this section, characterizes $\succsim_h$.

Theorem 13. *A bibliometric ranking $\succsim$ satisfies conditions* **C11**, **C12**, **C13**, **C14**, *and* **C15**, *iff, for all $a, b \in A$, $a \succsim b \Leftrightarrow f_h(a) \geq f_h(b)$.*

Proof. Necessity is clear. We show sufficiency.

Suppose that a bibliometric ranking $\succsim$ satisfies **C11**, **C12**, **C13**, **C14**, and **C15**. We have to show that, for all a, $b \in A$ $a \succsim b \Leftrightarrow f_h(a) \geq f_h(b)$.

If $f_h(a) = f_h(b) = 0$, then a and b are either null or quasi-null. The conclusion follows from **C11**. If $f_h(a) = k > 0$ and $f_h(b) = 0$, using **C11** and Lemma 24, we know that $a \sim k\mathbf{1}_k$ and $b \sim \mathbf{0}$. The conclusion follows from **A2** and Lemma 23. If $f_h(a) = k > 0$ and $f_h(b) = \ell > 0$. Using Lemma 24, we know that $a \sim k\mathbf{1}_k$ and $b \sim \ell\,\mathbf{1}_\ell$. If $k = \ell$, the conclusion follows. Suppose that $k > \ell$. We know from Lemma 23 that $\mathbf{1}_1 \succ \mathbf{1}_0$. Using **C15**, we obtain, for all $x \in \mathbb{N}_+$, $(x+1)\mathbf{1}_{x+1} \succ x\mathbf{1}_x$. Hence, we have $k\mathbf{1}_k \succ \ell\,\mathbf{1}_\ell$. This completes the proof. □

To our knowledge, Theorem 13 is the only characterization of $\succsim_h$, together with Marchant (2009, Th. 5, p. 336). We use conditions that are much simpler than the ones used in Marchant (2009a). Our proof is shorter.

Remark 20. Going through the proof of Theorem 13 shows that **C12** is only used when one of the two authors a and b is the square author of size n. Condition **C12** could therefore be weakened in this way.

We also give a characterization of the index f_h up to the multiplication by a positive constant. As shown above, this is at variance with the previous characterizations of f_h (Hwang, 2013; Miroiu, 2013; Quesada, 2009, 2010, 2011a, 2011b; Woeginger, 2008a, 2008b). Moreover our conditions do not suffer from the problems discussed in Section 1 since they never explicitly refer to the value of the index f.

It is clear that the index f_h satisfies conditions ORI and ES. The following lemma will be useful.

Lemma 25. *A bibliometric index f satisfies conditions* **C11**, **C12**, **C13**, **C14**, **C15**, *ORI, and ES, then, for all $x \in \mathbb{N}$, $f(x\mathbf{1}_x) = xf(\mathbf{1}_1)$*

Proof. Using Theorem 13, we know that $1\mathbf{1}_1$ is immediately above $\mathbf{0}$. Moreover, for all $y \in \mathbb{N}_+$, $(y+1)\mathbf{1}_{y+1}$ is immediately above $y\mathbf{1}_y$. Given ORI, the thesis follows from ES. □

Theorem 14. *A bibliometric index f satisfies conditions* **C11**, **C12**, **C13**, **C14**, **C15**, *ORI, and ES, iff, for all $a \in A$, $f(a) = \beta f_h(a)$, for some $\beta \in \mathbb{R}_+$.*

Proof. Necessity was shown above. We concentrate below on sufficiency. Suppose that a bibliometric index f satisfies **C11**, **C12**, **C13**, **C14**, ORI, and ES. We have to show that $f = \beta f_h$, for some $\beta \in \mathbb{R}_+$.

Using Lemma 23, we know that $f(\mathbf{1}_1) > 0$ and $f(\mathbf{0}) = 0$. Define $\beta = f(\mathbf{1}_1) > 0$. We will show that, for all $a \in A$, $f(a) = \beta f_h(a)$.

Suppose that $f_h(a) = 0$, so that a is null or quasi-null. The conclusion follows from ORI and **C11**.

Suppose that $f_h(a) = k > 0$. Using Lemma 24, we have $f(a) = f(k\mathbf{1}_k)$. Using Lemma 25, we know that $f(k\mathbf{1}_k) = kf(\mathbf{1}_1)$. Hence, the conclusion follows. □

9.5. Independence of conditions

Let us show that none of the conditions used in Theorem 14 is redundant.

Example 14 *(C11 Strong quasi-null authors).* Consider the bibliometric index f such that, for all $a \in A$,

$$f(a) = \begin{cases} 0 & \text{if } a \text{ is null or if } a = \mathbf{1}_0, \\ 1 & \text{if } a = x\mathbf{1}_0 \text{ with } x > 0, \\ f_h(a) + 1 & \text{otherwise,} \end{cases}$$

This index clearly violates **C11**. It is simple to check that all other conditions are satisfied.

Observe that conditions **C12**, **C13** and **C14** all involve constraints on the value of f for authors having a strictly positive h-index. Hence, they are all satisfied. Conditions **A1**, **A2**, **C15**, ORI, and ES clearly hold.

Example 15 *(C12 Tail independence).* Consider the bibliometric index such that, for all $a \in A$, $f(a) = \min(p_a, x^a_{(1)}) = \min(p_a, f_M(a))$ (using the convention that, for a null author a, $x^a_{(1)} = 0$). As observed in Quesada (2009), this index violates **C12** since, for instance, $f(3\mathbf{1}_8) = f(3\mathbf{1}_3) = 3$, while $f(3\mathbf{1}_8 + \mathbf{1}_1) = 4 > f(3\mathbf{1}_3 + \mathbf{1}_1) = 3$. It is easy to check that all other conditions are satisfied. Indeed, **A1**, **A2**,**C11**, ORI, and ES are clearly satisfied. The value of the index for a square author of size x is x, so that **C15** holds. It is simple to check that **C13** holds since all authors involved in this condition have a value x for this index. The same is true for **C14**.

Example 16 *(C13 Square upwards).* Consider the bibliometric index such that, for all $a \in A$, $f(a) = f_M(a) = x^a_{(1)}$ (using the convention that, for a null author a, $x^a_{(1)} = 0$).

Condition **C13** is clearly violated. It is easy to check that all other conditions are satisfied. Indeed, **A1**, **A2**, **C11**, ORI, and ES are clearly satisfied. The value of the index for a square author of size x is x, so that **C15** holds. Clearly the addition of a paper that is less cited than the least cited paper of an author leaves the value of this index unchanged. Hence, **C12** holds. It is simple to check that **C14** holds since all authors involved in this condition have a value x for this index.

Example 17 *(C14 Square rightwards)*. Consider the bibliometric index such that, for all $a \in A$, $f(a) = p'_a$ (i.e., the number of publications of a having received at least one citation). Condition **C14** is clearly violated. It is easy to check that all other conditions are satisfied. Indeed, **A1**, **A2**, **C11**, and ORI are clearly satisfied. The value of the index for a square author of size x is x, so that ES and **C15** holds. To check that **C12** holds, it suffices to observe that adding to any author a paper having at least one citation increases the value of the index by 1 and, moreover, that the addition of an uncited paper leaves the index unchanged.

