An axiomatic characterization of the ranking based on the h-index and some other bibliometric rankings of authors

Thierry Marchant*

February 20, 2008

Abstract

In the last few years, many new bibliometric rankings or indices have been proposed for comparing the output of scientific researchers. We propose a formal framework in which rankings can be axiomatically characterized. We then present a characterization of some popular rankings. We argue that such analyses can help the user of a ranking to choose one that is adequate in the context where she/he is working.

1 Introduction

As bibliometric rankings of scientific researchers, departments and universities become more and more popular, it is not surprising that many researchers propose new techniques for deriving these rankings. Most of these techniques are based on indexes, like the number of publications [van Raan, 2006], the total number of citations [Hirsch, 2005, van Raan, 2006], the average number of citations [van Raan, 2006], the maximal number of citations [Eto, 2003], the h-index [Hirsch, 2005], the generalized h-index [Sidiropoulos et al., 2007], the g-index [Egghe, 2006], the number of publications with at least α citations [Chapron and Husté, 2006], and so on. Since all these indices yield different rankings, anybody willing to use such a ranking faces an important question: which ranking shall he use ? Which one is the good one ? So far, researchers trying to motivate a given ranking have followed three (non-exclusive) main approaches.

The first one is often adopted by researchers proposing a new ranking. They first identify a weakness of some existing ranking and they then propose a new ranking that does not exhibit the same weakness. For instance, Egghe [2006] writes "The h-index is also robust in the sense that it is insensitive to a set of uncited (or lowly cited) papers but also it is insensitive to one or several outstandingly highly cited papers. This last aspect can be considered as a drawback

*Address: Ghent University, Dunantlaan 1, 9000 Ghent, Belgium. Email: thierry.marchant@ugent.be

of the h-index." He then uses this to motivate the g-index. Sidiropoulos et al. [2007] follows the same route. This approach suffers a severe drawback: it is extremely myopic because it focusses on a specific aspect of rankings and pays very little attention to other aspects. Suppose for instance a researcher proposes a new ranking, improving an older one on some aspect. It may happen that this new ranking has many other weaknesses that the old ranking does not have and these new weaknesses can eventually be severe ones. It is therefore not obvious that the new ranking is an overall improvement.

The second approach is based on comparisons of a given ranking or index with other well-known rankings. For instance, van Raan [2006] compares the h-index with several other indices and peer ratings. A difficulty with this interesting approach is that it can tell us which rankings correlate high with the one under scrutiny but, as there is no reference ranking or objectively true ranking, it does not tell us whether the one under scrutiny is right.

The third approach tries to motivate an index, using a deterministic [Hirsch, 2005, Egghe, 2006] or probabilistic [Glänzel, 2006] model of publication and citation, depending on some parameters reflecting productivity, quality or some other characteristic of authors. The main weakness of this interesting approach is that it relies on the choice of a publication and citation model and, unfortunatly, we do not know which model is right.

In the second and third approaches, one tries to motivate a ranking by external arguments (publication models or other rankings). In the first approach, one tries to motivate a ranking by its intrinsic qualities. For example, 'it is sensitive to one or several outstandingly highly cited papers' or 'it is difficult to manipulate' or 'it rewards high productivity.' In this paper, we will take the first route, but in a systematic and non-myopic way. We will not focus on some weakness or nice property of a ranking. Instead, we will consider a set of properties that entirely characterize a given ranking and we will do this for several rankings. So, for each of these rankings, we will know which properties they satisfy and which ones they do not. Since we will use characterizing properties, we will have a complete picture of the situation, not focussing on some aspect.

It is important to acknowledge that there is no objectively right ranking. Indeed, suppose a university wants to develop an incentive program in order to reward its tenured professors. To some extent, the quality of these professors is fixed. Some are genious, some are not. Giving more reward to the geniuses would be unfair (from an equal opportunity perspective). So, this university is going to reward productivity, because this is the only thing that professors can improve by working harder (an average researcher will not publish big hits just by working harder). So, this university might rank the professors by decreasing number of publications. Suppose now that this university wants to hire a new professor. Among all candidates, it will of course select the one that publishes many high quality papers (with many citations). So, this time, the university might rank the candidates by decreasing number of citations. According to the context, different rankings can be adequate.

In each context, the user of a ranking thus needs to choose an adequate ranking. The analysis that will be presented in this paper will help the user to make this choice. For a couple of popular rankings, we will present a list of properties (between 2 and 6) that completely characterize it. Choosing an adequate ranking will then be greatly facilitated by considering properties that seem desirable in a given context.

In Section 2, we will introduce the notation and the main concepts that we will use in Section 3 for characterizing some popular rankings. Section 4 summarizes and concludes.

2 Notation and definitions

We represent an author by a mapping f from $\mathbb{N}$ to $\mathbb{N}$ and we interpret $f(x)$ as the number of publications of author f with exactly x citations. Let $\#(f) = \sum_{x\in\mathbb{N}} f(x) = \sum_x f(x)$ denote[1] the total number of publications of f and $\&(f) = \sum_x xf(x)$, the total number of citations of f. Let X be the set of all mappings f from $\mathbb{N}$ to $\mathbb{N}$ such that $\#(f)$ is finite. This set is called the set of authors. The elements of X will usually be denoted by $f, f', g, \ldots$ We want to construct a ranking (a complete and transitive binary relation[2]) $\succsim$ on X. The statement '$x \succsim y$' is interpreted as 'given their publication/citation records, author x is at least as good as author y.' When $x \succsim y$ and $y \not\succsim x$, we write $x \succ y$ (x is strictly better than y). When $x \succsim y$ and $y \succsim x$, we write $x \sim y$ (x and y are equivalent).

For all $x \in \mathbb{N}$, we denote by $\mathbf{1}_x$ the author such that $\mathbf{1}_x(x') = 0$ for all $x' \neq x$ and $\mathbf{1}_x(x) = 1$. So, $\mathbf{1}_x$ represents an author with exactly one publication and such that this publication is cited x times. An author without publication is represented by $\mathbf{0}$. We write $f \sqsubseteq g$ when $f(x) \leq g(x),\ \forall x \in \mathbb{N}$.

We now present some desirable properties that should definitely be satisfied by any sensible bibliometric ranking. These properties will be called axioms.

A 1 Non-Triviality. *There are f and g such that $f \succ g$.*

This axiom just expresses the fact that we do not want a complete tie; we want to discriminate among authors.

A 2 CDNH. *For all x, x', $x \geq x'$ implies $\mathbf{1}_x \succsim \mathbf{1}_{x'}$.*

The name CDNH stands for 'Citations Do Not Harm'. In other words, if two authors have a single publication each, then the author that has more citations cannot be ranked in a lower position than the other one. Since these two axioms are so compelling, we do not want to consider rankings that do not satisfy them. That is why we include them in the next definition.

Definition 1 *A bibliometric ranking is a complete and transitive relation on X satisfying Non-Triviality and CDNH.*

[1] From now on, we write $\sum_x f(x)$ instead of $\sum_{x\in\mathbb{N}} f(x)$. Similarly, we will write $\max_x$, argmax_x or 'for all $x \ldots$'

[2] A binary relation $\succsim$ on a set X is transitive if, $\forall x, y, z \in X$, $x \succsim y$ and $y \succsim z$ imply $x \succsim z$. It is complete if, $\forall x, y \in X$, $x \succsim y$ or $y \succsim x$.

3 Characterization of some bibliometric rankings

In this section, we characterize some popular bibliometric rankings.

