Journal of Informetrics 7 (2013) 792–802

Contents lists available at ScienceDirect

Journal of Informetrics

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

Scientific impact assessment cannot be fair

Marek Gagolewski[a,b,*]

[a] Systems Research Institute, Polish Academy of Sciences, ul. Newelska 6, 01-447 Warsaw, Poland
[b] Faculty of Mathematics and Information Science, Warsaw University of Technology, ul. Koszykowa 75, 00-662 Warsaw, Poland

ARTICLE INFO

Article history:
Received 25 March 2013
Received in revised form 1 July 2013
Accepted 2 July 2013

Keywords:
Impact functions
Aggregation
Decision making
Preference modeling
Hirsch's *h*-index
Scientometrics
Bibliometrics

ABSTRACT

In this paper we deal with the problem of aggregating numeric sequences of arbitrary length that represent e.g. citation records of scientists. Impact functions are the aggregation operators that express as a single number not only the quality of individual publications, but also their author's productivity.

We examine some fundamental properties of these aggregation tools. It turns out that each impact function which always gives indisputable valuations must necessarily be trivial. Moreover, it is shown that for any set of citation records in which none is dominated by the other, we may construct an impact function that gives any a priori-established authors' ordering. Theoretically then, there is considerable room for manipulation in the hands of decision makers.

We also discuss the differences between the impact function-based and the multicriteria decision making-based approach to scientific quality management, and study how the introduction of new properties of impact functions affects the assessment process. We argue that simple mathematical tools like the *h*- or *g*-index (as well as other bibliometric impact indices) may not necessarily be a good choice when it comes to assess scientific achievements.

1. Introduction

Policy managers, decision makers, and scientists across all disciplines show great interest in the development of sensible, just, and transparent assessment methods of individual scientific achievements. At first glance, it may seem that the adoption of any mathematical formula puts an end to discretionary rankings. However, it was the introduction of one particular tool by J.E. Hirsch (2005) that brought new hopes for the fairness of the quality evaluation process. One of the most attractive features – enthusiastically received by the bibliometric community – of the *h*-index (and related indices) is that it expresses as a single number both the quality of individual papers, as well as the overall author's productivity.

However, some of the studies revealed that particular classes of scientific impact indices may easily be manipulated. One can think of at least two kinds of such devious influencing:

1. The first one occurs when a scientist tries to artificially improve his/her position in a ranking. For example, Bartneck and Kokkelmans (2011) as well as Zhivotovsky and Krutowsky (2008) note that the *h*-index may be "inflated" by a clever self-citation pattern.

* Correspondence address: Systems Research Institute, Polish Academy of Sciences, ul. Newelska 6, 01-447 Warsaw, Poland. Tel.: +48 22 38 10 393; fax: +48 22 38 10 105
E-mail address: gagolews@ibspan.waw.pl

1751-1577/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.joi.2013.07.001

2. The second kind occurs when a person in charge of choosing the assessment procedure decides to favor a predefined clique. For example, it has been indicated recently that the generalized Hirsch index is highly sensitive to the application of simple input data transformations, which may be prone to fraudulence (see Cena & Gagolewski, 2013b; Gagolewski & Mesiar, 2012).

The first class of manipulation techniques is related to input data being aggregated (Should we use citations from *Web of Knowledge* rather that *SciVerse Scopus*? Should self-citations be removed? Should we normalize a paper's citations with respect to the number its authors? Should journal quality measures be used to assess the quality of a paper? etc.). The second one, only loosely related to the former, is of our interest in this paper. It concerns the ranking/aggregation tool itself, and assumes that we are working on representative and valid data. In this perspective, the "*Google Scholar h*-index" is the same mathematical tool as the *h*-index normalized for the number of coauthors.

Literally hundreds of studies were performed to examine the behavior of the *h*-, *g*-, and similar indices on real-world data (see e.g. Alonso, Cabrerizo, Herrera-Viedma, & Herrera, 2009; Bornmann & Daniel, 2009; Egghe, 2010). Few of them also considered the analysis of indices' theoretical properties, for example from the axiomatic (e.g. Woeginger, 2008b) or probabilistic/statistical perspective (cf. Nair & Turlach, 2012 or Gagolewski, 2013b). It turns out that equivalent mathematical objects were already known in other scientific domains. For example, Torra and Narukawa (2008) showed that the *h*-index is a Sugeno integral with respect to a counting measure from the fuzzy/monotone measure theory. The others studied similar indices in the context of aggregation theory (cf. e.g. Grabisch, Marichal, Mesiar, & Pap, 2009).

However, the most fundamental questions concerning the aggregation methods of scientific output quality measures still remain open. Is the very nature of the assessment process such that it inevitably produces debatable results? If so, how to show it in a formal manner? When can we rely on the automatically generated valuations? On the other hand, in which cases is there room for manipulation and favoritism in the hands of decision makers?

The answer to these questions is crucial, because automated decision making is becoming more and more popular nowadays. It is still hoped that this form of assessment process may become the cure for not-rare cases of disappointment with the subjectivity of the "human factor".

The paper is organized as follows. Section 2 recalls the notion of an impact function and shows its connection to a particular (binary) preordering relation defined on the set of vectors representing citation records.

In Section 3 we study whether there exists a nontrivial impact function that gives us "noncontroversial" results in "disputable" cases. Moreover, we formally show how the introduction of additional properties modifies results of pairwise comparisons of citation records. It turns out that the most "sensitive" part of an impact function creation is the transformation of a preordering relation (in which there is still some room for indefiniteness) to a total preorder.

In Section 4 we explore the possibility of creating an impact function that generates an arbitrary, preselected ranking of a set of authors. Additionally, we present an illustration concerning a particular class of impact functions (a generalized *h*-index) applied on exemplary, real-world scientometric data set.

Finally, in Section 5 we discuss the implications of the results.

2. Impact functions

In order to study any real-world phenomenon, we have to establish its abstract *model*, preferably in the language of mathematics. Let us assume that some a priori chosen, reliable paper quality measure takes values in $\mathbb{I} = [0, \infty)$. These may of course be non-integers, for example when we consider citations of papers that are normalized with respect to the number of coauthors. Importantly, the values are not bounded from above (and thus cannot be sensibly transformed to a finite-length interval, e.g. [0, 1]).

Moreover, let $\mathbb{I}^{1,2,\cdots}$ denote the set of all sequences (of arbitrary length) with elements in $\mathbb{I}$, i.e. $\mathbb{I}^{1,2,\cdots} = \bigcup_{n=1}^{\infty} \mathbb{I}^n$. Thus, the whole information on an author's output is represented by exactly one element in $\mathbb{I}^{1,2,\cdots}$.

We are interested in constructing an aggregation operator, i.e. a function that maps $\mathbf{x} \in \mathbb{I}^{1,2,\cdots}$ to some $\mathsf{F}(\mathbf{x}) \in \mathbb{I}$, and which reflects the two following "dimensions" of an author's output quality:

(a) quality of his/her papers (his/her ability to write eagerly cited or highly valuated papers),
(b) his/her overall productivity.

Additionally, (c) we have no reason to treat any paper in a special way: only their quality measures should have impact on the results of evaluation, and not how they are ordered (e.g. by the number of coauthors, by time of publication, or how does the author want it). Of course, this is one of the many possible approaches, see e.g. (Grabisch et al., 2009) for the axiomatization of functions which only consider (a).