Hence, **C12** holds. It is simple to check that **C13** holds since all authors involved in this condition have a value x for this index.

Example 18 *(C15 Strong Uniformity)*. Consider the bibliometric index f such that $f(a) = 0$ if $f_h(a) = 0$ and $f(a) = 1$ otherwise.

Conditions **A1** and **A2** are clearly satisfied. It is clear that this index violates **C15**. It is easy to check that all other conditions are satisfied. Indeed, **C11** and ORI are clearly satisfied. Condition ES trivially holds. Conditions **C12**, **C13**, and **C14** are satisfied since they all involve non-strict comparisons of strictly non-null authors.

Rephrasing the above examples in terms of the bibliometric ranking induced by the bibliometric index shows that the conditions used in Theorem 13 are independent. The standard examples show that ORI and ES cannot be omitted in Theorem 14.

10. Egghe index (g-index)

This index was proposed by Egghe (2006). It has received much attention in the literature.

10.1. Additional notation

The *Lorenz vector* associated with a non-null author $a \in A$ is a function a^L from $\mathbb{N}_+$ to $\mathbb{N}_+$ such that $a^L(1) = x^a_{(1)}$, $a^L(2) = x^a_{(1)} + x^a_{(2)}$, $a^L(3) = x^a_{(1)} + x^a_{(2)} + x^a_{(3)}, \ldots, a^L(p_a) = x^a_{(1)} + x^a_{(2)} + \cdots + x^a_{(p_a)}$. For all $i > p_a$, we let $a^L(i) = a^L(p_a)$. Hence $a^L(i)$ is the total number of citations received by the i most cited papers of $a \in A$. The link between the g-index and Lorenz vectors was stressed in Egghe (2010b, 2012, 2013) and Woeginger (2008c).

For instance for the author $a \in A$ such that

$$a = \mathbf{1}_0 + \mathbf{1}_0 + \mathbf{1}_1 + \mathbf{1}_2 + \mathbf{1}_2 + \mathbf{1}_2 + \mathbf{1}_3 + \mathbf{1}_5 + \mathbf{1}_5,$$

we have $p_a = 9$, $c_a = 20$, and $a^L = (5, 10, 13, 15, 17, 19, 20, 20, 20, 20, \ldots)$.

We say that author $a \in A$ *Lorenz dominates* author $b \in A$, which we denote by $a \trianglerighteq^L b$ whenever $a^L(i) \geq b^L(i)$, for all $i \in \mathbb{N}_+$.

Remark 21. It is not difficult to check that if $a \trianglerighteq b$, then $a \trianglerighteq^L b$. Although this fact can be easily derived from classical results on majorization (Marshall et al., 2011, chap. 7), we sketch its proof for completeness.

Remember that $a \trianglerighteq b$ means that $a^+(x) \geq b^+(x)$, for all $x \in \mathbb{N}$, where $a^+(x)$ is the number of papers published by a having received at least x citations, i.e., $a^+(x) = \sum_{i \geq x} a(i)$. Suppose that $a \trianglerighteq b$. Let $x = x^a_{(1)}$. If b has a paper with x citations, we have $a^L(1) = b^L(1)$. Otherwise, we have $a^L(1) > b^L(1)$, so that, in any case $a^L(1) \geq b^L(1)$.

Using induction, suppose that $a^L(k) \geq b^L(k)$, for $k = 1, 2, \ldots, i$. We want to prove that this implies $a^L(i+1) \geq b^L(i+1)$. Let $\alpha = x^a_{(i+1)}$ and $\beta = x^b_{(i+1)}$. If $\alpha \geq \beta$, we obtain $a^L(i+1) \geq b^L(i+1)$. Suppose that $\alpha < \beta$, so that $a(\beta) = 0$. Since $a \trianglerighteq b$, we know that $a^+(\beta) \geq b^+(\beta)$. Hence, since $a(\beta) = 0$, we have $a^+(\beta+1) > b^+(\beta+1)$, i.e., a has strictly more papers having been cited at least $\beta+1$ times than b. This clearly implies that $a^L(i+1) \geq b^L(i+1)$.

Remark 22. Observe that the addition of uncited papers to an author does not modify her Lorenz vector. Hence, in order to compare two authors a and b using their Lorenz vectors, it is not restrictive to suppose that they have published the same number of papers. If this is not the case, we may add a number of uncited papers to the authors having published the least number of papers, without modifying her Lorenz vector.

It is well-known that when a and b have received the same number of citations, the fact that $a \trianglerighteq^L b$ means that it is possible to go from b to a by a succession of elementary transformations consisting in decreasing the number of citations of a paper published by b by one unit and, simultaneously, increasing by one unit the number of citations of a paper published by b having initially received at least as many citations (Marshall et al., 2011, p. 195). Each of these elementary transformations lead to an author that is above the initial one w.r.t. $\trianglerighteq^L$.

When the total number of citations received by the papers published by b exceeds the total number of citations received by the papers published by a, it is clearly impossible to have $a \trianglerighteq^L b$ since, when i is large, $a^L(i)$ (resp. $b^L(i)$) is equal to c_a (resp. c_b).

When the total number of citations received by the papers published by a exceeds the total number of citations received by the papers published by b and $a \trianglerighteq^L b$, we can go from b to a by the following steps. We first lower the total number of citations received by the papers published by a decreasing the number of citations received by the papers published by a

having received the least number of citations. We then obtain an author a' for which we have $c_{a'} = c_b$ and such that $a \trianglerighteq^L a'$. We can now go from b to a' using the above elementary transformations (Marshall et al., 2011, p. 177).

Let us illustrate this process on a simple example. Suppose that authors a and b are such that:

$$a = \mathbf{1}_0 + \mathbf{1}_1 + \mathbf{1}_2 + \mathbf{1}_2 + \mathbf{1}_2 + \mathbf{1}_3 + \mathbf{1}_5 + \mathbf{1}_5,$$
$$b = \mathbf{1}_1 + \mathbf{1}_1 + \mathbf{1}_1 + \mathbf{1}_2 + \mathbf{1}_2 + \mathbf{1}_3 + \mathbf{1}_4 + \mathbf{1}_5.$$

We have $a \trianglerighteq^L b$ since

$$a^L = (5, 10, 13, 15, 17, 19, 20, 20, 20, \ldots) \text{ and}$$
$$b^L = (5, 9, 12, 14, 16, 17, 18, 19, 19, \ldots).$$

Observe that we have: $a^+(1) = 7$, $b^+(1) = 8$, $a^+(5) = 2$, $b^+(5) = 1$, so that these two authors related by $\trianglerighteq^L$ are not related by $\trianglerighteq$.