3.1 The number of publications

The ranking based on the total number of publications will be denoted by $\succsim_{\#}$. Formally, $f \succsim_{\#} g$ iff $\#(f) \geq \#(g)$. In order to characterize this ranking, we will use the following axioms.

A 3 Lower Bound. *For all x, $\mathbf{0} \precsim \mathbf{1}_x$.*

In other words, one publication can not be worse than zero. This is an extremely reasonable condition.

A 4 Independence. *For all $f, g, e \in X$, $f \succsim g$ iff $f + e \succsim g + e$.*

In the statement of this axiom, $f + e$ is the sum of two functions. It is therefore a function and represents also an author. Intuitively, Independence can be understood as follows. Suppose an author f is at least as good as g. Suppose also both of them publish some additional papers: the same number of papers, each with the same number of citations. So, both made the same improvement. Then these two authors (now represented by $f+e$ and $g+e$) should compare in the same way as previously, i.e., $f+e$ is at least as good as $g+e$. Independence will sometimes be weakened to

A 5 Weak Independence. *For all $f, g, e \in X$, $f \succsim g$ implies $f + e \succsim g + e$.*

With this condition, $f + e$ and $g + e$ need not to compare in the same way as f and g but no reversal can occur. In other words, if $f \succsim g$, we can have $f + e \sim g + e$ but not $f + e \prec g + e$. Obviously, Independence implies Weak Independence.

If Lower Bound and Weak Independence are satisfied, then the following condition holds.

A 6 Weak Publication Monotonicity. *For all $f \in X$ and all x, $f + \mathbf{1}_x \succsim f$.*

We formalize this in a lemma.

Lemma 1 *A bibliometric ranking satisfying Lower Bound and Weak Independence also satisfies Weak Publication Monotonicity.*

Proof. By Lower Bound, $\mathbf{1}_x \succsim \mathbf{0}$. Then, by Weak Independence, $f + \mathbf{1}_x \succsim f + \mathbf{0} = f$. □

This fact will be often used in the proofs of our results. Let us now analyze the normative content of Weak Publication Monotonicity. It means that you can not be worse off if you publish a new paper. At first sight, this is an extremely compelling condition but there might be contexts where it is not desirable. Suppose for instance we want to compare two outstanding authors f and g such that $f(100) = 100 = g(100)$. But $f = 100 \cdot \mathbf{1}_0 + 100 \cdot \mathbf{1}_{100}$ while $g = 100 \cdot \mathbf{1}_{100}$. So, both authors have the same number of highly cited papers but f has also many never cited papers. In this case, we might conclude that $g \succ f$, contradicting Weak Publication Monotonicity.

A 7 OIO. *For all* x, x', $\mathbf{1}_x \sim \mathbf{1}_{x'}$.

The name OIO stands for 'One Is One'. This axiom says that two authors, each with exactly one publication, are equivalent irrespective of their number of citations. This condition is quite strong. Many people will find it not reasonable but some others, doubting that citations are a sign of quality, will find it appealing.

Theorem 1 *The bibliometric ranking* $\succsim$ *satisfies Lower Bound (A3), Independence (A4) and OIO (A7) if and only if the authors are ranked according to their total number of publications, that is* $f \succsim g \iff \#(f) \geq \#(g)$, *for all* $f, g \in X$.

Before proving this theorem, we prove the following lemma.

Lemma 2 *If the bibliometric ranking* $\succsim$ *satisfies Weak Independence (A5), then there is* x *such that* $\mathbf{1}_x \succ \mathbf{0}$.

Proof. Suppose, on the contrary, $\mathbf{1}_x \sim \mathbf{0}$ for all x. By Weak Independence, $\mathbf{1}_x + \mathbf{1}_{x'} \sim \mathbf{0} + \mathbf{1}_{x'} = \mathbf{1}_{x'} \sim \mathbf{0}$, for all x, x'. Similarly, $\mathbf{1}_x + \mathbf{1}_{x'} + \mathbf{1}_{x''} \sim \mathbf{0} + \mathbf{1}_{x''} \sim \mathbf{0}$, for all x, x', x''. And we can easily go on adding more single-paper authors. Since any author f can be written as a finite sum of single-paper authors, we have $f \sim \mathbf{0}$. This contradicts Non-Triviality and concludes the proof. □

Proof of Theorem 1. We first prove the necessity of the axioms. Clearly, the relation $\succsim_\#$ satisfies Lower Bound and OIO. We now prove that it satisfies Independence. Suppose $f \succsim_\# g$. So, we know that $\#(f) \geq \#(g)$. Take any e in X. We have $\#(f + e) = \#(f) + \#(e)$. Similarly, $\#(g + e) = \#(g) + \#(e)$. This implies $\#(f+e) = \#(f)+\#(e) \geq \#(g)+\#(e) = \#(g+e)$. Hence, $f+e \succsim_\# g+e$. Let us now prove the converse implication. Suppose $f + e \succsim_\# g + e$. Using the same argument as in the preceding lines, it is easy to find that $f \succsim_\# g$.

We now prove the sufficiency. If we combine Independence and OIO, we find

$$f + \mathbf{1}_x \sim f + \mathbf{1}_{x'},\ \forall x, x'. \tag{1}$$

Let f and g be such that $\#(f) \geq \#(g)$. Let f' be such that $f' \sqsubseteq f$ and $\#(f') = \#(g)$. Since $\succsim$ is complete, we have $\mathbf{0} \sim \mathbf{0}$. Repeatedly using (1), we easily obtain $f' \sim g$. If $\#(f) = \#(g)$, then $f' = f$ and $f \sim g$. If $\#(f) > \#(g)$, then applying $(\#(f) - \#(g))$ times Weak Publication Monotonicity (Lemma 1) and transitivity yields $f \succsim f'$. If $f \sim f'$, then $f' + \mathbf{1}_x \sim f'$ for any x and, by

Independence, $\mathbf{1}_x \sim \mathbf{0}$ for any x. This contradicts Lemma 2. So, $f \succ f'$ and, by transitivity, $f \succ g$. $\square$

It is important to know whether the set of axioms characterizing $\succsim_{\#}$ is minimal, in the sense that no subset would do the job. We therefore check the independence of the conditions of Theorem 1. Each of the following bibliometric rankings satisfies all conditions of Theorem 1 but one. So, no condition can be dropped.

Lower Bound. Define $\succsim$ by $f \succsim g$ iff $\#(f) \leq \#(g)$. This bibliometric ranking (thus verifying Non-Triviality and CDNH) satisfies Independence and OIO but not Lower Bound. Indeed, in this ranking, any author with at least one publication is worse than $\mathbf{0}$.

Independence. Define $\succsim$ by $f \succsim g$ iff $\max_x f(x) \geq \max_x g(x)$. This bibliometric ranking satisfies Lower Gound and OIO but not Independence. Indeed, in this ranking, $\mathbf{1}_1 \succ \mathbf{0}$ but $\mathbf{1}_1 + 2 \cdot \mathbf{1}_2 \sim \mathbf{0} + 2 \cdot \mathbf{1}_2$.

OIO. Define $\succsim$ by $f \succsim g \Leftrightarrow \&(f) \geq \&(g)$. This bibliometric ranking satisfies Lower Bound and Independence but not OIO. Indeed, in this ranking, $\mathbf{1}_1 \succ \mathbf{1}_0$.

3.2 The total number of citations

The ranking based on the total number of publications will be denoted by $\succsim_{\&}$. Formally, $f \succsim_{\&} g$ iff $\&(f) \geq \&(g)$. In order to characterize it, we need one more axiom.