It is often assumed (see e.g. Franceschini & Maisano, 2011; Gagolewski & Grzegorzewski, 2011; Quesada, 2009, 2010; Rousseau, 2008; Woeginger, 2008a, 2008b, 2008c) that each *impact function* – i.e. an aggregation operator $\mathsf{F} : \mathbb{I}^{1,2,\cdots} \to \mathbb{I}$ to be applied in the assessment process that follows the above-mentioned characteristics – should *at least* be:

(a) nondecreasing with respect to each variable (additional citations received by a paper or an improvement of its quality measure does not result in a decrease in the author's overall evaluation),
(b) arity-monotonic (by publishing a new paper we never decrease the overall valuation of the entity),
(c) symmetric (independent of the order of elements' presentation, i.e. we may always assume that we aggregate vectors that are already sorted).

More formally, condition (a) holds if and only if for each n and $\mathbf{x}, \mathbf{y} \in \mathbb{I}^n$ such that $(\forall i)\ x_i \le y_i$ we have $\mathsf{F}(\mathbf{x}) \le \mathsf{F}(\mathbf{y})$. On the other hand, axiom (b) is fulfilled iff for any $\mathbf{x} \in \mathbb{I}^{1,2,\cdots}$ and $y \in \mathbb{I}$ it holds $\mathsf{F}(\mathbf{x}) \le \mathsf{F}(x_1, \ldots, x_n, y)$. Lastly, requirement (c) holds iff for all n, $\mathbf{x} \in \mathbb{I}^n$, and any permutation σ of the set $\{1, \ldots, n\}$, we have $\mathsf{F}(\mathbf{x}) = \mathsf{F}(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$. It may be shown (see Grabisch et al., 2009), that symmetry is equivalent to stating that for each $\mathbf{x}$ it holds $\mathsf{F}(\mathbf{x}) = \mathsf{F}(x_{\{1\}}, \ldots, x_{\{n\}})$, where $x_{\{i\}}$ denotes the ith largest value from $\mathbf{x}$, i.e. its $(n-i+1)$th order statistic.

From now on, let $\mathcal{I}$ denote the set of all impact functions.

Please, note that in our model, we abstract from the very nature of aggregated data (construction of paper quality measures is a problem of its own). The aggregation operator may be conceived as a kind of "black box", which takes a numeric sequence in input, and gives a single number in its output. This computational tool may be expressed as a simple mathematical formula or e.g. by a very complex computer algorithm.

2.1. (Weak) dominance relation

Let us consider the following binary relation $\trianglelefteq \subseteq \mathbb{I}^{1,2,\cdots} \times \mathbb{I}^{1,2,\cdots}$. For any $\mathbf{x} \in \mathbb{I}^n$ and $\mathbf{y} \in \mathbb{I}^m$ we write $\mathbf{x} \trianglelefteq \mathbf{y}$ (or, equivalently, $(\mathbf{x}, \mathbf{y}) \in \trianglelefteq$) if and only if

$$n \le m \text{ and } x_{\{i\}} \le y_{\{i\}} \quad \text{for } i = 1, \ldots, n.$$

In other words, we say that an author X is (weakly) dominated by an author Y, if X has no more papers than Y and each of the ith most cited paper by X has no more citations than the ith most cited paper by Y. Note that the $m-n$ least cited Y's papers are not taken into account here at all.

Of course, $\trianglelefteq$ is a preorder, i.e. it is:

- reflexive, i.e. $(\forall \mathbf{x})\ \mathbf{x} \trianglelefteq \mathbf{x}$,
- transitive, i.e. $(\forall \mathbf{x}, \mathbf{y}, \mathbf{z})\ \mathbf{x} \trianglelefteq \mathbf{y}$ and $\mathbf{y} \trianglelefteq \mathbf{z}$ implies that $\mathbf{x} \trianglelefteq \mathbf{z}$.

It turns out that there is a strong relationship between the introduced preorder and the impact functions, see Section 2.3. First, however, we should note that there exist pairs of vectors that are *incomparable* with respect to $\trianglelefteq$.

2.2. Illustration

Let us consider the following vectors. They may for example represent citations of papers published by 5 imaginary authors:

- $\mathbf{u} = (10, 9, \ldots, 1)$, *(some upper bound)*
- $\mathbf{a} = (7, 6)$, *(high quality, small productivity)*
- $\mathbf{b} = (4, 4, 4, 4)$, *(moderate quality and productivity)*
- $\mathbf{c} = (3, 3, 3, 2, 1, 0, 0, 0)$, *(small quality but high productivity)*
- $\mathbf{l} = (1, 0)$. *(some lower bound)*

These vectors are depicted in Fig. 1a – they are represented by so-called citation curves. Formally, if $\mathbf{z} \in \mathbb{I}^n$ for some n, then a *citation curve* $C_{\mathbf{z}}(x, y)$ is created by joining the points $((0, z_{\{1\}}), (1, z_{\{1\}}), (1, z_{\{2\}}), (2, z_{\{2\}}), (2, z_{\{3\}}), (3, z_{\{3\}}), \ldots, (n, z_{\{n\}}))$ with line segments.

The Hasse diagram for the preordered set $(\{\mathbf{l}, \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{u}\}, \trianglelefteq)$ is depicted in Fig. 1b. Nodes of the graph represent given vectors, and edges join the nodes for which our (weak) dominance holds. We have $\mathbf{l} \trianglelefteq \mathbf{a}$, $\mathbf{l} \trianglelefteq \mathbf{b}$, $\mathbf{l} \trianglelefteq \mathbf{c}$, and $\mathbf{a} \trianglelefteq \mathbf{u}$, $\mathbf{b} \trianglelefteq \mathbf{u}$, $\mathbf{c} \trianglelefteq \mathbf{u}$. Additionally, by the transitivity of $\trianglelefteq$, as e.g. $\mathbf{l} \trianglelefteq \mathbf{b}$ and $\mathbf{b} \trianglelefteq \mathbf{u}$, we of course have $\mathbf{l} \trianglelefteq \mathbf{u}$.

What is more, note that $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ is the (maximal) independent set with respect to the relation $\trianglelefteq$, i.e. no pair of vectors from $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ is comparable (w.r.t. $\trianglelefteq$). In other words, we cannot state which of $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ is "better" (w.r.t. $\trianglelefteq$) than the others. For example, we have $a_{\{1\}} > c_{\{1\}}$ and $a_{\{2\}} > c_{\{2\}}$, but, on the other hand, $\text{length}(\mathbf{a}) < \text{length}(\mathbf{c})$. That is, it neither holds $\mathbf{a} \trianglelefteq \mathbf{c}$ nor $\mathbf{c} \trianglelefteq \mathbf{a}$.

2.3. Weak dominance and impact functions

The following result was given in (Gagolewski & Grzegorzewski, 2011).

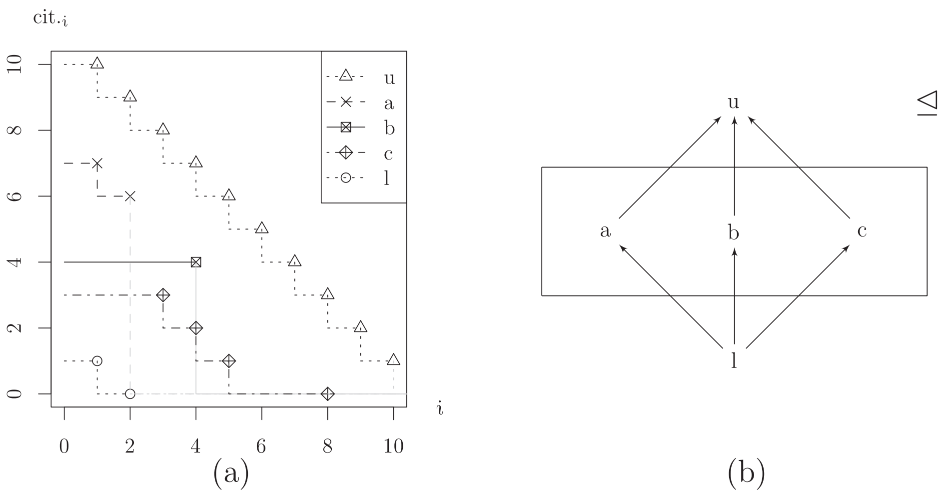

Fig. 1. Citation curves and the Hasse diagram for 5 vectors discussed in Section 2.2

Theorem 1. *Let* $\mathsf{F}:\mathbb{I}^{1,2,\cdots}\to\mathbb{I}$ *be an aggregation operator. Then* F *is symmetric, nondecreasing in each variable and arity-monotonic if and only if for any* $\mathbf{x},\mathbf{y}$ *if* $\mathbf{x}\trianglelefteq\mathbf{y}$, *then* $\mathsf{F}(\mathbf{x})\le\mathsf{F}(\mathbf{y})$.