We have $c_a = 20$ while $c_b = 19$. We first build an author a' that has the same number of citations as b by decreasing the number of citations of a paper having been cited at least once and having received the least number of citations. Hence, we have:

$$a\prime = \mathbf{1}_0 + \mathbf{1}_0 + \mathbf{1}_2 + \mathbf{1}_2 + \mathbf{1}_2 + \mathbf{1}_3 + \mathbf{1}_5 + \mathbf{1}_5,$$

so that

$$a\prime^L = (5, 10, 13, 15, 17, 19, 19, 19, 19, \ldots).$$

Observe that we have $a \trianglerighteq^L a' \trianglerighteq^L b$. Now we can go from b to a' by, first, taking a citation from one of the papers $\mathbf{1}_1$ published by b, transforming it into a paper $\mathbf{1}_0$, and adding this extra citation to the paper $\mathbf{1}_4$, transforming it into a paper $\mathbf{1}_5$, and, second, taking a citation from one of the remaining papers $\mathbf{1}_1$ published by b, transforming it into a paper $\mathbf{1}_0$, and adding this extra citation to the other remaining paper $\mathbf{1}_1$, transforming it into a paper $\mathbf{1}_2$.

Contrary to the usual economic interpretation of Lorenz dominance that is often associated with the idea of promoting "equality" (Perny, Spanjaard, & Storme, 2006; Shorrocks, 1983), we use it in a dual way that promotes "elitism" (Bazen & Moyes, 2012). This point was also stressed by Albarran et al. (2011a, bottom of p. 53). This dual perspective has been criticized in Ravallion and Wagstaff (2011).

10.2. Setting

Consider an author $a \in A$. If author a is null or quasi-null, she has an Egghe (2006) index (henceforth g-index) of 0. Otherwise, a is strictly non-null so that $p_a > 0$ and $c_a > 0$. Suppose that $a \in A$ is such that:

$$a = \mathbf{1}_{x^a_{(1)}} + \mathbf{1}_{x^a_{(2)}} + \cdots + \mathbf{1}_{x^a_{(p_a)}}.$$

If $k > p_a$, we extend our previous notation and define $x^a_{(k)} = 0$. The g-index of a strictly non-null author a is defined as the largest $k \in \mathbb{N}_+$ such that

$$\sum_{i=1}^{k} x^a_{(i)} \geq k^2.$$

Hence, an author a has a g-index of $f_g(a) = k$ if she has k papers having collected a number of citations at least equal to k^2 and she does not have a set of $k+1$ papers having collected a number of citations at least equal to $(k+1)^2$.

As with the h-index, the g-index of a square author of size k is exactly k. When $f_g(a) = k > 0$, any set of k papers having received a total number of k^2 citations is called the *g-core* of $a \in A$.

Remark 23. We have defined $x^a_{(k)} = 0$ for $k > p_a$. This fact is important to compute the g-index. For instance, we have $f_g(\mathbf{1}_5 + \mathbf{1}_4) = 3$, although this author has only two papers.

As detailed in Woeginger (2008c) (see also Woeginger, 2009) there are in fact two versions of the g-index that were introduced by Egghe (2006). The one that we use here corresponds to the "Note added in proof" in Egghe (2006, p. 145) and is similar to the version studied in Woeginger (2008c). The other version, in which f_g cannot exceed p_a, was investigated in Quesada (2011a).

Remark 24. As was the case with $\succsim_h$ the ranking $\succsim_g$ violates independence (**C1**) and even weak independence (**C7**) (Waltman & van Eck, 2012). For instance, we have $f_g(\mathbf{1}_8) = f_g(\mathbf{1}_4) = 2$, while $f_g(\mathbf{1}_8 + \mathbf{1}_1) = 3 > f_g(\mathbf{1}_4 + \mathbf{1}_1) = 2$. Notice that the above example also shows that $\succsim_g$ violates **C12**.

The ranking $\succsim_g$ does not satisfy **C13**. Indeed, the g-index rewards authors for having highly cited papers. For instance, an author with a single paper that is highly cited may have an arbitrarily large g-index, whereas her h-index is always 1. On the contrary, it is easy to check that the g-index satisfies **C14**, since adding papers being cited at most x times to $x\mathbf{1}_x$ cannot increase the g-index. Condition **C15** is clearly satisfied by f_g.

It is not difficult to check that f_g satisfies **C4**, **C6** and **C9**. On the contrary, f_g violates **C2** and **C3** (since $f_g(\mathbf{1}_4) = 2 > f_g(\mathbf{1}_1) = 1$), **C8** (since $f_g(\mathbf{1}_2) = f_g(\mathbf{1}_1) = 1$), and **C10** (since $f_g(\mathbf{1}_2) = 1$, while $f_g(2\mathbf{1}_2) = 2$).

10.3. Conditions

C16 (Lorenz Monotonicity) *For all* $a, b \in A$,

$$a \trianglerighteq^L b \Rightarrow a \succsim b.$$

The above condition should be unsurprising. It is clearly satisfied by $\succsim_g$. Remembering that $a \trianglerighteq b$ implies $a \trianglerighteq^L b$, we know that **C16** is stronger than **A2**. It is useful to observe that, if $f_g(a) = x$, then $a \trianglerighteq^L x\mathbf{1}_x$. Moreover, if $f_g(a) = x$ and $p_a \le x$, then $\mathbf{1}_{x^2+2x} \trianglerighteq^L a$. This condition is violated by the ranking induced by the h-index. For instance, we have $\mathbf{1}_8 \trianglerighteq^L 2\mathbf{1}_4$, while $f_h(\mathbf{1}_8) = 1$ and $f_h(2\mathbf{1}_4) = 2$. It is clear that $\succsim_0$ and $\succsim_\tau$) (with $\tau > 0$) violate **C16**. This condition is clearly satisfied by $\succsim_{c_\tau}$, $\succsim_{t_\tau}$ (whatever τ), and $\succsim_M$.

C17 (Single paper author) *For all* $x \in \mathbb{N}_+$, *we have*

$$x\mathbf{1}_x \succsim \mathbf{1}_{(x+1)^2-1}.$$

The above condition says what is the maximal (with respect to Lorenz dominance) author with a single paper that is not above the square author $x\mathbf{1}_x$. It is also satisfied by the ranking induced by the h-index. It will be used as a replacement of **C13**. Observe that combining **C16** with **C17** implies, for instance, that $x\mathbf{1}_x \sim (x+1)\mathbf{1}_x$. Indeed, $(x+1)\mathbf{1}_x \trianglerighteq^L x\mathbf{1}_x$, so that $(x+1)\mathbf{1}_x \succsim x\mathbf{1}_x$. At the same time, we know that $\mathbf{1}_{x^2+2x} \trianglerighteq^L (x+1)\mathbf{1}_x$ and $x\mathbf{1}_x \succsim \mathbf{1}_{x^2+2x}$.

Condition **C17** is clearly violated by $\succsim_c$ and $\succsim_M$. It is also violated by $\succsim_{c_\tau}$ and $\succsim_{t_\tau}$ (with $\tau > 1$), and by $\succsim_\tau$ (with $\tau > 0$). It is satisfied by $\succsim_0$ and $\succsim_h$.