A 8 Additivity. *For all x, $\mathbf{1}_x + \mathbf{1}_1 \sim \mathbf{1}_{x+1}$.*

In other words, if you have only one publication, then getting one more citation for that paper or publishing one additional paper with one citation makes you equally better off.

Theorem 2 *The relation $\succsim$ satisfies Independence (A4) and Additivity (A8) if and only if the authors are ranked according to their total number of citations, that is $f \succsim g \iff \&(f) \geq \&(g)$.*

Proof. Clearly, the relation $\succsim_{\&}$ satisfies Additivity and Independence. We now prove the converse implications.

Suppose $x \leq y$ and let $f = x \cdot \mathbf{1}_0 + (y - x) \cdot \mathbf{1}_1$. By Independence and Additivity, $f + x \cdot \mathbf{1}_1 \sim f - x \cdot \mathbf{1}_0 + x \cdot \mathbf{1}_1$. So,

$$x \cdot \mathbf{1}_0 + y \cdot \mathbf{1}_1 \sim y \cdot \mathbf{1}_1,\ \forall x \leq y. \tag{2}$$

Suppose $x > y$ and let $f = x \cdot \mathbf{1}_0$. By Independence and Additivity, $f + y \cdot \mathbf{1}_1 \sim f - y \cdot \mathbf{1}_0 + y \cdot \mathbf{1}_1$. So,

$$x \cdot \mathbf{1}_0 + y \cdot \mathbf{1}_1 \sim (x - y) \cdot \mathbf{1}_0 + y \cdot \mathbf{1}_1,\ \forall x > y. \tag{3}$$

If $2y \geq x > y$, then $(x-y) \leq y$ and we can apply (2). This yields $(x-y) \cdot \mathbf{1}_0 + y \cdot \mathbf{1}_1 \sim y \cdot \mathbf{1}_1$. By transitivity, $x \cdot \mathbf{1}_0 + y \cdot \mathbf{1}_1 \sim y \cdot \mathbf{1}_1$. If $x > 2y$, we replace x by $x-y$ in (3) and we obtain

$$(x-y) \cdot \mathbf{1}_0 + y \cdot \mathbf{1}_1 \sim (x-2y) \cdot \mathbf{1}_0 + y \cdot \mathbf{1}_1, \ \forall x > 2y. \tag{4}$$

If $3y \geq x > 2y$, then $(x-2y) \leq y$ and we can apply (2). This yields $(x-2y) \cdot \mathbf{1}_0 + y \cdot \mathbf{1}_1 \sim y \cdot \mathbf{1}_1$. By transitivity, $x \cdot \mathbf{1}_0 + y \cdot \mathbf{1}_1 \sim y \cdot \mathbf{1}_1$. Applying the same reasoning for $x \in]3y, 4y],]4y, 5y], \ldots$ yields

$$x \cdot \mathbf{1}_0 + y \cdot \mathbf{1}_1 \sim y \cdot \mathbf{1}_1, \ \forall x, y. \tag{5}$$

Let f and g be such that $\&(f) \geq \&(g)$. Define f' and g' by $f' = f - f(0) \cdot \mathbf{1}_0$ and $g' = g - g(0) \cdot \mathbf{1}_0$. By (5), $f \sim f'$ and $g \sim g'$. By reflexivity, $\#(g') \cdot \mathbf{1}_1 \sim \#(g') \cdot \mathbf{1}_1$. Applying $(\&(g') - \#(g'))$ times Independence, Additivity and transitivity, we obtain $\&(g) \cdot \mathbf{1}_1 = \&(g') \cdot \mathbf{1}_1 \sim g' \sim g$. Applying now $\&(g)$ times Additivity, we easily obtain $\mathbf{1}_{\&(g)} \sim \&(g) \cdot \mathbf{1}_1 \sim g$. Following the same reasoning, $\mathbf{1}_{\&(f)} \sim \&(f) \cdot \mathbf{1}_1 \sim f$. By CDNH, $\&(f') \cdot \mathbf{1}_1 \succsim \&(g') \cdot \mathbf{1}_1$ and, so, $f \succsim g$.

Suppose $\&(f) = \&(g)$. Then $f \sim g$ because $f \sim \mathbf{1}_{\&(f)} = \mathbf{1}_{\&(g)} \sim g$. Suppose now $\&(f) > \&(g)$ and $f \sim g$. Then $(\&(f) - \&(g)) \cdot \mathbf{1}_1 \sim \mathbf{0}$. By Additivity, $\mathbf{1}_{(\&(f)-\&(g))} \sim \mathbf{0}$. So, by CDNH, $\mathbf{1}_1 \precsim \mathbf{0}$. Suppose $\mathbf{1}_1 \prec \mathbf{0}$. Then, by Independence, $\mathbf{1}_1 + \mathbf{1}_1 \prec \mathbf{1}_1$. By Additivity, $\mathbf{1}_2 \prec \mathbf{1}_1$. This contradicts CDNH. So, $\mathbf{1}_1 \sim \mathbf{0}$. Then, by Independence $\mathbf{1}_1 + \mathbf{1}_1 \sim \mathbf{1}_1 \sim \mathbf{0}$ and, by Additivity, $\mathbf{1}_2 \sim \mathbf{0}$. Repeating this reasoning, we find $\mathbf{1}_x \sim \mathbf{0}$ for any x, contradicting Lemma 2. Hence, $f \succ g$. □

We now check the independence of the conditions of Theorem 2. Each of the following bibliometric rankings satisfies all conditions of Theorem 2 but one. The verification is left to the reader.

Independence. Let $\succsim$ be the weak order with two equivalence classes, defined by

- $f \succ g$ iff $\#(f) \geq 3$ and $\#(g) \leq 2$,
- $f \sim g$ iff $\min\{\#(f), \#(g)\} \geq 3$ or , $\max\{\#(f), \#(g)\} \leq 2$.

Additivity. Define $\succsim$ by means of $f \succsim g \Leftrightarrow \&(f) - f(0) \geq \&(g) - g(0)$.

3.3 The maximal number of citations

The ranking based on the maximal number of citations will be denoted by $\succsim_M$. Formally, $f \succsim_M g$ iff $M(f) \geq M(g)$, where $M(f) = \max\{x : f(x) > 0\}$. In order to characterize this ranking, we will use the following axioms.

A 9 OPOEO. *For all x, $\mathbf{1}_x + \mathbf{1}_x \sim \mathbf{1}_x$.*

Informally, if you have only one publication, then a second publication with the same number of citations as the first one does not make you better off. The name OPOEO stands for 'One Plus One Equals One'. It favours quality over quantity, but in such an extreme way that there are probably few contexts where it will be judged appealing.

A 10 Uniformity. *For all* $x \geq 1$, $\mathbf{1}_{x-1} \prec \mathbf{1}_x$ *iff* $\mathbf{1}_x \prec \mathbf{1}_{x+1}$.

In words, if you have only one publication and, by getting an additional citation, you improve your position in the ranking, then getting a second additional citation should also improve your position. This condition is quite weak. Nevertheless, if we want a ranking with only a few equivalence classes (e.g., excellent, good, average, poor), then Uniformity is not relevant. On the contrary, if we want a very discriminating or fine-grained ranking, then Uniformity is almost compelling.

Theorem 3 *The relation* $\succsim$ *satisfies Lower Bound (A3), Weak Independence (A5), OPOEO (A9) and Uniformity (A10) if and only if the authors are ranked according to their maximal number of citations, that is* $f \succsim g \iff M(f) \geq M(g)$.