Therefore, the class of impact functions is equivalent to all preorder preserving mappings between the preordered set $(\mathbb{I}^{1,2,\cdots},\trianglelefteq)$ and $(\mathbb{I},\le)$.

What is crucial, the space $(\mathbb{I},\le)$ is totally ordered (a binary relation $R\subseteq V\times V$ is *total* if and only if for any $v,v'\in V$ it holds vRv' or $v'Rv$ or both). It means that with the use of $\mathsf{F}\in\mathcal{I}$ we lose information on vectors' incomparability.

In fact, the construction of any impact function F may be decomposed into the two following steps:

1. Setting $\trianglelefteq''_{\mathsf{F}}\supseteq\trianglelefteq$ (a total preorder on $\mathbb{I}^{1,2,\cdots}\times\mathbb{I}^{1,2,\cdots}$).
2. As each total preorder $\trianglelefteq''_{\mathsf{F}}$ induces an equivalence relation $\simeq_{\mathsf{F}}$ (such that $\mathbf{x}\simeq_{\mathsf{F}}\mathbf{y}$ iff $\mathbf{x}\trianglelefteq''_{\mathsf{F}}\mathbf{y}$ and $\mathbf{y}\trianglelefteq''_{\mathsf{F}}\mathbf{x}$), we may easily obtain a function that for each vector in $\mathbb{I}^{1,2,\cdots}$ outputs a real number in $\mathbb{I}$. This is done by means of some strictly increasing function φ_{F} defined on the set of equivalence classes of $\simeq_{\mathsf{F}}$.

Thus, $\mathsf{F}\equiv(\trianglelefteq''_{\mathsf{F}},\varphi_{\mathsf{F}})$.

For the purpose of our discussion, we will be only interested in step 1: at this point it should not bother us too much whether e.g. it holds $\mathsf{F}(\mathbf{x})=1$ and $\mathsf{F}(\mathbf{y})=2$, or maybe $\mathsf{F}(\mathbf{x})=0$ and $\mathsf{F}(\mathbf{y})=1000$ ("three times better" is just "better"). Step 2 is therefore performed only to "calibrate" the function, for example to achieve boundary conditions like $\mathsf{F}(n,n,\ldots,n)=n$ or $\mathsf{F}(0)=0$.

In other words, by means of the presented model, we may focus only on how to compare pairs of vectors, i.e. how to establish a "reasonable" relation $\trianglelefteq''_{\mathsf{F}}\supseteq\trianglelefteq$. The same idea is used in computer science: algorithms for sorting/ranking elements of some set rely only on pairwise comparisons of objects.

3. Does there exist a fair impact function?

Let us depart from the dominance relation $\trianglelefteq$. This relation defines the (weakest) set of necessary comparison results. In order to construct an impact function, $\trianglelefteq$ has to be *totalized.*

Fact 1. *Each total preorder* $\trianglelefteq''_{\mathsf{F}}\supseteq\trianglelefteq$ *and a strictly increasing function* $\varphi_{\mathsf{F}}:\mathbb{I}\to\mathbb{I}$ *define a unique impact function.*

Example 1. Let us consider the set of authors $\mathcal{A}=\{(0),(1),(2,1),(1,2),(1,0,0)\}\subset\mathbb{I}^{1,2,\cdots}$. Here are the results of comparisons between each pair of elements in $\mathcal{A}$: "1" means that $\mathbf{x}\trianglelefteq\mathbf{y}$.

$$\trianglelefteq|_{\mathcal{A}}=\begin{bmatrix}\mathbf{x}\backslash\mathbf{y} & (0) & (1) & (2,1) & (1,2) & (1,0,0)\\ (0) & 1 & 1 & 1 & 1 & 1\\ (1) & * & 1 & 1 & 1 & 1\\ (2,1) & * & * & 1 & 1 & *\\ (1,2) & * & * & 1 & 1 & *\\ (1,0,0) & * & * & * & * & 1\end{bmatrix}$$

Note that the "1"s on the diagonal are due to the reflexivity property. "*"s denote the pairs for which we will obtain "0" ($\mathbf{x} \ntrianglelefteq \mathbf{y}$) or "1" after totalization. By the above-mentioned fact, we are free to do anything with "*" (again, $\trianglelefteq$ defines only "must have" comparison results), as long as we are not violating the transitivity property. In this example, we have two incomparable pairs: (2,1) and (1,0,0), as well as (1,2) and (1,0,0) – such situation corresponds to having two "*" on both sides of the diagonal. □

A total preorder $\trianglelefteq''_{\mathsf{F}}$ lets us state whether for a pair ($\mathbf{x}$, $\mathbf{y}$) it either holds $\mathbf{x} \prec_{\mathsf{F}} \mathbf{y}$ ($\mathbf{x}$ is "worse" than $\mathbf{y}$), $\mathbf{x} \simeq_{\mathsf{F}} \mathbf{y}$ (they are equivalent), or $\mathbf{x} \succ_{\mathsf{F}} \mathbf{y}$ ($\mathbf{x}$ is "better" than $\mathbf{y}$). Here is how each possible situation between ($\mathbf{x}$, $\mathbf{y}$) may be resolved in $\trianglelefteq''_{\mathsf{F}}$:

$\mathbf{x} \trianglelefteq \mathbf{y}$	$\mathbf{y} \trianglelefteq \mathbf{x}$	may become in $\trianglelefteq''_{\mathsf{F}}$
1	*	$\prec_{\mathsf{F}}$ or $\simeq_{\mathsf{F}}$
*	1	$\succ_{\mathsf{F}}$ or $\simeq_{\mathsf{F}}$
1	1	$\simeq_{\mathsf{F}}$
*	*	$\prec_{\mathsf{F}}$, $\simeq_{\mathsf{F}}$ or $\succ_{\mathsf{F}}$

3.1. Fairness and $\trianglelefteq$

The question is, of course, how to perform the totalization of $\trianglelefteq$ to do "least harm". Of course, uncomparable vectors are the most problematic here, as for them no preference is even suggested. Perhaps the least controversial decision would be to give them an equal valuation. For example, consider vectors $\mathbf{a}$ and $\mathbf{c}$ from the illustration in Section 2.2: it is questionable whether – in general – stating either $\mathsf{F}(\mathbf{a}) < \mathsf{F}(\mathbf{c})$ or $\mathsf{F}(\mathbf{a}) > \mathsf{F}(\mathbf{c})$ should be preferred.

Thus, we will call an impact function F **fair** if for all $\mathbf{x}, \mathbf{y}$ such that $\neg(\mathbf{x} \trianglelefteq \mathbf{y} \text{ or } \mathbf{y} \trianglelefteq \mathbf{x})$ it holds $\mathsf{F}(\mathbf{x}) = \mathsf{F}(\mathbf{y})$. We have the following result.