Our next condition will be used as a replacement for **C12** and **C14**. Consider a strictly non-null $a \in A$. Suppose that a has at least $k+1$ papers and is equivalent to a square author of size k, i.e., $a \sim k\mathbf{1}_k$. Suppose now that we add additional papers to a that are all less cited than the least cited paper of a and all have at most k citations. The condition implies that the addition of such weak papers to a cannot change its position in the ranking. It is simple to check that this condition is satisfied by $\succsim_g$. Indeed, if $a \sim_g k\mathbf{1}_k$ and a has at least $k+1$ papers, the least cited paper of a has at most k citations. The g-index of a is not altered after of the addition of weak papers. Note that the hypothesis that a has at least $k+1$ papers is crucial here. Indeed, for instance, we have $\mathbf{1}_7 + \mathbf{1}_1 \sim_g 2\mathbf{1}_2$ but $\mathbf{1}_7 + \mathbf{1}_1 + \mathbf{1}_1 \succ_g 2\mathbf{1}_2$.

C18 (Modified Tail Independence) *Let* $a \in A$ *be a strictly non-null author so that* $p_a > 0$. *Suppose that* $a = \mathbf{1}_{x^a_{(1)}} + \mathbf{1}_{x^a_{(2)}} + \cdots + \mathbf{1}_{x^a_{(p_a)}}$. *Suppose that* $k\mathbf{1}_k \sim a$, *with* $k < p_a$. *Let* $y \in \mathbb{N}$ *be such that* $y \le k$ *and* $y \le x^a_{(n)}$. *We have* $k\mathbf{1}_k \sim a + \mathbf{1}_y$.

Using **C16** with **C17**, we know that $(x+1)\mathbf{1}_x \sim x\mathbf{1}_x$. Repeated applications of **C18** imply that $(x+y)\mathbf{1}_x \sim x\mathbf{1}_x$, with $y \ge 2$. Combining this last expression with **C16** shows that **C16**, **C17**, and **C18** imply **C14**.

Condition **C18** is satisfied by $\succsim_h$. It is clearly violated by $\succsim_c$, $\succsim_{c_\tau}$, and $\succsim_{t_\tau}$ (with $\tau > 1$). It is satisfied by $\succsim_\tau$, whatever τ.

The following lemmas will be useful.

Lemma 26. *If a bibliometric ranking* $\succsim$ *satisfies* **C16**, **C17**, *and* **C18**, *then, for all* $a \in A$, $f_g(a) = k > 0$ *implies* $a \sim k\mathbf{1}_k$.

Proof. We distinguish two cases depending on the fact that $p_a \ge k$ or $p_a < k$.

(1) $p_a \le k$. The fact that $f_g(a) = k$ implies that $c_a \ge k^2$ and $c_a < (k+1)^2$. Since $p_a \le k$, we have:

$$\mathbf{1}_{k^2+2k} \trianglerighteq^L a \trianglerighteq^L k\mathbf{1}_k.$$

Using **C16** and **C17**, we obtain that $a \sim k\mathbf{1}_k$.

(2) $p_a > k$. We distinguish two cases.

Suppose that $p_a = k+1$. By construction, since $f_g(a) = k$, we have $\mathbf{1}_{k^2+2k} \trianglerighteq^L a$. Because we also have $a \trianglerighteq^L k\mathbf{1}_k$, we obtain, using **C16** and **C17**, $a \sim k\mathbf{1}_k$.

Suppose that $p_a > k+1$. Let $a^{(k+1)}$ be the author consisting of the $(k+1)$ most cited papers of a. Let $a^{-(k+1)}$ be the author consisting of all other papers of a. From the above case, we know that $a^{(k+1)} \sim k\mathbf{1}_k$. By construction, we know that $x^a_{(k+1)} \le k$ (Woeginger, 2008c). Hence, we have $x^a_{(k+2)} \le k$. Repeated applications of **C18** lead to $k\mathbf{1}_k \sim a^{(k+1)} + a^{-(k+1)} = a$.

□

10.4. Results

Our first result, in this section, characterizes $\succsim_g$.

Theorem 15. *A bibliometric ranking* $\succsim$ *satisfies conditions* **C11**, **C15**, **C16**, **C17**, *and* **C18**, *iff, for all* $a, b \in A$, $a \succsim b \Leftrightarrow f_g(a) \ge f_g(b)$.

Proof. Necessity is clear. We show sufficiency. Observe that, since **C11** and **C15** hold, we can use Lemma 23.

Suppose that a bibliometric ranking $\succsim$ satisfies **C11**, **C15**, **C16**, **C17**, and **C18**. We have to show that, for all $a, b \in A$ $a \succsim b \Leftrightarrow f_g(a) \geq f_g(b)$.

If $f_g(a) = f_g(b) = 0$, the a and b are either null or quasi-null and the conclusion follows from **C11**. If $f_g(a) = k > 0$ and $f_g(b) = 0$, using **C11** and Lemma 26, we know that $a \sim k\mathbf{1}_k$ and $b \sim \mathbf{0}$. The conclusion follows from **C16** and Lemma 23. If $f_g(a) = k > 0$ and $f_g(b) = \ell > 0$. Using Lemma 26, we know that $a \sim k\mathbf{1}_k$ and $b \sim \ell\,\mathbf{1}_\ell$. If $k = \ell$, the conclusion follows. Suppose that $k > \ell$. We know from Lemma 23 that $\mathbf{1}_1 \succ \mathbf{1}_0$. Using **C15**, we obtain, for all $x \in \mathbb{N}_+$, $(x+1)\mathbf{1}_{x+1} \succ x\mathbf{1}_x$. Hence, we have $k\mathbf{1}_k \succ \ell\,\mathbf{1}_\ell$. This completes the proof. □

To our knowledge, Theorem 15 is the only existing characterization of $\succsim_g$. We now turn to the characterization of f_g. This requires to bring conditions ORI and ES into the picture.

Theorem 16. *A bibliometric index f satisfies conditions* **C11**, **C15**, **C16**, **C17**, **C18**, *ORI, and ES, iff, for all* $a \in A$, $f(a) = \beta f_g(a)$, *for some* $\beta \in \mathbb{R}_+$.

Proof. Necessity was shown above. We concentrate below on sufficiency. Suppose that a bibliometric index f satisfies **C11**, **C15**, **C16**, **C17**, **C18**, ORI, and ES. We have to show that $f = \beta f_g$, for some $\beta \in \mathbb{R}_+$.

Using Lemma 23, we know that $f(\mathbf{1}_1) > 0$ and $f(\mathbf{0}) = 0$. Define $\beta = f(\mathbf{1}_1)$. We will show that, for all $a \in A$, $f(a) = \beta f_g(a)$.

Suppose that $f_g(a) = 0$, so that a is either null of quasi-null. The conclusion follows from ORI and Theorem 15.