Proof. Clearly, the relation $\succsim_M$ satisfies Weak Independence, OPOEO and Uniformity. We now prove the converse implications. Combining Lemma 2 and Uniformity yields

$$x < y \text{ iff } \mathbf{1}_x \prec \mathbf{1}_y. \tag{6}$$

Successive applications of OPOEO and Weak Independence easily yield

$$y \cdot \mathbf{1}_x \sim \mathbf{1}_x, \; \forall x, y. \tag{7}$$

Let y and x_i ($i = 1 \ldots m$) be $m+1$ natural numbers such that $x_i \leq y$ for $i = 1 \ldots m$. By Weak Publication Monotonicity (Lemma 1), $\mathbf{1}_y \precsim \mathbf{1}_y + \mathbf{1}_{x_i}$ for all $i = 1 \ldots m$. By CDNH and Weak Independence, $\mathbf{1}_y + \mathbf{1}_{x_i} \precsim 2 \cdot \mathbf{1}_y$ and, by (7), $\mathbf{1}_y + \mathbf{1}_{x_i} \precsim \mathbf{1}_y$. So, $\mathbf{1}_y \sim \mathbf{1}_y + \mathbf{1}_{x_i}$. This, setting $i = 1$ and by Weak Independence, implies $\mathbf{1}_y + \mathbf{1}_{x_2} \sim \mathbf{1}_y + \mathbf{1}_{x_1} + \mathbf{1}_{x_2}$. By CDNH, Weak Independence and (7), $\mathbf{1}_y + \mathbf{1}_{x_1} + \mathbf{1}_{x_2} \precsim 3 \cdot \mathbf{1}_y \sim \mathbf{1}_y$. By transitivity, $\mathbf{1}_y \sim \mathbf{1}_y + \mathbf{1}_{x_1} + \mathbf{1}_{x_2}$. It is now clear that we can repeat the same process in order to obtain

$$\mathbf{1}_y \sim \mathbf{1}_y + \sum_{i=1}^{m} \mathbf{1}_{x_i}. \tag{8}$$

Let f and g be such that $M(f) > M(g)$. By (6), for all $y : g(y) > 0$, we have $\mathbf{1}_{M(g)} \succsim \mathbf{1}_y$. Applying (8) and transitivity yields $\mathbf{1}_{M(g)} \sim g$. A similar reasoning yields $\mathbf{1}_{M(f)} \sim f$. By (6), $\mathbf{1}_{M(f)} \succ \mathbf{1}_{M(g)}$ and, by transitivity, $f \succ g$. The case where $M(f) = M(g)$ is similar. □

We now check the independence of the conditions of Theorem 3. Each of the following bibliometric rankings satisfies all conditions of Theorem 3 but one.

Lower Bound. Define $\succsim$ by

- for all $f, g \neq \mathbf{0}$, $f \succsim g$ iff $M(f) \geq M(g)$;
- for all $f : M(f) \geq 1$, $f \succ \mathbf{0}$;
- for all $f : M(f) = 0$, $\mathbf{0} \succ f$.

Weak Independence. $f \succsim g$ iff $\operatorname{argmax}_x f(x) \geq \operatorname{argmax}_x g(x)$.

Uniformity. $f \sim \mathbf{0}$ for all $f : M(f) \leq 10$, $f \sim \mathbf{1}_{11}$ for all $f : M(f) > 10$ and $\mathbf{0} \prec \mathbf{1}_{11}$.

OPOEO. $f \succsim g$ iff $\&(f) \geq \&(g)$.

3.4 The number of papers with at least α citations

The ranking based on the number of papers with at least α citations ($\alpha > 0$) will be denoted by $\succsim_\alpha$. Formally, $f \succsim_\alpha g$ iff $\#_\alpha(f) \geq \#_\alpha(g)$, where $\#_\alpha(f) = \sum_{x \geq \alpha} f(x)$. In order to characterize this ranking, we will need two new axioms.

A 11 2-Gradedness. *For all $x < x' < x''$, $\mathbf{1}_x \sim \mathbf{1}_{x'}$ or $\mathbf{1}_{x'} \sim \mathbf{1}_{x''}$.*

In words, if your only publication is cited x times today, x' times tomorrow and x'' times the day after tomorrow, then it is possible that your position in the ranking the day after tomorrow is better than today. But it can not happen that your position tomorrow is better than today *and*, conjointly, that your position the day after tomorrow is better than tomorrow. Put in another way, increasing the number of citations of a unique publication can help once but not twice. This condition might be relevant when one wants a ranking with a small number of equivalence classes.

This condition is strongly antagonistic to Uniformity: if a bibliometric satisfies Uniformity, then it does not satisfy 2-Gradedness and if it satisfies 2-Gradedness, then it does not satisfy Uniformity. Yet, 2-Gradedness is not the negation of Uniformity: a ranking can violate both Uniformity and 2-Gradedness.

Theorem 4 *The relation $\succsim$ satisfies Independence (A4), 2-Gradedness (A11) and Lower Bound (A3) if and only if the authors are ranked according to the number of papers with at least α citations, that is $f \succsim g \iff \#_\alpha(f) \geq \#_\alpha(g)$.*

Proof. Clearly, the relation $\succsim_\alpha$ satisfies Independence, 2-Gradedness and Lower Bound. We now prove the converse. Among all authors f such that $f \succ g$ for some g, choose one such that $\#(f)$ is minimal (by Non-Triviality, it exists). We denote this author by f_*. By Independence and CDNH, $\#(f_*) \cdot \mathbf{1}_{M(f_*)} \succsim f_*$. By construction, for any $x, y < \#(f_*)$, we have $x \cdot \mathbf{1}_{M(f_*)} \sim y \cdot \mathbf{1}_{M(f_*)} \prec \#(f_*) \cdot \mathbf{1}_{M(f_*)}$. Suppose now that $\#(f_*) > 1$. This leads to a contradiction. Indeed, Independence and $(\#(f_*) - 1) \cdot M(f_*) + \mathbf{1}_{M(f_*)} \succ (\#(f_*) - 1) \cdot M(f_*)$ yield $\mathbf{1}_{M(f_*)} \succ \mathbf{0}$ which, again by Independence, yields $(\#(f_*) - 2) \cdot M(f_*) + \mathbf{1}_{M(f_*)} \succ (\#(f_*) - 2) \cdot M(f_*)$. But this is not true and it therefore implies $\#(f_*) = 1$.

Choose now the smallest x such that $\mathbf{1}_x \succ \mathbf{0}$. Call it α. By construction and Lower Bound, for any $x < \alpha$, $\mathbf{1}_x \sim \mathbf{0} \prec \mathbf{1}_\alpha$. Since $\mathbf{1}_x = \mathbf{1}_x + \mathbf{0} \sim \mathbf{0}$ for any $x < \alpha$, we have, by Independence, $f + \mathbf{1}_x \sim f + \mathbf{0} = f$, for any f. Repeatedly using this reasoning, we find

$$f \sim g \ \forall f, g \text{ s.t. } g(x) \leq f(x) \ \forall x < \alpha \text{ and } f(x) = g(x) \ \forall x \geq \alpha. \tag{9}$$

As we have seen, $\mathbf{1}_\alpha = \mathbf{0} + \mathbf{1}_\alpha \succ \mathbf{0}$. So, by Independence, $2 \cdot \mathbf{1}_\alpha = \mathbf{1}_\alpha + \mathbf{1}_\alpha \succ \mathbf{1}_\alpha$. Repeatedly using the same argument, $x > y$ implies $x \cdot \mathbf{1}_\alpha \succ y \cdot \mathbf{1}_\alpha$. Using this, CDNH, 2-Gradedness and Independence, we have, for any x, $x \cdot \mathbf{1}_\alpha \sim g$ whenever $\#_\alpha(g) = x$ and $g(y) = 0 \ \forall y < \alpha$. Combining this with (9) yields $f \succsim g$ if and only if $\#_\alpha(f) \geq \#_\alpha(g)$. □

We now check the independence of the conditions of Theorem 4. Each of the following bibliometric rankings satisfies all conditions of Theorem 4 but one.