Theorem 2. *Each fair impact function is trivial; there exists $c \in \mathbb{I}$ such that for all $\mathbf{x} \neq (0, \ldots, 0)$ we have $\mathsf{F}(\mathbf{x}) = c$.*

Proof. Assume the opposite. Let there exist $\mathbf{x} \neq (0, \ldots, 0)$, $\mathbf{y} \neq (0, \ldots, 0)$ such that $\mathsf{F}(\mathbf{x}) < \mathsf{F}(\mathbf{y})$. Let $\mathbf{z} = (0, \ldots, 0) \in \mathbb{I}^{k+1}$, where $k = \max\{\text{length}(\mathbf{x}), \text{length}(\mathbf{y})\}$. We have $\neg(\mathbf{z} \trianglelefteq \mathbf{y} \text{ or } \mathbf{y} \trianglelefteq \mathbf{z})$, thus, by fairness of F, $\mathsf{F}(\mathbf{z}) = \mathsf{F}(\mathbf{y})$. But also $\neg(\mathbf{z} \trianglelefteq \mathbf{x} \text{ or } \mathbf{x} \trianglelefteq \mathbf{z})$. Therefore, $\mathsf{F}(\mathbf{x}) = \mathsf{F}(\mathbf{y})$, a contradiction. □

Therefore, a fair totalization $\trianglelefteq''_{\mathsf{F}}$ of $\trianglelefteq$ resolves each $(\mathbf{x}, \mathbf{y}) \ntrianglelefteq$, $\mathbf{x}, \mathbf{y} \neq (0, \ldots, 0)$, to $\mathbf{x} \simeq_{\mathsf{F}} \mathbf{y}$. Note that this result holds under only very mild conditions – we only assume that we are given a nondecreasing, symmetric, and arity-monotonic function. We may thus conclude that no impact function gives sensible results with respect to $\trianglelefteq$ – there will always be a set of independent vectors for which one may ask: why are they ranked in a given way and not another?

3.2. Properties of impact functions

One may argue that the problem with the above result is that the class of impact functions is very broad: the relation $\trianglelefteq$ provides too many "degrees of freedom" when resolving it to a total preorder. Hence, in practice we would like to narrow down the class of interesting impact functions by introducing some additional properties.

Formally, a property of impact functions is equivalent to some subset $\mathcal{P}$ of $\mathcal{I}$ such that $\mathcal{P} = \{\mathsf{G} \in \mathcal{I} : \mathsf{G} \text{ fulfills a given property}\}$. Clearly, we should ask ourselves the question: from the point of view of our model, what actually happens when we establish a property?

Let us fix $\mathcal{P} \subseteq \mathcal{I}$. Denote by $\trianglelefteq_{\mathcal{P}} = \bigcap_{\mathsf{F} \in \mathcal{P}} \trianglelefteq''_{\mathsf{F}}$ the set of all "necessary" results of comparisons that are generated by all functions from $\mathcal{P}$. Moreover, let $\vdash_{\mathcal{P}} = (\bigcup_{\mathsf{F} \in \mathcal{P}} \trianglelefteq''_{\mathsf{F}})^c = (\mathbb{I}^{1,2,\cdots} \times \mathbb{I}^{1,2,\cdots}) \setminus (\bigcup_{\mathsf{F} \in \mathcal{P}} \trianglelefteq''_{\mathsf{F}})$ denote the set of results of comparisons that are excluded from all functions from $\mathcal{P}$. Note that $\trianglelefteq_{\mathcal{I}} = \trianglelefteq$ and $\vdash_{\mathcal{I}} = \emptyset$.

For any $\mathcal{P}$, it holds $\trianglelefteq \subseteq \trianglelefteq_{\mathcal{P}}$, $\trianglelefteq_{\mathcal{P}}$ is a preorder, but not necessarily total. On the other hand, $\vdash_{\mathcal{P}}$ is at least antireflexive ($(\forall \mathbf{x})\ \mathbf{x} \nvdash_{\mathcal{P}} \mathbf{x}$) and antisymmetric ($(\forall \mathbf{x}, \mathbf{y})\ \mathbf{x} \nvdash_{\mathcal{P}} \mathbf{y}$ or $\mathbf{y} \nvdash_{\mathcal{P}} \mathbf{x}$). Additionally, $\trianglelefteq_{\mathcal{P}} \cap \vdash_{\mathcal{P}} = \emptyset$.

Example 2 *(Max-based indices).* Let the family of impact functions (i.e. a property) $\mathcal{M}$ be defined as follows. For any $\mathbf{x}, \mathbf{y} \in \mathbb{I}^{1,2,\cdots}$ we have $\mathsf{F}(\mathbf{x}) = \mathsf{F}(\mathbf{y})$ iff $x_{\{1\}} = y_{\{1\}}$. We have $\trianglelefteq_{\mathcal{M}} := \left\{(\mathbf{x}, \mathbf{y}) : x_{\{1\}} \le y_{\{1\}}\right\} \supseteq \trianglelefteq$, and $\vdash_{\mathcal{M}} := \left\{(\mathbf{x}, \mathbf{y}) : x_{\{1\}} > y_{\{1\}}\right\}$. Thus, before introducing the property, we had no knowledge on how to compare e.g. the vectors (3, 2, 1) and (4, 0), i.e. $(3, 2, 1) \ntrianglelefteq (4, 0)$ and $(3, 2, 1) \ntrianglerighteq (4, 0)$. Now, however, we have $(3, 2, 1) \trianglelefteq_{\mathcal{M}} (4, 0)$ and $(3, 2, 1) \ntrianglerighteq_{\mathcal{M}} (4, 0)$. Also, $(3, 2, 1) \dashv_{\mathcal{M}} (4, 0)$. Hence, $(3, 2, 1) \prec_{\mathcal{M}} (4, 0)$. □

Example 3 *(Zero-insensitiveness).* Let $\mathcal{Z}$ denote the set of all impact functions F such that for any $\mathbf{x} \in \mathbb{I}^{1,2,\cdots}$ we have $\mathsf{F}(\mathbf{x}) = \mathsf{F}(\mathbf{x}, 0)$, i.e. those which does not discriminate between uncited and nonexistent papers. Note that e.g. the h-index $\in \mathcal{Z}$. We have $\vdash_{\mathcal{Z}} = \emptyset$ and

$$\trianglelefteq_{\mathcal{Z}} := \trianglelefteq \cup \left\{ (\mathbf{x}, \mathbf{y}) : \begin{array}{lll} x_{\{i\}} = y_{\{i\}} & \text{for} & i = 1, \ldots, \min\{n_x, n_y\}, \\ x_{\{i\}} = 0 & \text{for} & i = \min\{n_x, n_y\} + 1, \ldots, n_x, \\ y_{\{i\}} = 0 & \text{for} & i = \min\{n_x, n_y\} + 1, \ldots, n_y \end{array} \right\}.$$

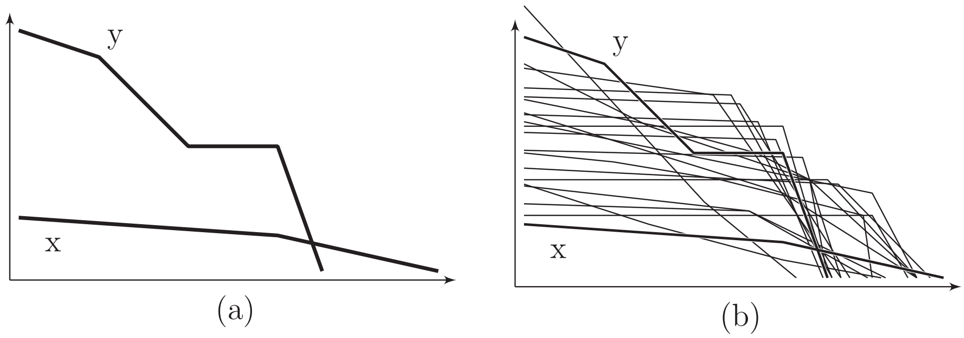

Fig. 2. The effect of context.