Suppose that $f_g(a) = k > 0$. Using Lemma 26, we have $f(a) = f(k\mathbf{1}_k)$. Using Theorem 15, we know that $1\mathbf{1}_1$ is immediately above $\mathbf{0}$. Moreover, for all $y \in \mathbb{N}_+$, $(y+1)\mathbf{1}_{y+1}$ is immediately above $y\mathbf{1}_y$. Given ORI, the thesis follows from ES. □

Remark 25. Quesada (2011a) studies the variant of the g-index mentioned above. In this variant, we have $f'_g(a) = \min(f_g(a), p_a)$. As observed in Woeginger (2008c) using the example of the citation profile of John F. Nash, this variant of the g-index is not really satisfactory. Indeed, it seems to be in contradiction with the initial motivation for introducing the g-index, i.e., rewarding authors for having highly cited papers that are neglected by the h-index.

Quesada (2011a, Propositions 2.7 and 2.8) claims to characterize f'_g using, among others, an axiom S called Subadditivity. Unfortunately, this condition is not satisfied by f'_g. Indeed, let $a = \mathbf{1}_{1000}$ and $b = \mathbf{1}_1 + \mathbf{1}_1 + \mathbf{1}_1$. We clearly have $f'_g(a) = f'_g(b) = 1$. Condition S requires that, letting $c = \mathbf{1}_{1001} + \mathbf{1}_1 + \mathbf{1}_1$, we should have $f'_g(a) + f'_g(b) = 2 \geq f'_g(c)$. This is false since $f'_g(c) = 3$. Hence, Propositions 3.7 and 3.8 in Quesada (2011a) are incorrect as they stand. We are not presently aware of any simple way to fix them. Since we think that f'_g is not completely satisfactory, we do not pursue this point here.

10.5. Independence of conditions

Example 19 *(**C11** Quasi-null authors).* Consider the bibliometric index f such that, for all $a \in A$,

$$f(a) = \begin{cases} 0 & \text{if} \quad a \text{ is null or if } a = \mathbf{1}_0, \\ 1 & \text{if} \quad a = x\mathbf{1}_0 \text{ with } x > 0, \\ f_g(a) + 1 & \text{otherwise,} \end{cases}$$

This index clearly violates **C11**. It is simple to check that all other conditions are satisfied.

Example 20 *(**C15** Strong Uniformity).* Consider the bibliometric index f such that $f(a) = 0$ if $f_g(a) = 0$ and $f(a) = 1$ otherwise. Conditions **A1** and **A2** are clearly satisfied.

It is clear that this index violates **C15**. It is easy to check that all other conditions are satisfied. Indeed, **C11**, **C17**, **C16** and ORI are clearly satisfied. Conditions **C18** is satisfied since it involves non-strict comparisons of strictly non-null authors. Condition ES trivially holds.

Example 21 *(**C18** Modified Tail Independence).* Let $a \in A$ such that $f_g(a) = k$. We denote by $c^+(a)$ the total number of citations of the $k+1$ most cited papers of a and $c^-(a) = c_a - c^+(a)$.

Let us define the index f as follows. If $f_g(a) = 0$, then $f(a) = 0$. If $[f_g(a) = k > 0$ and $c^+(a) < k^2 + 2k]$ or if $[f_g(a) = k > 0$ and $c^+(a) = k^2 + 2k$ and $c^-(a) = 0]$, we have $f(a) = 2f_g(a) - 1$. If $[f_g(a) = k > 0$ and $c^+(a) = k^2 + 2k$ and $c^-(a) > 0]$, we let $f(a) = 2f_g(a)$.

Clearly the above definition covers all possible cases since, if $f_g(a) = k$, it is impossible that $c^+(a) > k^2 + 2k$.

It is not difficult to check that this index is a bibliometric index since it satisfies **A1** and **A2**.

It is clear that **C11** and ORI hold. It is not difficult to check that ES holds.

Transferring a citation from a paper beyond the $k+1$ most cited papers to a more cited paper remaining beyond the $k+1$ most cited papers cannot decrease f. Similarly, transferring a citation from a paper belonging to the $k+1$ most cited papers to a more cited paper cannot decrease f. Transferring a citation from a paper beyond the $k+1$ most cited papers to a paper belonging to the $k+1$ most cited papers has the following effect. If initially, we had $c^+(a) < k^2 + 2k - 1$, this transfer has no effect. If initially, we had $c^+(a) = k^2 + 2k - 1$, this transfer increases f by 1 if initially $c^-(a) \geq 2$. Otherwise, this transfer has no effect on f. If initially, we had $c^+(a) = k^2 + 2k$, this transfer increases f by 2. Hence, **C16** holds.

Condition **C18** is violated. For instance, if $a = 4\mathbf{1}_5 + \mathbf{1}_4$, we have $f(a) = 7$, since $f_g(a) = 4$, $c^+(a) = 24$ and $c^-(a) = 0$. Yet we have $f(a + \mathbf{1}_1) = 8$ since $f_g(a + \mathbf{1}_1) = 4$, $c^+(a + \mathbf{1}_1) = 24$ and $c^-(a + \mathbf{1}_1) > 0$.

Example 22 *(**C16** Lorenz Monotonicity).* Consider the bibliometric index f_h. It clearly violates **C16**. All other conditions are clearly satisfied.

Example 23 *(**C17** Single paper author).* The index f_M violates **C17**. It clearly satisfies **A1**, **A2**, **C11**, ORI, ES, and **C16**. It satisfies **C18** since, if a is such that $k\mathbf{1}_k \sim_M a^{(k+1)}$, then we have $a \sim_M a^{(k+1)}$, with $a^{(k+1)}$ defined as in the proof of Lemma 26.

Rephrasing the above examples in terms of the bibliometric ranking induced by the bibliometric index shows that the conditions used in Theorem 15 are independent. The standard examples show that ORI and ES cannot be omitted in Theorem 16.

11. Remarks

11.1. Independence and monotonicity

Using **C1** (Independence), what happens with authors having a single paper has consequences for authors having more than one paper. Hence, when **C1** holds, the full strength of **A2** is not really needed. The only implication of **A2** that is used in all theorems using **C1** is that, for all $x, y \in \mathbb{N}$ such that $x \geq y$,

$$\mathbf{1}_x \succsim \mathbf{1}_y \succsim \mathbf{0},$$

This above condition is exactly equivalent to the combination of Lower Bound and CDNH in Marchant (2009a) that we recall below.

A3 (CDNH) *For all* $x, y \in \mathbb{N}$, $x \geq y \Rightarrow \mathbf{1}_x \succsim \mathbf{1}_y$.

A4 (Lower bound) *For all* $x \in \mathbb{N}$, $\mathbf{1}_x \succsim \mathbf{0}$.

These two conditions may replace **A2** for the study of rankings and indices that are independent (a formal proof of this fact can be deduced using the lemmas in the next section). This may be seen as an advantage, since they are more elementary than **A2**. But, since many rankings and indices discussed in this paper are not independent, we need **A2** in many results and, for the sake of consistency, we decided to use it in all results.