Independence. Let $\alpha \in \mathbb{N}$. $\mathbf{1}_\alpha \succ \mathbf{0}$, $f \sim \mathbf{1}_\alpha$ if $\#_\alpha(f) > 0$ and $f \sim \mathbf{0}$ if $\#_\alpha(f) = 0$.

2-Gradedness. $f \succsim g$ iff $\sum_{x \geq \alpha} x f(x) \geq \sum_{x \geq \alpha} x f(x)$.

Lower Bound. Define $\succsim$ by

- for all $f, g \neq \mathbf{0}$, $f \succsim g$ iff $\#_\alpha(f) \geq \#_\alpha(g)$;
- for all $f \neq \mathbf{0}$, $\mathbf{0} \succ f$.

3.5 The h-index

An author has h-index (or h-number) x if x of his/her papers have at least x citations each, and the other $(\#(f) - x)$ papers have no more than x citations each. The h-index of author f is denoted by $h(f)$. The ranking based on the h-index, denoted by $\succsim_h$, is defined by $f \succsim_h g$ iff $h(f) \geq h(g)$. This ranking is quite different from those we have seen so far; indeed, it does not satisfy Independence and even not Weak Independence, as we now show. Let $f = 4 \cdot \mathbf{1}_4$ and $g = 3 \cdot \mathbf{1}_6$. Then $h(f) = 4$ and $h(g) = 3$ so that $f \succ_h g$. We also have $h(f + g) = 4$ and $h(g + g) = 6$ so that $f + g \prec_h g + g$, thereby contradicting Weak Independence. In this example, using the h-index, we find f better than g. Then, a few months later, both authors have published 3 additional papers, each with 3 citations and then, suddenly, we find g better than f. This is quite difficult to motivate and shows that the ranking based on the h-index is in many circumstances probably not reasonable.

Since $\succsim_h$ does not satisfy Weak Independence, we need many new axioms to characterize it.

A 12 Strong 2-Gradedness. *For all $x < x' < x''$ and all $f \in X$, $f + \mathbf{1}_x \sim f + \mathbf{1}_{x'}$ or $f + \mathbf{1}_{x'} \sim f + \mathbf{1}_{x''}$.*

The idea underlying this condition is the same as for 2-Gradedness. The main difference is that 2-Gradedness applies only to authors with a single publication whereas Strong 2-Gradedness applies to all authors. Independence and 2-Gradedness imply Strong 2-Gradedness.

A 13 Condition Z. $f + \mathbf{1}_0 \precsim f$.

Publishing papers with zero citations cannot improve your position in the ranking. This is a very mild axiom.

A 14 Strong Non-Triviality. *For all $x \neq y$, $x \cdot \mathbf{1}_x \not\sim y \cdot \mathbf{1}_y$.*

Suppose $x > y$. Then $x \cdot \mathbf{1}_x$ and $y \cdot \mathbf{1}_y$ are two very different authors: the former has more publications than the latter and each publication of the former is more cited than any publication of the latter. It is then reasonable to impose that these two authors be not considered as equivalent (unless one wishes a ranking with few classes).

A 15 Condition H. *$f + \mathbf{1}_x \succ f + \mathbf{1}_{x-1}$ and $g + \mathbf{1}_{x+1} \succ g + \mathbf{1}_x$ implies $f + \mathbf{1}_x \sim g$.*

The rationale for this condition is extremely unclear but it seems to be an essential feature of the ranking based on the h-index. Among all axioms that we will use to characterize the ranking $\succ_h$, Condition H is the only one that is not satisfied by any of the other rankings we have seen so far.

A 16 Weak OPOEO. $\mathbf{1}_1 + \mathbf{1}_1 \sim \mathbf{1}_1$.

This is a weakening of OPOEO (A9). It also favours quality over quantity, but in a very mild way: only for publications with one citation.

Theorem 5 *The bibliometric ranking $\succsim$ satisfies Weak Publication Monotonicity (A6), Strong 2-Gradedness (A11), Weak OPOEO (A16), Strong Non-Triviality (A14), Condition H (A15) and Condition Z (A13) if and only if the authors are ranked according to their h-index, that is $f \succsim g \iff h(f) \geq h(g)$.*

Proof. Clearly, the relation $\succsim_h$ satisfies Weak Publication monotonicity, Strong 2-Gradedness, Weak OPOEO, Strong Non-Triviality, Condition H and Condition Z. We now turn to the converse implication.

We first prove that

$$x \geq x' \Rightarrow f + \mathbf{1}_x \succsim f + \mathbf{1}_{x'} \tag{10}$$

Suppose, on the contrary, $f + \mathbf{1}_x \prec f + \mathbf{1}_{x'}$. Strong 2-Gradedness then implies $f + \mathbf{1}_0 \sim f + \mathbf{1}_{x'}$. By Condition Z and Weak Publication Monotonicity, $f + \mathbf{1}_0 \sim f$ and, by transitivity, $f + \mathbf{1}_{x'} \sim f$. So, $f + \mathbf{1}_x \prec f$. This contradicts Weak Publication Monotonicity. So, (10) must hold.

We now prove that all authors with h-index equal to 0 are equivalent to $\mathbf{0}$.

An author f has h-index equal to zero if and only if $f = x \cdot \mathbf{1}_0$, with $x \geq 0$. It is then clear, by Condition Z and Weak Publication Monotonicity, that $f \sim \mathbf{0}$.

We then prove that all authors with h-index equal to 1 *are equivalent to* $\mathbf{1}_1$ *and strictly better than* $\mathbf{0}$.

We distinguish three cases:

(0) $f = \mathbf{1}_y$ $(y \geq 1)$**.** By Strong Non-Triviality, Condition Z and Weak Publication Monotonicity, $\mathbf{1}_1 \succ \mathbf{1}_0$. By Strong 2-Gradedness and (10), $\mathbf{1}_1 \succ \mathbf{1}_0$ implies $\mathbf{1}_2 \sim \mathbf{1}_1$ and, similarly, $\mathbf{1}_y \sim \mathbf{1}_1$, $\forall y > 1$.