For example, we have e.g. $(3, 2) \trianglelefteq (3, 2, 0, 0)$ and $(3, 2) \ntrianglerighteq (3, 2, 0, 0)$, but $(3, 2) \trianglelefteq_{\mathcal{Z}} (3, 2, 0, 0)$ and $(3, 2) \trianglerighteq_{\mathcal{Z}} (3, 2, 0, 0)$, i.e. $(3, 2) \simeq_{\mathcal{Z}} (3, 2, 0, 0)$. □

Example 4 *(Some lower bound).* Let $\mathcal{L}$ denote the set of all impact functions F such that for any $\mathbf{x} \in \mathbb{I}^{1,2,\cdots}$ we have $\mathsf{F}(\mathbf{x}) \geq 3.14$. Here $\trianglelefteq_{\mathcal{L}} = \trianglelefteq$ and $\vdash_{\mathcal{L}} = \emptyset$. We see that this property does not affect the way we order the elements in $\mathbb{I}^{1,2,\cdots}$ and thus is not interesting for us. □

Example 5 *(Degenerated subclass).* Let $\mathcal{D} = \{\mathsf{F} \in \mathcal{I} : \mathsf{F}$ is proportional to the h- or to the g-index$\}$. This property is *degenerated* in the following sense. There exists a total preorder $\trianglelefteq''_{\mathsf{F}} \supseteq \trianglelefteq_{\mathcal{D}}$, $\trianglelefteq''_{\mathsf{F}} \cap \vdash_{\mathcal{D}} = \emptyset$, such that for some φ we have $(\trianglelefteq''_{\mathsf{F}}, \varphi) \notin \mathcal{D}$, cf. Section 2.3. □

Since degenerate properties are not in the scope of our interest, the establishment of a property $\mathcal{P}$ may be viewed as the introduction of the pair $(\trianglelefteq_{\mathcal{P}}, \vdash_{\mathcal{P}})$. This pair may be used to generate the set of impact functions fulfilling $\mathcal{P}$, by constructing total preorders $\trianglelefteq''_{\mathsf{F}} \supseteq \trianglelefteq_{\mathcal{P}}$ such that $\trianglelefteq''_{\mathsf{F}} \cap \vdash_{\mathcal{P}} = \emptyset$.

Thus, $(\trianglelefteq_{\mathcal{P}}, \vdash_{\mathcal{P}})$ tells us (a) between which pairs the relation $\trianglelefteq''_{\mathsf{F}}$ must hold ("1"s, by $\trianglelefteq_{\mathcal{P}}$) and (b) between which it is forbidden ("0"s, through $\vdash_{\mathcal{P}}$). Thus, now the set $(\trianglelefteq_{\mathcal{P}} \cup \vdash_{\mathcal{P}})^c$ represents the pairwise comparison results that we may set freely when constructing a total preorder $\trianglelefteq''_{\mathsf{F}}$ (denoted as "*" in Example 1).

3.3. "Reasonable" impact functions

Now let us assume that a group of experts decided to establish the class of "reasonable" impact functions, $\mathcal{R}$. They are rather not interested in introducing a mathematical formula for how an impact function shall be calculated, but in having full control on the results of pairwise comparisons (see e.g. Marchant, 2009a, 2009b; Quesada, 2011 for some interesting approaches to the construction of bibliometric preorders).

For example, they may agree that (101) is "worse" than (100, 100, . . ., 100), (55.001) is the same as (55.002), (4, 4, 4) is not better than (5, 4, 3), and they have undefined preference for (5, 4, 3, 2, 1, 0, 0, 0, 0) and (10, 5). In fact, they specify their favorite pair $(\trianglelefteq_{\mathcal{R}}, \vdash_{\mathcal{R}})$. As long as they do not use a pre-defined scheme like "let's base our decisions solely on the total number of citations" (which does not take into account author's productivity), it is highly possible that they will obtain $\trianglelefteq_{\mathcal{R}}$ being not a total preorder, and $(\trianglelefteq_{\mathcal{R}} \cup \vdash_{\mathcal{R}})^c \neq \emptyset$. Note that the knowledge represented in $(\trianglelefteq_{\mathcal{R}}, \vdash_{\mathcal{R}})$ may be implemented e.g. as a very complex computer algorithm – contemporarily, this is not problematic at all.

It is easily seen that with $(\trianglelefteq_{\mathcal{R}}, \vdash_{\mathcal{R}})$ we state whether between $\mathbf{x}$ and $\mathbf{y}$ holds either $\prec_{\mathcal{R}}$, $\preceq_{\mathcal{R}}$, $\simeq_{\mathcal{R}}$, $\succeq_{\mathcal{R}}$, $\succ_{\mathcal{R}}$, or $???_{\mathcal{R}}$ (i.e. they are incomparable/we have not preference at all). In order to construct an impact function, $\trianglelefteq_{\mathcal{R}}$ has to be totalized. That is, we have to resolve e.g. $\preceq_{\mathcal{R}}$ to either $\prec_{\mathsf{F}}$ or $\simeq_{\mathsf{F}}$.

$\mathbf{x} \trianglelefteq_{\mathcal{R}} / \vdash_{\mathcal{R}} \mathbf{y}$	$\mathbf{y} \trianglelefteq_{\mathcal{R}} / \vdash_{\mathcal{R}} \mathbf{x}$	Notation	May become in $\trianglelefteq''_{\mathsf{F}}$
1	0	$\prec_{\mathcal{R}}$	$\prec_{\mathsf{F}}$
1	*	$\preceq_{\mathcal{R}}$	$\prec_{\mathsf{F}}$ or $\simeq_{\mathsf{F}}$
1	1	$\simeq_{\mathcal{R}}$	$\simeq_{\mathsf{F}}$
*	1	$\succeq_{\mathcal{R}}$	$\succ_{\mathsf{F}}$ or $\simeq_{\mathsf{F}}$
0	1	$\succ_{\mathcal{R}}$	$\succ_{\mathsf{F}}$
*	*	$???_{\mathcal{R}}$	$\prec_{\mathsf{F}}$, $\simeq_{\mathsf{F}}$ or $\succ_{\mathsf{F}}$
0	0	–	– (impossible)

Here is an intuitive argument why such totalization cannot be done fairly. Fig. 2a represents two vectors between which we possibly would establish $\mathbf{x} \prec_{\mathsf{F}} \mathbf{y}$. However, when we create an impact function, we have to take the whole assessment context into account. In Fig. 2b we depict many other vectors which also play an important role. Assume we observe a sequence $\mathbf{z}_1, \ldots, \mathbf{z}_k$ such that $\mathbf{x} ???_{\mathcal{R}} \mathbf{z}_1 ???_{\mathcal{R}} \ldots \mathbf{z}_k ???_{\mathcal{R}} \mathbf{y}$. In order to be fair, we should substitute $???_{\mathcal{R}}$ for $\simeq_{\mathsf{F}}$. This is in at least

Table 1
Exemplary ways to generate each of the possible rankings for incomparable vectors from Section 2.2

desired ordering	exemplary impact function $\mathsf{F}(\mathbf{x})$
$\mathsf{F}(\mathbf{a})>\mathsf{F}(\mathbf{b})>\mathsf{F}(\mathbf{c})$	$x_{\{1\}}$ (maximal number of citations)
$\mathsf{F}(\mathbf{c})>\mathsf{F}(\mathbf{b})>\mathsf{F}(\mathbf{a})$	n (number of papers)
$\mathsf{F}(\mathbf{b})>\mathsf{F}(\mathbf{a})>\mathsf{F}(\mathbf{c})$	$\sum_i x_i$ (total number of citations)
$\mathsf{F}(\mathbf{b})>\mathsf{F}(\mathbf{c})>\mathsf{F}(\mathbf{a})$	h-index
$\mathsf{F}(\mathbf{a})>\mathsf{F}(\mathbf{c})>\mathsf{F}(\mathbf{b})$	$\max\{x_{\{1\}}, 0.75n_x\}$
$\mathsf{F}(\mathbf{c})>\mathsf{F}(\mathbf{a})>\mathsf{F}(\mathbf{b})$	$\max\{x_{\{1\}}, n_x\}$

one case impossible, as it would violate the requirement $\mathbf{x}\prec_{\mathcal{R}}\mathbf{y}$ (and/or transitivity). Thus, we are forced to make an unfair decision, e.g. by assuming $\mathbf{z}_i \prec_{\mathsf{F}} \mathbf{z}_{i+1}$ for some i.