11.2. Weak independence and monotonicity

Basically there are two distinct forms of monotonicity. The first one deals with citations, while the other one deals with publications. A first form of monotonicity deals with citations and says what happens to the performance of an author when one of her papers receives an additional citation. Clearly, this should not lead to a lower performance. In order for an author to receive an additional citation, she must be non-null. It is always possible to write a non-null author as $c + \mathbf{1}_x$, where $c \in A$ is a possibly null author and $x \in \mathbb{N}$. Suppose that the paper $\mathbf{1}_x$ receives an additional citation, everything else remaining unchanged. The resulting author is $c + \mathbf{1}_{x+1}$. In such a case, it is tempting to consider that the performance of $c + \mathbf{1}_{x+1}$ should not be lower than the performance of $c + \mathbf{1}_x$. This is a first form of monotonicity that only concerns citations: receiving an additional citation, ceteris paribus, cannot decrease performance. All the indices (and rankings) that we have analyzed satisfy this first form of monotonicity. We formalize this condition below

A5 (Weak Citation Monotonicity) *For all* $b \in A$ and $x \in \mathbb{N}$, $b + \mathbf{1}_{x+1} \succsim b + \mathbf{1}_x$.

Another form of monotonicity deals with publications and says what happens to the performance of an author when she publishes a new paper. Consider an author $a \in A$. Suppose that this author publishes a new paper receiving x citations, with $x \in \mathbb{N}$. Hence, a becomes $a + \mathbf{1}_x$. It is then tempting to conclude that the performance of $a + \mathbf{1}_x$ should not be inferior to the performance of a. This is a form of monotonicity that only concerns publications: publishing an extra paper, ceteris paribus, cannot decrease performance. All the indices (and rankings) that we have analyzed satisfy this second form of monotonicity. We formalize this condition below. It is identical to condition A6 in Marchant (2009a, p. 328).

A6 (Weak Publication Monotonicity) *For all* $a \in A$ and $x \in \mathbb{N}$, $a + \mathbf{1}_x \succsim a$.

The next two lemmas show that, in the presence of **C7**, **A3** implies **A5** and **A4** implies **A6**.

Lemma 27. *If a bibliometric ranking* $\succsim$ *satisfies* **A3** *and* **C7**, *it also satisfies* **A5**.

Proof. Using **A3**, we know that $\mathbf{1}_{x+1} \succsim \mathbf{1}_x$, for all $x \in \mathbb{N}$. Let $b \in A$. Using **C7**, we obtain $\mathbf{1}_{x+1} + b \succsim \mathbf{1}_x + b$. □

Lemma 28. *If a bibliometric ranking* $\succsim$ *satisfies* **A4** *and* **C7**, *it also satisfies* **A6**.

Proof. Using **A4**, we know that $\mathbf{1}_x \succsim \mathbf{0}$, for all $x \in \mathbb{N}$. For all $a \in A$, using **C7**, we obtain $\mathbf{1}_x + a \succsim \mathbf{0} + a = a$. □

Finally, we show that, in the presence of **C7**, **A3** and **A4** imply **A2**.

Lemma 29. *If a bibliometric ranking* $\succsim$ *satisfies* **A3**, **A4**, *and* **C7**, *it also satisfies* **A2**.

Proof. We know from the two preceding lemmas that **A5** and **A6** hold. It is clear that the conjunction of **A5** and **A6** implies **A2**. □

Using Lemma 29, it is therefore possible to formulate variants of our results that use **C7** or **C1**, replacing **A2** with **A3** and **A4**.

Table 2
Summary of results for rankings: "**Y**" indicates a characterizing condition, "y" indicates a condition that is satisfied, "n" indicates a condition that is violated. In the columns for $\succsim_\tau$ (resp. $\succsim_{c_\tau}$ and $\succsim_{t_\tau}$), a "y" indicates that the condition holds for all $\tau > 0$ (resp. $\tau > 1$).

	$\succsim_\tau$	$\succsim_0$	$\succsim_{c_\tau}$	$\succsim_{t_\tau}$	$\succsim_c$	$\succsim_M$	$\succsim_h$	$\succsim_g$
C1	**Y**	**Y**	**Y**	**Y**	**Y**	n	n	n
C2	**Y**	y	n	y	n	n	y	n
C3	n	**Y**	n	n	n	n	n	n
C4	n	n	**Y**	n	y	n	n	y
C5	n	y	n	**Y**	y	n	n	y
C6	n	n	n	n	**Y**	n	y	y
C7	y	y	y	y	y	**Y**	n	n
C8	n	n	n	n	y	**Y**	n	n
C9	y	n	y	**Y**	y	**Y**	y	y
C10	n	n	n	n	n	**Y**	n	n
C11	y	n	y	y	y	y	**Y**	**Y**
C12	y	y	y	y	y	y	**Y**	n
C13	n	y	n	n	n	n	**Y**	n
C14	n	n	n	n	n	y	**Y**	y
C15	n	y	n	n	y	y	**Y**	**Y**
C16	n	n	y	y	y	y	n	**Y**
C17	y	y	n	n	n	n	y	**Y**
C18	y	y	n	n	n	y	y	**Y**

12. Conclusion

Table 2 gives a summary of our results for rankings. Entries marked with "**Y**" result from Theorems 1, 3, 5, 9, 11, 13, and 15. The proof for the "y" and "n" entries was noted in the text. In order to avoid redundancies with the columns $\succsim_0$ and $\succsim_c$, the column for $\succsim_\tau$ (resp. $\succsim_{c_\tau}$ and $\succsim_{t_\tau}$) states what happens when $\tau > 0$ (resp. $\tau > 1$).

12.1. Classic and "modern" rankings and indices

It has often been argued (Alonso et al., 2009; Egghe, 2010a; Hirsch, 2005; Norris & Oppenheim, 2010; Ruscio et al., 2012) that the h-index was combining "quantity" (number of papers) and "quality" (number of citations) in a way that leads to a *robust* ranking and index. Indeed, the h-index does not reward the publication of lowly cited papers. Similarly, it is not sensitive to the existence of a few papers having attracted many citations. As forcefully stressed in van Eck and Waltman (2008), Waltman and van Eck (2009a), these properties are shared by the index f_τ (we will limit ourselves here to the discussion of indices since the situation for rankings is similar).

It has often been argued that an advantage of the h-index over the number of highly cited papers f_τ is that is does not require to determine the value of τ. Indeed, the index f_τ is clearly dependent upon the choice of τ. van Eck and Waltman (2008) have argued that the fact that the h-index does not use a parameter such as τ is largely due to an artifact. Instead of considering square authors as central, one may have chosen to work instead with rectangle authors (rectangle authors are also considered in the *step-based indices* studied in Chambers & Miller, 2014). The shape of the rectangle authors would require setting a parameter just like τ. Deciding that among all possible rectangle authors, square authors play a central part, involves a great deal of arbitrariness, as stressed in van Eck and Waltman (2008). Hence, the "modern" h-index does not differ from its classical counterpart f_τ on this account. We refer to Schreiber (2013b) for an empirical study of the impact varying τ for f_τ and the impact of the use of the variants of the h-index suggested in van Eck and Waltman (2008).