(1) $f = x \cdot \mathbf{1}_1$ $(x \geq 2)$**.** Suppose $\mathbf{1}_1 + \mathbf{1}_2 \succ \mathbf{1}_1 + \mathbf{1}_1$. We then also have $\mathbf{1}_2 + \mathbf{1}_1 \succ \mathbf{1}_2 + \mathbf{1}_0$ (by Condition Z, Weak Publication Monotonicity, Weak OPOEO and case (0)). So, by Condition H, $\mathbf{1}_2 + \mathbf{1}_1 \sim \mathbf{1}_1$. A contradiction. This implies $\mathbf{1}_1 + \mathbf{1}_2 \sim \mathbf{1}_1 + \mathbf{1}_1$ and, by Weak OPOEO,

$$\mathbf{1}_1 + \mathbf{1}_2 \sim \mathbf{1}_1. \tag{11}$$

By Strong Non-Triviality and (11), $2 \cdot \mathbf{1}_2 \succ \mathbf{1}_1 \sim \mathbf{1}_1 + \mathbf{1}_2$. Put differently,

$$\mathbf{1}_2 + \mathbf{1}_2 \succ \mathbf{1}_1 + \mathbf{1}_2. \tag{12}$$

Suppose $x \cdot \mathbf{1}_1 \succ (x-1) \cdot \mathbf{1}_1$ for some $x > 1$. Using Condition Z and Weak Publication Monotonicity, this amounts to $(x-1) \cdot \mathbf{1}_1 + \mathbf{1}_1 \succ (x-1) \cdot \mathbf{1}_1 + \mathbf{1}_0$. Combining this with (12) and Condition H yields $x \cdot \mathbf{1}_1 \sim \mathbf{1}_2$. But we know from higher that $\mathbf{1}_2 \sim \mathbf{1}_1$. So, $x \cdot \mathbf{1}_1 \sim \mathbf{1}_1$. This contradiction proves that $x \cdot \mathbf{1}_1 \sim (x-1) \cdot \mathbf{1}_1$ for all $x > 1$. So, $x \cdot \mathbf{1}_1 \sim \mathbf{1}_1$ for all $x \geq 1$.

(2) $f = \mathbf{1}_y + x \cdot \mathbf{1}_1$ $(x \geq 1, y \geq 1)$**.** Suppose now $\mathbf{1}_y + x \cdot \mathbf{1}_1 \succ \mathbf{1}_y + (x-1) \cdot \mathbf{1}_1$ with $x \geq 1, y \geq 1$. Then, by Condition Z and Weak Publication Monotonicity, $\mathbf{1}_y + x \cdot \mathbf{1}_1 \succ \mathbf{1}_y + (x-1) \cdot \mathbf{1}_1 + \mathbf{1}_0$. Combining this with (12) and Condition H yields $\mathbf{1}_y + x \cdot \mathbf{1}_1 \sim \mathbf{1}_2$. But we have already seen (case (0)) that $\mathbf{1}_2 \sim \mathbf{1}_y$ and, by Weak Publication Monotonicity, $\mathbf{1}_y \precsim \mathbf{1}_y + (x-1) \cdot \mathbf{1}_1$. Hence $\mathbf{1}_y + x \cdot \mathbf{1}_1 \precsim \mathbf{1}_y + (x-1) \cdot \mathbf{1}_1$. This is a contradiction and therefore implies $\mathbf{1}_y + x \cdot \mathbf{1}_1 \sim \mathbf{1}_y + (x-1) \cdot \mathbf{1}_1$ for all $x \geq 1, y \geq 1$. This combined with case (0) leads to $\mathbf{1}_y + x \cdot \mathbf{1}_1 \sim \mathbf{1}_1$ for all $x \geq 1, y \geq 1$.

If an author has some publications with zero citations, Condition Z and Weak Publication Monotonicity imply that those publications can be ignored and, so, we can always use one of the cases (0–2). This completes the proof that $h(f) = 1 \Rightarrow f \sim \mathbf{1}_1$.

We then prove that all authors with h-index equal to 2 *are equivalent to* $2 \cdot \mathbf{1}_2$ *and strictly better than* $\mathbf{1}_1$.

We distinguish three cases:

(0) $f = \mathbf{1}_x + \mathbf{1}_y$ $(2 \leq x \leq y)$**.** By (10), $f \succsim 2 \cdot \mathbf{1}_2$ and by Strong Non-Triviality, $f \succ \mathbf{1}_1$. Because $h(\mathbf{1}_2 + \mathbf{1}_1) = 1$, we have $\mathbf{1}_2 + \mathbf{1}_2 \succ \mathbf{1}_2 + \mathbf{1}_1$. So, by

Strong 2-Gradedness, $\mathbf{1}_2 + \mathbf{1}_x \sim \mathbf{1}_2 + \mathbf{1}_2$. Because $h(\mathbf{1}_1 + \mathbf{1}_x) = 1$, we have $\mathbf{1}_2 + \mathbf{1}_x \succ \mathbf{1}_1 + \mathbf{1}_x$. So, by Strong 2-Gradedness, $\mathbf{1}_y + \mathbf{1}_x \sim \mathbf{1}_2 + \mathbf{1}_x$. By transitivity, $\mathbf{1}_y + \mathbf{1}_x \sim \mathbf{1}_2 + \mathbf{1}_2$.

(1) $f = z \cdot \mathbf{1}_1 + \mathbf{1}_x + \mathbf{1}_y$ $(z \geq 1, 2 \leq x \leq y)$**.** Let $g = f - \mathbf{1}_1$ and suppose $g + \mathbf{1}_1 \succ g$. By Condition Z and Weak Publication Monotonicity, $g + \mathbf{1}_1 \succ g + \mathbf{1}_0$. Because $h(\mathbf{1}_2 + \mathbf{1}_1) = 1$ and by Strong Non-Triviality, we know that $\mathbf{1}_2 + \mathbf{1}_2 \succ \mathbf{1}_2 + \mathbf{1}_1$. Then Condition H implies $f = g + \mathbf{1}_1 \sim \mathbf{1}_2$. This is a contradiction. Indeed, we know from case (0) and Weak Publication Monotonicity that $f \succsim 2 \cdot \mathbf{1}_2$. This contradiction implies $f \sim g$.

(2) $f = z \cdot \mathbf{1}_1 + w \cdot \mathbf{1}_2 + \mathbf{1}_x + \mathbf{1}_y$ $(z \geq 0, w \geq 1, 2 \leq x \leq y)$**.** Let $g = f - \mathbf{1}_2$ and suppose $g + \mathbf{1}_1 \succ g$. By Condition Z and Weak Publication Monotonicity, $g + \mathbf{1}_1 \succ g + \mathbf{1}_0$. We know by Strong Non-Triviality that $\mathbf{1}_2 + \mathbf{1}_2 \succ \mathbf{1}_2 + \mathbf{1}_1$. Then Condition H implies $g + \mathbf{1}_1 \sim \mathbf{1}_2$. This is a contradiction. Indeed, we know from case (0) and Weak Publication Monotonicity that $g + \mathbf{1}_1 \succsim 2 \cdot \mathbf{1}_2$. This contradiction implies $g + \mathbf{1}_1 \sim g$.

Suppose now $f = g + \mathbf{1}_2 \succ g$. So, $g + \mathbf{1}_2 \succ g + \mathbf{1}_1$. We also have $\mathbf{1}_0 + \mathbf{1}_1 \succ \mathbf{1}_0 + \mathbf{1}_0$. Then, by Condition H, $\mathbf{1}_0 + \mathbf{1}_1 \sim g$. But we know from case (0) and Weak Publication Monotonicity that $g \succsim 2 \cdot \mathbf{1}_2$. This contradiction implies $f \sim g$.

If $w = 1$, we know from cases (0–1) that $g \sim 2 \cdot \mathbf{1}_2$. So, $f \sim 2 \cdot \mathbf{1}_2$.

If $w > 1$, we find by induction that $g \sim 2 \cdot \mathbf{1}_2$. So, $f \sim 2 \cdot \mathbf{1}_2$.

We then prove that all authors with h-index equal to 3 *are equivalent to* $3 \cdot \mathbf{1}_3$ *and strictly better than* $2 \cdot \mathbf{1}_2$.