On the other hand, assuming that $\mathbf{x}\preceq_{\mathcal{R}}\mathbf{y}$ instead of $\mathbf{x}\prec_{\mathcal{R}}\mathbf{y}$ will inevitably lead to $\mathbf{x}\simeq_{\mathsf{F}}\mathbf{y}$ (as far as fair rule is concerned).

Therefore, the problem with impact functions is that they must be adjusted "globally": to a situation in which we assess all the possible vectors, $\mathbb{I}^{1,2,\cdots}$. This is different to the standard approach in multicriteria decision making, in which we assume that we consider only a distincted set of vectors, $\mathcal{A}\subset\mathbb{I}^{1,2,\cdots}$. The first approach will not necessarily lead to "locally" optimal decisions.

Example 6. Let $\trianglelefteq_{\mathcal{R}}$ be such that $\mathbf{x}\trianglelefteq_{\mathcal{R}}\mathbf{y}$ if and only if $\sum_i x_i \le \sum_i y_i + 1$, and $\vdash_{\mathcal{R}} = \emptyset$.

We have $(3, 2, 1)\preceq_{\mathcal{R}}(4, 3, 1)$. For $\mathcal{A} = \{(3, 2, 1), (4, 3, 1)\}$, we are free to resolve $\preceq_{\mathcal{R}}$ either to $\prec_{\mathsf{F}}$ or $\simeq_{\mathsf{F}}$. However, in $\mathcal{A}' = \{(3, 2, 1), (4, 3, 1), (4, 3)\}$ (as well as in $\mathbb{I}^{1,2,\cdots}$), the only possible decision is $\simeq_{\mathsf{F}}$ for each possible pair. □

4. Can the scientific assessment process be manipulated?

Of course, fair strategy of totalization is not the sole possibility. On the other side of the spectrum, the indefiniteness of $(\trianglelefteq_{\mathcal{R}}, \vdash_{\mathcal{R}})$ may be a tool in the hands of dishonest decision makers.

Again, let us depart from the "weakest" set of comparison results, i.e. $\trianglelefteq$ (thus, we have $(\trianglelefteq_{\mathcal{R}}, \vdash_{\mathcal{R}}) = (\trianglelefteq_{\mathcal{I}}, \emptyset)$). Assume we are given a set of n authors that are incomparable with respect to $\trianglelefteq$. We would like to assess their scientific output. However, for some (not necessarily ethical) reason we have some preferences towards which of them should be ranked higher than the others. Is it possible to construct an impact function that overrates an arbitrarily chosen clique and which, simultaneously, debases some other group of authors?

The answer to this question is positive. What is more, it turns out that there always exists an impact function which exactly generates a predefined ranking, provided that it does not contradict the dominance relation.

Theorem 3. *Let $\mathcal{X} = (\mathbf{x}^1, \ldots, \mathbf{x}^n)$ be a sequence of n vectors such that $\mathbf{x}^1, \ldots, \mathbf{x}^n \in \mathbb{I}^{1,2,\cdots}$, and for all $i \neq j$ it holds $\neg(\mathbf{x}^i \trianglelefteq \mathbf{x}^j)$. Then for any permutation σ of the set $\{1, \ldots, n\}$, there exists an impact function F such that $\mathsf{F}(\mathbf{x}^{\sigma(1)}) < \ldots < \mathsf{F}(\mathbf{x}^{\sigma(n)})$.*

Proof. Let us fix $\mathcal{X}$ and σ. Moreover, let $\mathsf{F}: \mathbb{I}^{1,2,\cdots} \to \mathbb{I}$ be an aggregation operator such that $\mathsf{F}(\mathbf{y}) = \max\{i = 0, 1, \ldots, n : \mathbf{x}^{\sigma(i)} \trianglelefteq \mathbf{y}\}$, where, by convention, $\mathbf{x}^{\sigma(0)} = (0)$. It is easily seen that $\mathsf{F}(0) = 0$, and for all i we have $\mathsf{F}(\mathbf{x}^{\sigma(i)}) = i$, i.e. F generates the desired ordering of $\mathcal{X}$. We now only have to show that F is monotonic with respect to the $\trianglelefteq$ relation. First of all, it is evident that F is symmetric. Next, consider any $\mathbf{y}, \mathbf{z}$ such that $\mathbf{y} \trianglelefteq \mathbf{z}$. We have $\mathsf{F}(\mathbf{y}) = j$ for some $j \in \{0, 1, \ldots, n\}$. Therefore, $\mathbf{x}^{\sigma(j)} \trianglelefteq \mathbf{y}$. By transitivity of $\trianglelefteq$, it also holds $\mathbf{x}^{\sigma(j)} \trianglelefteq \mathbf{z}$. Thus, we have $\mathsf{F}(\mathbf{z}) \ge j$. As a consequence, F is indeed an impact function, and the proof is complete.
□

For the sake of illustration, let us go back to the data set discussed in Section 2.2. Table 1 shows exemplary impact functions that may be used to generate each possible ranking of vectors **a**, **b**, and **c**.

As we see, some of the aggregation operators do not seem to have an "artificial form" and, psychologically speaking, one may have no doubt about their "reasonableness".

4.1. A case study: the generalized h-index and Price medalists

Fig. 3 shows the Hasse diagram for 11 Price medalists that were analyzed in (Gagolewski & Mesiar, 2012).[1] We have highlighted each of the five maximal independent (w.r.t. $\trianglelefteq$) sets $S_1, \ldots, S_5$.

From the results presented in the previous section we know that it is always possible to find an impact function that generates an ordering of each subset of incomparable authors in any preferred way. Once more, one may point out that Theorem 3 is so strong because the conditions provided by $(\trianglelefteq_{\mathcal{I}}, \emptyset)$ are very mild.

We shall thus restrict the space of possible total orderings. Now however, to be less abstract than in the previous section, we will focus on a particular class given by some simple mathematical formula: the so-called OM3 aggregation operators

[1] This data set is available for download from the author's Web page.

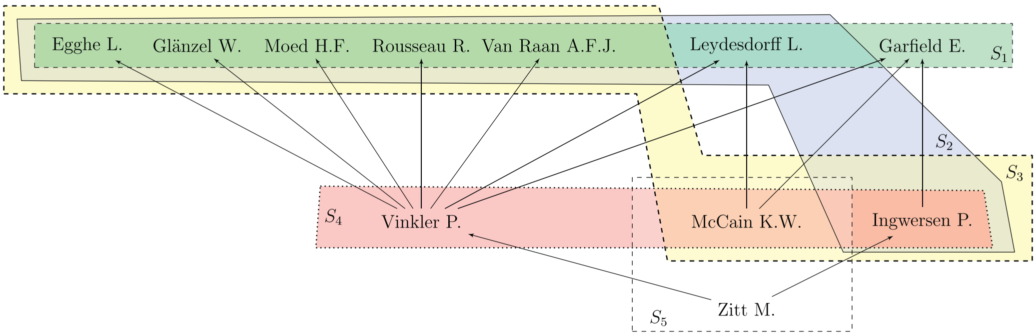

Fig. 3. Hasse diagram for the Price medalists.