We may hence try to compare f_h and f_τ using their formal properties. In our view, the comparison of Theorems 2 and 14 is enlightening. The index f_h does not satisfy independence (we refer to Ye, 2013, for a skeptical view on independence). Its characterizing properties are complex and not easy to motivate. On the contrary, the index f_τ satisfies independence and its characterizing properties are simple and appealing. Schreiber (2013a) offers[7] a more critical view on the index f_τ. Indeed, this index may violate a property stating that the comparison of two authors should not be altered by a common absolute improvement, i.e., the addition of the same number of citations to each paper of these two authors. We do not think that this property is very compelling however. Indeed, since the two authors may have, at the beginning, a very different number of papers and of citations, the "common" improvement consisting of adding the same number of citations to each paper of these two authors, may be *quite* different for the two authors. The same is true for the case of a common relative improvement that is also studied by Schreiber (2013a).

Both f_τ and f_h emphasize the number of "important" papers, i.e., papers that have been cited sufficiently often. On the contrary, the indices $f^*_{c_\tau}$, $f^*_{t_\tau}$ and f_g emphasize the total number of citations of "important" papers, i.e., papers that have been cited sufficiently often (this is also the case for f_{c_τ}, but this index violates ES). This interpretation of f_g was proposed and discussed in Schreiber et al. (2011). It is not consensual however, see Bornmann et al. (2008), Bornmann, Mutz, Daniel,

[7] We thank Ludo Waltman for having brought this paper to our attention.

Wallon, and Ledin (2009). The above comparison between f_h and f_τ carries over to the comparison between $f^*_{c_\tau}$ (or f_{t_τ}) and f_g. The fact that the g-index does not use a parameter, contrary to $f^*_{c_\tau}$ (or f_{t_τ}) is again largely due to an artifact. The index f_g does not satisfy independence and its characterizing properties are not easy to motivate. On the contrary, the indices $f^*_{c_\tau}$ and f_{t_τ} satisfy independence and their characterizing properties seem simple and appealing.

The above critical remarks on f_h or f_g are based on formal arguments. As already mentioned, such arguments only consider one aspect of bibliometric rankings and indices. The analysis of f_h or f_g should also be based on empirical evidence. However, the empirical evidence that we are aware does not seem, up to now, to give clear cut arguments in favor of f_h (resp. f_g) over f_τ (resp. $f^*_{c_\tau}$).

12.2. Limitations and directions for future research

We have stressed above that we agreed with the limitations of the axiomatic analysis of bibliometric rankings and indices put forward in Glänzel and Moed (2013). Hence, we simply view our results as providing a general framework for the comparison of the formal properties of several indices and rankings. The study of these formal properties does not exhaust the analysis of these indices and rankings.

Besides this clear limitation of our work, we would like to stress some others.

First, we have not studied the incredibly many variants of the h- and g-index that have been proposed in the literature (Alonso et al., 2009; Egghe, 2010a; Norris & Oppenheim, 2010; Ruscio et al., 2012; Schreiber et al., 2011). This would have been clearly impossible within a single paper. The h-index and the g-index represent two focal elements in this literature, the first focusing on counting highly-cited papers, the second rewarding authors having highly cited papers. We feel that it should not be a major difficulty to adapt our conditions for the study of the many variants (Bornmann et al., 2011, p. 349 present no less than 37 variants of the h-index) of these indices (see Kosmulski, 2013, for a synthetic presentation). The empirical literature is not consensual on the way to categorize these variants (compare, e.g., Bornmann et al., 2008, with (Schreiber et al., 2011).

Second, we would like to stress that our paper leaves completely untouched some of the most difficult problems that are to be faced in evaluative bibliometrics:

- the field normalization of indices,
- the adequate treatment of various types of publications (articles, reviews, letters, notes),
- the correction of indices for the length of careers,
- the proper treatment of multiple authors,
- the proper treatment of multiple affiliations.

Our analysis offers no clue on how to deal with these “bibliometric nightmares”. Nevertheless, we believe that an axiomatic approach could shed some light on these difficulties. This will be the subject of further studies.

References

Albarrán, P., Ortuño, I., & Ruiz-Castillo, J. (2011a). The measurement of low- and high-impact in citation distributions: Technical results. *Journal of Informetrics*, *5*(1), 48–63. Corrigendum: 2012. *Journal of Informetrics, 6*(4), 515.

Albarrán, P., Ortuño, I., & Ruiz-Castillo, J. (2011b). High- and low-impact citation measures: Empirical applications. *Journal of Informetrics*, *5*(1), 122–145.

Alonso, S., Cabreziro, F. J., Herrera-Viedma, E., & Herrera, F. (2009). *h*-Index: A review focused on its variants, computation and standardization for different scientific fields. *Journal of Informetrics*, *3*, 273–289. See also the companion website http://sci2s.ugr.es/hindex/ (accessed 10.07.13)

Bazen, S., & Moyes, P. (2012). Elitism and stochastic dominance. *Social Choice and Welfare*, *39*, 207–251.

Bornmann, L., & Leydesdorff, L. (2013). The validation of (advanced) bibliometric indicators through peer assessments: A compartive study using data from InCites and F1000. *Journal of Informetrics*, *7*, 286–291.

Bornmann, L., Mutz, R., & Daniel, H.-D. (2008). 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), 830–837.

Bornmann, L., Mutz, R., Daniel, H.-D., Wallon, G., & Ledin, A. (2009). Are there really two types of *h* index variants? A validation study by using molecular life scince data. *Research Evaluation*, *18*(3), 185–190.

Bornmann, L., Mutz, R., Hug, S. E., & Daniel, H.-D. (2011). A multilevel meta-analysis of studies reporting correlations between the *h* index and 37 different *h* index variants. *Journal of Informetrics*, *5*, 348–359.

Bouyssou, D., & Marchant, T. (2010). Consistent bibliometric rankings of authors and journals. *Journal of Informetrics*, *4*(3), 365–378.

Bouyssou, D., & Marchant, T. (2011a). Bibliometric rankings of journals based on impact factors: An axiomatic approach. *Journal of Informetrics*, *5*(1), 75–86.

Bouyssou, D., & Marchant, T. (2011b). Ranking scientists and departments in a consistent manner. *Journal of the American Society for Information Science and Technology*, *62*(9), 1761–1769.

Chambers, C. P., & Miller, A. D. (2014). Scholarly influence. *Journal of Economic Theory*, *151*(1), 571–583.

Chapron, G., & Husté, A. (2006). Open, fair, and free journal ranking for researchers. *Bioscience*, *56*, 558–559.