We distinguish four cases:

(0) $f = \mathbf{1}_x + \mathbf{1}_y + \mathbf{1}_z$ $(3 \leq x \leq y \leq z)$**.** By (10), $f \succsim 3 \cdot \mathbf{1}_3$ and by Strong Non-Triviality, $f \succ 2 \cdot \mathbf{1}_2$. Because $h(2 \cdot \mathbf{1}_3 + \mathbf{1}_2) = 2$, we have $3 \cdot \mathbf{1}_3 \succ 2 \cdot \mathbf{1}_3 + \mathbf{1}_2$. So, by Strong 2-Gradedness, $2 \cdot \mathbf{1}_3 + \mathbf{1}_x \sim 3 \cdot \mathbf{1}_3$. Because $h(\mathbf{1}_x + \mathbf{1}_3 + \mathbf{1}_2) = 2$, we have $\mathbf{1}_x + 2 \cdot \mathbf{1}_3 \succ \mathbf{1}_x + \mathbf{1}_3 + \mathbf{1}_2$. So, by Strong 2-Gradedness, $\mathbf{1}_x + \mathbf{1}_y + \mathbf{1}_3 \sim \mathbf{1}_x + 2 \cdot \mathbf{1}_3$. Because $h(\mathbf{1}_x + \mathbf{1}_y + \mathbf{1}_2) = 2$, we have $\mathbf{1}_x + \mathbf{1}_y + \mathbf{1}_3 \succ \mathbf{1}_x + \mathbf{1}_y + \mathbf{1}_2$. So, by Strong 2-Gradedness, $\mathbf{1}_x + \mathbf{1}_y + \mathbf{1}_z \sim \mathbf{1}_x + \mathbf{1}_y + \mathbf{1}_3$. By transitivity, $\mathbf{1}_x + \mathbf{1}_y + \mathbf{1}_z \sim 3 \cdot \mathbf{1}_3$.

(1) $f = z \cdot \mathbf{1}_1 + \mathbf{1}_x + \mathbf{1}_y + \mathbf{1}_w$ $(z \geq 1, 3 \leq x \leq y \leq w)$**.** Let $g = f - \mathbf{1}_1$ and suppose $g + \mathbf{1}_1 \succ g$. By Condition Z and Weak Publication Monotonicity, $g + \mathbf{1}_1 \succ g + \mathbf{1}_0$. Because $h(\mathbf{1}_2 + \mathbf{1}_1) = 1$ and by Strong Non-Triviality, we know that $\mathbf{1}_2 + \mathbf{1}_2 \succ \mathbf{1}_2 + \mathbf{1}_1$. Then Condition H implies $f = g + \mathbf{1}_1 \sim \mathbf{1}_2$. This is a contradiction. Indeed, we know from case (0) and Weak Publication Monotonicity that $f \succsim 3 \cdot \mathbf{1}_3$. This contradiction implies $f \sim g$.

(2) $f = z \cdot \mathbf{1}_1 + p \cdot \mathbf{1}_2 + \mathbf{1}_x + \mathbf{1}_y + \mathbf{1}_w$ $(z \geq 0, p \geq 1, 3 \leq x \leq y \leq w)$**.** Let $g = f - \mathbf{1}_2$ and suppose $g + \mathbf{1}_1 \succ g$. By Condition Z and Weak Publication Monotonicity, $g + \mathbf{1}_1 \succ g + \mathbf{1}_0$. We know by Strong Non-Triviality that

$\mathbf{1}_2 + \mathbf{1}_2 \succ \mathbf{1}_2 + \mathbf{1}_1$. Then Condition H implies $g + \mathbf{1}_1 \sim \mathbf{1}_2$. This is a contradiction. Indeed, we know from case (0) and Weak Publication Monotonicity that $g + \mathbf{1}_1 \succsim 3 \cdot \mathbf{1}_3$. This contradiction implies $g + \mathbf{1}_1 \sim g$.

Suppose now $f = g + \mathbf{1}_2 \succ g$. So, $g + \mathbf{1}_2 \succ g + \mathbf{1}_1$. We also have $\mathbf{1}_0 + \mathbf{1}_1 \succ \mathbf{1}_0 + \mathbf{1}_0$. Then, by Condition H, $\mathbf{1}_0 + \mathbf{1}_1 \sim g$. But we know from case (0) and Weak Publication Monotonicity that $g \succsim 3 \cdot \mathbf{1}_3$. This contradiction implies $f \sim g$.

If $p = 1$, we know from cases (0–1) that $g \sim 3 \cdot \mathbf{1}_3$. So, $f \sim 3 \cdot \mathbf{1}_3$.

If $p > 1$, we find by induction that $g \sim 3 \cdot \mathbf{1}_3$. So, $f \sim 3 \cdot \mathbf{1}_3$.

(3) $f = z \cdot \mathbf{1}_1 + p \cdot \mathbf{1}_2 + q \cdot \mathbf{1}_3 + \mathbf{1}_x + \mathbf{1}_y + \mathbf{1}_w$ $(z \geq 0, p \geq 0, q \geq 1, 3 \leq x \leq y \leq w)$**.**
Let $g = f - \mathbf{1}_3$ and suppose $g + \mathbf{1}_1 \succ g$. By Condition Z and Weak Publication Monotonicity, $g + \mathbf{1}_1 \succ g + \mathbf{1}_0$. We know by Strong Non-Triviality that $\mathbf{1}_2 + \mathbf{1}_2 \succ \mathbf{1}_2 + \mathbf{1}_1$. Then Condition H implies $g + \mathbf{1}_1 \sim \mathbf{1}_2$. This is a contradiction. Indeed, we know from case (0) and Weak Publication Monotonicity that $g + \mathbf{1}_1 \succsim 3 \cdot \mathbf{1}_3$. This contradiction implies $g + \mathbf{1}_1 \sim g$.

Suppose now $g + \mathbf{1}_2 \succ g$. So, $g + \mathbf{1}_2 \succ g + \mathbf{1}_1$. We also have $\mathbf{1}_0 + \mathbf{1}_1 \succ \mathbf{1}_0 + \mathbf{1}_0$. Then, by Condition H, $\mathbf{1}_0 + \mathbf{1}_1 \sim g$. But we know from case (0) and Weak Publication Monotonicity that $g \succsim 3 \cdot \mathbf{1}_3$. This contradiction implies $g + \mathbf{1}_2 \sim g$.

Suppose now $f = g + \mathbf{1}_3 \succ g$. So, $g + \mathbf{1}_3 \succ g + \mathbf{1}_2$. We also have $\mathbf{1}_2 + \mathbf{1}_2 \succ \mathbf{1}_2 + \mathbf{1}_1$. Then, by Condition H, $\mathbf{1}_2 + \mathbf{1}_2 \sim g$. But we know from case (0) and Weak Publication Monotonicity that $g \succsim 3 \cdot \mathbf{1}_3$. This contradiction implies $f \sim g$.

If $q = 1$, we know from cases (0–2) that $g \sim 3 \cdot \mathbf{1}_3$. So, $f \sim 3 \cdot \mathbf{1}_3$.

If $q > 1$, we find by induction that $g \sim 3 \cdot \mathbf{1}_3$. So, $f \sim 3 \cdot \mathbf{1}_3$.

The proof goes on treating all authors with h-index 4, then 5, 6, 7 and son on. □

We now check the independence of the conditions of Theorem 5. Each of the following bibliometric rankings satisfies all conditions of Theorem 5 but one.

Weak Publication Monotonicity. Let $\succsim$ be defined as follows.

- $Mf \neq 0$ implies $f \succ \mathbf{0}$.
- $Mf \neq 0 \neq Mg$ implies $f \succsim g \iff h(f) \geq h(g)$.
- $Mf = 0$ and $f \neq \mathbf{0}$ implies $\mathbf{0} \succsim f$.