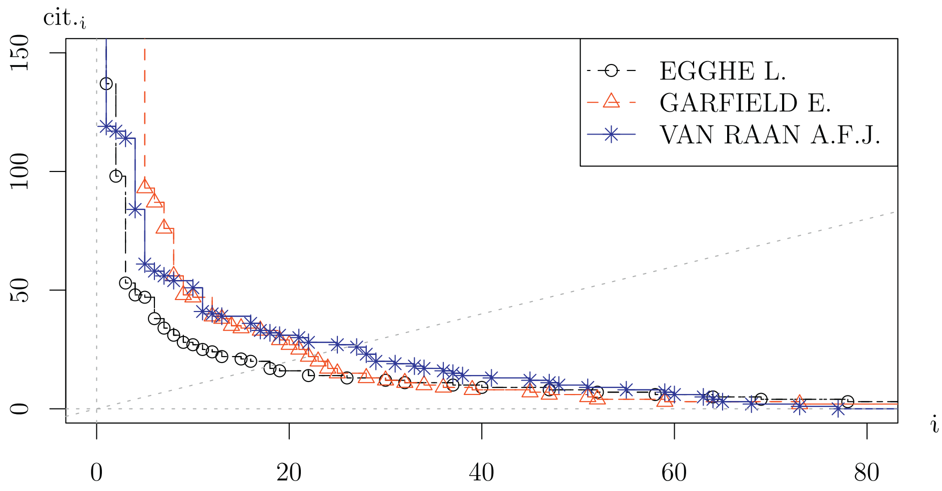

Fig. 4. Citation functions of 3 selected authors.

(Cena & Gagolewski, 2013a, 2013b; Gagolewski, 2013a). Given continuous, increasing functions w, v : $\mathbb{I} \to \mathbb{I}$, and $\mathbf{x} \in \mathbb{I}^{1,2,\ldots}$, let

$$\mathrm{OM3}_{\mathrm{w,v}}(x_1, \ldots, x_n) = \mathrm{v}(\max\{\min\{x_{\{1\}}, \mathrm{w}(1)\}, \ldots, \min\{x_{\{n\}}, \mathrm{w}(n)\}\}).$$

Such impact functions not only have many interesting theoretical properties (like maxitivity, minitivity, modularity, and zero-insensitivity, cf. Gagolewski (2013a), but also a very nice graphical intepretation: these are all and the only aggregation operators that are calculated by transforming the ordinate (y) at which the citation curve intersects some increasing continuous function w. Most importantly, for $\mathrm{w}(x) = x$ and $\mathrm{v}(x) = \lfloor x \rfloor$ we get the h-index.

Fig. 4 depicts the citation curves of 3 randomly chosen, incomparable Price medalists. Note that the OM3-generated ranking is determined by considering the order at which a given function w intersects consecutive citation curves. We have highlighted the case of $\mathrm{w}(x) = x$. Here, $\mathrm{OM3}_{\mathrm{w},\cdot}(\texttt{EGGHE}) < \mathrm{OM3}_{\mathrm{w},\cdot}(\texttt{GARFIELD}) < \mathrm{OM3}_{\mathrm{w},\cdot}(\texttt{VAN RAAN})$. Note that as far as the ranking problem is concerned, the choice of v is irrelevant.

Let us reduce our investigation, for simplicity, to the class of functions $\mathrm{w}(x) = cx$ for some $c > 0$. Of course, a broader class could only increase the set of possible rankings. Table 2 shows which rankings of the authors from Fig. 4 may be obtained. Here $c = 0$ denotes the Max-based ranking and $c = \infty$ the sample length-based one (both are limiting cases of OM3 operators for $\mathbf{x} > 0$).

Additionally, Table 3 gives some excerpts of the authors' citation sequences. Note that if there exists i such that some author's ith citation has the greatest value among all ith citations of group, there exists an OM3 operator for which that

Table 2
All possible OM3-generated rankings for (Egghe, Garfield, Van Raan) and $\mathrm{w}(x) = cx$; "213" means that we may obtain the ordering (Van Raan < Egghe < Garfield)

c	.000	.027	.143	.164	.167	.170	.255	.295	.302	.529	.576	.594	.645	2.44	3.42	10.2	16.8	28.5	∞
EGG	2	3	3	3	3	3	3	3	3	3	3	3	3	3	2	2	1	1	2
GAR	1	1	2	1	2	2	1	2	2	2	1	2	2	3	3	3	3	2	1
VAN	3	2	2	2	2	1	2	2	1	2	2	2	1	1	1	2	2	3	3

Table 3
The number of citations of the *i*th most cited articles of the 3 selected authors.

i	Egghe	Garfield	Van Raan
1	255	**582**	*180*
3	*98*	**386**	117
10	*28*	48	**54**
12	*25*	**47**	41
13	*24*	39	**40**
19	*17*	**33**	32
20	*16*	29	**31**
35	11	*10*	**17**
64	**6**	*3*	5
69	**5**	3	*2*
91	**3**	1	*0*
162	0	0	–
211	–	0	–

author has the most favorable valuation. A dual property holds for the smallest value. The attainability of a single author's promotion or depreciation can then easily be verified only by studying raw citations sequences or examining their plots.

Let us now consider all possible rankings for every combination of 3 Price medalists in S_1 that are incomparable w.r.t. $\trianglelefteq$. The number of independent combinations of triples of authors is $\binom{7}{3} = 35$ (cf. Fig. 3). Table 4 shows which orderings may be generated by using the considered subclass of OM3 impact functions (recall that there always exists an impact function that gives us any desired ranking provided that it does not contradict $\trianglelefteq$). Even though we have concentrated on a fixed set of impact functions, there is still considerable room for manipulation (we could adjust *c* e.g. to normalize citations with respect to scientific domain).

Table 4
All possible OM3-generated rankings for triples in S_1; e.g. for (EGG, GAR, GLA) "132" denotes the ranking (GAR < GLA < EGG) and "322" that (EGG < GAR = GLA).

	123	132	231	213	321	312	333	133	313	331	322	232	223
EGG,GAR,GLA			+	+	+	+				+	+		
EGG,GAR,LEY			+	+	+	+			+	+	+		
EGG,GAR,MOE	+	+	+	+	+	+				+	+		
EGG,GAR,ROU			+	+	+	+			+	+			
EGG,GAR,VAN	+	+	+	+	+	+				+	+	+	
EGG,GLA,LEY	+		+		+	+					+	+	
EGG,GLA,MOE	+	+		+		+					+		
EGG,GLA,ROU	+				+	+		+	+		+		
EGG,GLA,VAN	+	+		+		+			+				
EGG,LEY,MOE				+	+	+							
EGG,LEY,ROU		+	+	+	+	+		+		+	+	+	
EGG,LEY,VAN		+		+	+	+			+		+		
EGG,MOE,ROU	+	+	+		+	+	+				+		
EGG,MOE,VAN	+	+	+		+	+					+	+	
EGG,ROU,VAN	+	+		+	+	+			+		+		
GAR,GLA,LEY	+	+		+	+	+		+	+		+		+
GAR,GLA,MOE	+	+		+		+		+	+				+
GAR,GLA,ROU	+			+	+	+		+					+
GAR,GLA,VAN	+	+		+		+			+				+
GAR,LEY,MOE	+	+		+		+			+				
GAR,LEY,ROU	+	+		+		+		+					
GAR,LEY,VAN	+	+		+		+		+	+				
GAR,MOE,ROU	+	+	+	+	+	+		+			+		+
GAR,MOE,VAN	+	+	+		+	+		+		+	+	+	
GAR,ROU,VAN	+	+	+	+	+	+		+			+	+	
GLA,LEY,MOE	+	+		+		+			+				+
GLA,LEY,ROU	+	+		+		+		+	+				+
GLA,LEY,VAN	+	+		+		+		+					+
GLA,MOE,ROU	+	+	+	+			+	+				+	+
GLA,MOE,VAN	+	+			+			+					+
GLA,ROU,VAN	+	+	+	+				+					+
LEY,MOE,ROU	+	+		+	+	+		+	+	+	+		+
LEY,MOE,VAN	+	+		+	+	+		+	+	+			+
LEY,ROU,VAN	+	+			+	+		+	+				
MOE,ROU,VAN	+	+	+	+	+	+	+	+		+	+	+	+

5. Discussion and conclusions

The implications of the results presented in this paper may seem unfavorable to automated scientific impact assessment. In other words, it is the very nature of the assessment process that it produces disputable results – in some cases there is always doubt why some vectors are ordered in a given way and not the other. The world, as well as decision making processes in particular, is not ideal.