Deineko, V. G., & Woeginger, G. J. (2009). A new family of scientific impact measures: The generalized Kosmulski-indices. *Scientometrics*, *80*(3), 819–828.

Egghe, L. (2005). *Power laws in the information production process: Lotkaian informetrics*. Elsevier: Amsterdam.

Egghe, L. (2006). Theory and practice of the *g*-index. *Scientometrics*, *69*(1), 131–152.

Egghe, L. (2010a). The Hirsch-index and related impact measures. *Annual Review of Information Science and Technology*, *44*, 65–114.

Egghe, L. (2010b). Conjugate partitions in informetrics: Lorenz surve, *h*-type indices, Ferrers graphs and Durfee sqare in a discrete and continuous setting. *Journal of Informetrics*, *4*, 320–330.

Egghe, L. (2012). An econometric property of the *g*-index. *Information Processing and Management*, *45*(4), 484–489.

Egghe, L. (2013). A mathematical characterization of the Hirsch-index by means of minimal increments. *Journal of Informetrics*, *7*, 388–393.

Eto, H. (2003). Interdisciplinary information input and output of a nano-technology project. *Scientometrics*, *58*, 5–33.

Franceschini, F., & Maisano, D. A. (2010). Analysis of the Hirsch index's operational properties. *European Journal of Operational Research*, *203*, 494–504.

Gagolewski, M., & Grzegorzewski, P. (2009). A geometric approach to the construction of scientific impact indices. *Scientometrics*, *81*(2), 617–634.
Glänzel, W., & Moed, H. F. (2013). Opinion paper: Thoughts and facts on bibliometric indicators. *Scientometrics*, *96*(1), 381–394.
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.
Hwang, Y.-A. (2013). An axiomatization of the Hirsch-index without adopting monotonicity. *Applied Mathematics & Information Sciences. An International Journal*, *7*(4), 1317–1322.
Kosmulski, M. (2013). Family-tree of bibliometric indices. *Journal of Informetrics*, *7*(2), 313–317.
Liu, Y., Zuo, W., Goa, Y., & Qian, Y. (2013). Comprehensive geometrical interpretation of *h*-type indices. *Scientometrics*, *96*, 605–615.
Marchant, T. (2009a). An axiomatic characterization of the ranking based on the *h*-index and some other bibliometric rankings of authors. *Scientometrics*, *80*(2), 327–344.
Marchant, T. (2009b). Score-based bibliometric rankings of authors. *Journal of the American Society for Information Science and Technology*, *60*, 1132–1137.
Marshall, A. W., Olkin, I., & Arnold, B. C. (2011). *Inequalities: Theory of Majorization and Its Applications* (2nd ed.). Heidelberg: Springer., ISBN 978-0-387-40087-7.
Miroiu, A. (2013). Axiomatizfing the Hirsch index: Quantity and quality disjoined. *Journal of Informetrics*, *7*(1), 10–15.
Norris, M., & Oppenheim, C. (2010). The *h*-index: A broad review of a new bibliometric indicator. *Journal of Documentation*, *66*(5), 681–705.
Perny, P., Spanjaard, O., & Storme, L.-X. (2006). A decision-theoretic approach to robust optimization in multivalued graphs. *Annals of Operations Research*, *147*, 317–341.
Quesada, A. (2009). Monotonicity and the Hirsch index. *Journal of Informetrics*, *3*(2), 158–160.
Quesada, A. (2010). More axiomatics for the Hirsch index. *Scientometrics*, *82*, 413–418.
Quesada, A. (2011a). Axiomatics for the Hirsch index and the Egghe index. *Journal of Informetrics*, *5*, 476–480.
Quesada, A. (2011b). Further characterizations of the Hirsch index. *Scientometrics*, *87*, 107–114.
van, A. F. J., & Raan. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*, *67*, 491–502.
Ravallion, M., & Wagstaff, A. (2011). On measuring scholarly influence by citations. *Scientometrics*, *88*, 321–337.
Rousseau, R. (2008). Woeginger's axiomatisation of the *h*-index and its relation to the *h*-index, the $h^{(2)}$-index and the $r^{(2)}$-index. *Journal of Informetrics*, *2*(4), 335–340.
Rousseau, R., García-Zoritad, C., & Sanz-Casadod, E. (2013). The *h*-bubble. *Journal of Informetrics*, *7*(2), 294–300.
Ruscio, J., Seaman, F., D'Oriano, C., Stremlo, E., & Mahalchik, K. (2012). Measuring scholarly impact using modern citation-based indices. *Measurement: Interdisciplinary Research and Perspectives*, *10*(3), 123–146.
Schreiber, M. (2013a). Inconsistencies of the highly-cited-publications indicator. *Journal of the American Society for Information Science and Technology*, *84*(8), 12982–21302.
Schreiber, M. (2013b). A case study of the arbitrariness of the *h*-index and the highly-cited-publications indicator. *Journal of Informetrics*, *7*(2), 379–387.
Schreiber, M., Malesios, C. C., & Psarakis, S. (2011). Categorizing *h*-index variants. *Research Evaluation*, *20*(5), 397–408.
Shorrocks, A. F. (1983). Ranking income distributions. *Economica*, *50*, 3–17.
van Eck, N. J., & Waltman, L. (2008). Generalizing the *h*- and *g*-indices. *Journal of Informetrics*, *2*, 263–271.
Waltman, L., & van Eck, N. J. (2009a). A simple alternative to the *h*-index. *ISSI Newsletter*, *5*(3), 46–48.
Waltman, L., & van Eck, N. J. (2009b). A taxonomy of bibliometric performance indicators based on the property of consistency. Technical Report ERS-2009-014-LIS, Erasmus University Rotterdam, Erasmus Research Institute of Management, Rotterdam, The Netherlands. *Presented at the 12th International Conference on Scientometrics and Informetrics* Rio de Janeiro.
Waltman, L., & van Eck, N. J. (2012). The inconsistency of the *h*-index. *Journal of the American Society for Information Science and Technology*, *63*(2), 406–415.
Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. J. (2011). Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics*, *5*(1), 37–47.
Waltman, L., Costas, R., & van Eck, N. J. (2012). Some limitations of the *H* index: A commentary on Ruscio and colleagues' analysis of bibliometric indices. *Measurement: Interdisciplinary Research and Perspectives*, *10*(3), 173–175.
Woeginger, G. J. (2008a). An axiomatic characterization of the Hirsch-index. *Mathematical Social Sciences*, *56*, 224–232.
Woeginger, G. J. (2008b). A symmetry axiom for scientific impact indices. *Journal of Informetrics*, *2*(4), 298–303.
Woeginger, G. J. (2008c). An axiomatic analysis of Egghe's *g*-index. *Journal of Informetrics*, *2*(4), 364–368.
Woeginger, G. J. (2009). Generalizations of Egghe's *g*-index. *Journal of the American Society for Information Science and Technology*, *60*(6), 1267–1273.
Ye, F. Y. (2013). H-inconsistency is not an issue in dynamical systems. *ISSI Newsletter*, *8*(2), 22–24.