Strong 2-Gradedness. Let $\succsim$ be defined as follows.

- $x > y$ implies $x \cdot \mathbf{1}_x \succ y \cdot \mathbf{1}_y$.
- $\mathbf{0} \prec \mathbf{1}_1 \prec \mathbf{1}_2 \prec 2 \cdot \mathbf{1}_2$.
- $h(f) = x \neq 1$ implies $f \sim x \cdot \mathbf{1}_x$.

- $f = x \cdot \mathbf{1}_0 + y \cdot \mathbf{1}_1$ with $x \geq 0$ and $y \geq 1$ implies $f \sim \mathbf{1}_1$.
- $f = x \cdot \mathbf{1}_0 + z \cdot \mathbf{1}_1 + \mathbf{1}_y$ with $x \geq 0$, $z \geq 0$ and $y \geq 2$ implies $f \sim \mathbf{1}_2$.

Strong Non-Triviality. Let $\succsim$ be defined as follows.

- $\&(f), \&(g) \geq 1$ implies $f \sim g \succ \mathbf{0}$.
- $\&(f) = 0$ implies $f \sim \mathbf{0}$.

Condition H. Let $\succsim$ be defined as follows.

- $x > y$ implies $x \cdot \mathbf{1}_x \succ y \cdot \mathbf{1}_y$.
- $h(f) = x \neq 1$ implies $f \sim x \cdot \mathbf{1}_x$.
- $f = x \cdot \mathbf{1}_0 + \mathbf{1}_y$ with $x \geq 0$ and $y \geq 1$ implies $f \sim \mathbf{1}_1$.
- $f = x \cdot \mathbf{1}_0 + y \cdot \mathbf{1}_1$ with $x \geq 0$ and $y \geq 1$ implies $f \sim \mathbf{1}_1$.
- $f = x \cdot \mathbf{1}_0 + z \cdot \mathbf{1}_1 + \mathbf{1}_y$ with $x \geq 0$, $z \geq 1$ and $y \geq 2$ implies $f \sim 2 \cdot \mathbf{1}_2$.

Condition Z. Let $\succsim$ be the same ranking as $\succsim_h$ except that the equivalence class of $\mathbf{0}$ is split into two consecutive equivalence classes. The lowest one contains $\mathbf{0}$. The highest one contains all other authors with h-index equal to zero.

Weak OPOEO. Let $\succsim$ be defined as follows.

- $x > y$ implies $x \cdot \mathbf{1}_x \succ y \cdot \mathbf{1}_y$.
- $h(f) = x \neq 1$ implies $f \sim x \cdot \mathbf{1}_x$.
- $f = x \cdot \mathbf{1}_0 + \mathbf{1}_y$ with $x \geq 0$ and $y \geq 1$ implies $f \sim \mathbf{1}_1$.
- $f = x \cdot \mathbf{1}_0 + z \cdot \mathbf{1}_1 + \mathbf{1}_y$ with $x \geq 0$, $z \geq 1$ and $y \geq 1$ implies $f \sim 2 \cdot \mathbf{1}_2$.

4 Discussion

4.1 Summary

As we have shown in the introduction, there is no objectively best or true ranking. According to the context and the aim of the person (or institution) using a ranking, one may use different rankings. How can then our work help this person or user ? If the user knows what he wants to do with the ranking, then it should be possible for him/her to decide, for each axiom, whether it is desirable or not. Then, given this list of desirable axioms, the user just has to find a ranking that satisfies them all. In order to help him in this task, we have summarized our results in Table 1. This table is of course not complete: there are many more rankings and we do not know yet what their characteristic properties are. Further research should help complete this table. It should also help ranking designers better understand what users need. It should therefore help them design new rankings that fit the needs of the users.

Note that a ranking satisfying all axioms that a user finds desirable does not necessarily exist. For instance, there is no bibliometric ranking satisfying Uniformity and 2-Gradedness.

	LB	I	WI	WPM	OI0	A	OPOEO	U	2G	S2G	Z	SNT	H	WOPOEO
$\succsim_\#$	Y	Y	y	y	Y	n	n	y	y	y	n	y	n	n
$\succsim_\&$	Y	y	y	y	n	Y	n	y	n	n	y	y	n	n
$\succsim_M$	Y	n	Y	y	n	n	Y	Y	n	n	y	y	n	y
$\succsim_\alpha$	Y	Y	y	y	n	n	n	n	Y	y	y	n	n	y[1]
$\succsim_h$	y	n	n	Y	n	n	n	n	y	Y	Y	Y	Y	Y

Table 1: The letter 'y' (resp. 'n') represents an axiom that is satisfied (not satisfied). Characterizing axioms are indicated by 'Y'.
LB = Lower Bound; I = Independence; WI = Weak Independence; WPM = Weak Publication Monotonicity; A = Additivity; U = Uniformity; 2G = 2-Gradedness; S2G = Strong 2-Gradedness; Z = Condition Zero; SNT = Strong Non-Triviality; H = Condition H; WOPOEO = Weak OPOEO.
[1] The ranking $\succsim_\alpha$ satisfies Weak OPOEO if $\alpha > 1$.

4.2 Index vs ranking

In this paper, we have axiomatically characterized several rankings. All of them are obtained by first computing an index and then ranking the authors by decreasing value of their index. It is important to note that we did not characterize the corresponding indices, but only the rankings. In order to characterize indices, we would need stronger or more axioms. Indeed, consider the h-index and the squared h-index. These are two different indices. But the rankings derived from those two indices are the same. So, the axioms characterizing the ranking based on the h-index are the same as those characterizing the ranking based on the squared h-index. Actually, the axioms characterizing $\succsim_h$ characterize all rankings that are based on an increasing transformation (like the square) of the h-index. Put differently, the axioms characterizing the ranking based on the h-index characterize the family of all indices that are related to the h-index by an increasing transformation. So, if we want to characterize a single member of this family, we need stronger conditions.

In this paper, we have chosen to work with rankings and not indices because, for most people, indices are not an end but just a way to derive rankings. Of course, it does sometimes make sense to consider an index *per se*; in that case, it would be interesting to know which axioms characterize it. This deserves more research.

4.3 Future research

In addition to the research questions mentioned above, let us enumerate a few other directions for future research.

In this paper, an author f is described by his publication/citation records and we consider that this is given. But which publications shall be counted? Books, articles in peer reviewed journals, all articles, ...? And which citations shall be counted? There is a vast literature on this topic. It would be interesting to see whether some rankings or some axioms make more sense when, e.g., all articles are counted.

Papers with multiple authors are very common and many researchers advocate that such papers should not weight as much as single-author papers. In this paper we did not address this issue. It is not a weakness of our paper (we just analyzed some popular rankings, that do not take multiple authors into account) but a weakness of many rankings. An axiomatic analysis of rankings differentiating between single and multiple authors would be very useful.

In most rankings, implicitely, the quality of a paper is judged from the number of citations it gets. But the quality of a paper can also be judged from the journal where it is published and from the journals that cite it. An axiomatic analysis of rankings taking journals quality into account would also be very interesting.

Acknowledgements

I wish to thank Patrice Perny that introduced me to this topic.

References

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

L. Egghe. Theory and practise of the g-index. *Scientometrics*, 69:131–152, 2006.

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

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

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

A. Sidiropoulos, D. Katsaros, and Y. Manolopoulos. Generalized hirsch h-index for disclosing latent facts in citation networks. *Scientometrics*, 72(2):253–280, 2007.

A. F. J. van Raan. Comparison of the hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*, 67:491–502, 2006.