Of course, although arity-monotonicity and symmetry are very interesting and attractive properties, one may sometimes prefer to consider other axiomatizations, i.e. such that the output of a new, very weak paper may *decrease* overall valuation. In such cases we expect to obtain, however, similar conclusions.

In fact, there are better tools to rank scientists than the impact functions: instead of concentrating on the whole set of possible outcomes, one rather should think of using methods that concern only a fixed set of authors. In the time of computers, comparing n authors with each other (even for quite large n) is definitely not a problem. Having a simple mathematical definition is no more a good argument for some impact function. Moreover, this may not give us full control on the way we compare pairs of authors.

What is more, we should ask ourselves whether the set $(\mathbb{I}, \leq)$ (the space to which $(\mathbb{I}^{1,2,\cdots}, \trianglelefteq)$ is mapped) is *really* linearly ordered. "h-index 4 is greater than 3" – this may be just the "magic of numbers", a psychological effect which makes us believe that we are on the way towards some kind of "objectivity" (cf. Gagolewski, 2013b).

We suggest that impact functions shall rather be conceived as ways to describe a citation sequence. For example h-index of 4 means that an author has 4 papers with at least for citations – and nothing more (definitely not that it is "worse" than having 5 papers cited no less than 5 times). Impact indices may also be used to depict current state-of-the-art of science from the bibliometrical perspective.

Perhaps the practical usage of impact indices should be reduced to the provision of a set of requirements needed to gain some "level of excellence". For example, a director may say to the employee: reach g-index of 7 and we will grant you some money bonus. This is the kind of transparency which is also needed, but not very often applied. In such approaches, impact functions could be used to provide *binary* outcomes: yes, he/she reached the level we discussed; *or* no, not yet.

Of course, this does not mean that we ought to resign from using automated methods for decision making in scientific evaluation (as well as related disciplines in which impact functions are applied, see e.g. Franceschini & Maisano, 2009). Nevertheless, these methods still aim at giving a much higher level of transparency. However, the greater the awareness of general limitations of such tools and of their particular advantages, the more rational choice of an appropriate tool can be made.

Nevertheless, it is important to note that any method to be used in the real assessment will modify (in the long term) the human behavior towards the direction of the impact function's increment. Thus, it should be introduced to the concerned groups in advance so that there is enough time to adapt. Moreover, these groups have to be sure that the method will be used for a substantial amount of time and will not be likely to change too often.

Acknowledgments

The author would like to thank the anonymous reviewers whose precious, helpful comments have lead to a significant improvement of the manuscript. Thanks also to Prof. Radko Mesiar, Anna Cena and Bartek Tartanus for stimulating discussion. This study was partially supported by the European Union from resources of the European Social Fund, Project PO KL "Information technologies: Research and their interdisciplinary applications", agreement UDA-POKL.04.01.01-00-051/10-00, and by FNP START 2013 Scholarship from the Foundation for Polish Science.

References

Alonso, S., Cabrerizo, 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.

Bartneck, C., & Kokkelmans, S. (2011). Detecting *h*-index manipulation through self-citation analysis. *Scientometrics*, *87*, 85–98.

Bornmann, L., & Daniel, H.-D. (2009). The state of *h* index research? *EMBO Reports*, *10*(1), 2–5.

Cena, A., & Gagolewski, M. (2013a). OM3: Ordered maxitive, minitive, and modular aggregation operators – Part I: Axiomatic analysis under arity-dependence. In H. Bustince, et al. (Eds.), *Aggregation functions in theory and in practise (AISC 228)* (pp. 93–103). Springer-Verlag.

Cena, A., & Gagolewski, M. (2013b). OM3: Ordered maxitive, minitive, and modular aggregation operators – Part II: A simulation study. In H. Bustince, et al. (Eds.), *Aggregation functions in theory and in practise (AISC 228)* (pp. 105–115). Springer-Verlag.

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

Franceschini, F., & Maisano, D. A. (2009). The Hirsch index in manufacturing and quality engineering. *Quality and Reliability Engineering International*, *25*, 987–995.

Franceschini, F., & Maisano, D. A. (2011). Structured evaluation of the scientific output of academic research groups by recent *h*-based indicators. *Journal of Informetrics*, *5*, 64–74.

Gagolewski, M. (2013a). On the relationship between symmetric maxitive, minitive, and modular aggregation operators. *Information Sciences*, *221*, 170–180.

Gagolewski, M. (2013b). Statistical hypothesis test for the difference between hirsch indices of two pareto-distributed random samples. In R. Kruse, et al. (Eds.), *Synergies of soft computing and statistics for intelligent data analysis (AISC 190)* (pp. 359–367). Springer-Verlag.

Gagolewski, M., & Grzegorzewski, P. (2011). Possibilistic analysis of arity-monotonic aggregation operators and its relation to bibliometric impact assessment of individuals? *International Journal of Approximate Reasoning*, *52*(9), 1312–1324.

Gagolewski, M., & Mesiar, R. (2012). Aggregating different paper quality measures with a generalized *h*-index? *Journal of Informetrics*, *6*(4), 566–579.

Grabisch, M., Marichal, J.-L., Mesiar, R., & Pap, E. (2009). *Aggregation functions*. Cambridge University Press.

Hirsch, J. E. (2005). An index to quantify individual's scientific research output? *Proceedings of the National Academy of Sciences*, *102*(46), 16569–16572.
Marchant, T. (2009a). An axiomatic characterization of the ranking based on the *h*-index and some other bibliometric rankings of authors? *Scientometrics*, *80*(2), 325–342.
Marchant, T. (2009b). Score-based bibliometric rankings of authors? *Journal of the American Society for Information Science and Technology*, *60*(6), 1132–1137.
Nair, G. M., & Turlach, B. A. (2012). The stochastic *h*-index? *Journal of Informetrics*, *6*(1), 80–87.
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. (2011). Further characterizations of the Hirsch index. *Scientometrics*, *88*, 107–114.
Rousseau, R. (2008). Woeginger's axiomatisation of the *h*-index and its relation to the *g*-index, the *h*(2)-index and the r^2-index. *Journal of Informetrics*, *2*(4), 335–340.
Torra, V., & Narukawa, Y. (2008). The *h*-index and the number of citations: Two fuzzy integrals? *IEEE Transactions on Fuzzy Systems*, *16*(3), 795–797.
Woeginger, G. J. (2008a). An axiomatic analysis of Egghe's *g*-index? *Journal of Informetrics*, *2*(4), 364–368.
Woeginger, G. J. (2008b). An axiomatic characterization of the Hirsch-index? *Mathematical Social Sciences*, *56*(2), 224–232.
Woeginger, G. J. (2008c). A symmetry axiom for scientific impact indices. *Journal of Informetrics*, *2*, 298–303.
Zhivotovsky, L. A., & Krutowsky, K. V. (2008). Self-citation can inflate *h*-index? *Scientometrics*, *77*(2), 373–375.