H-index and Other Sugeno Integrals: Some Defects and Their Compensation

Radko Mesiar and Marek Gagolewski

***Abstract*—The famous Hirsch index has been introduced just ca. 10 years ago. Despite that, it is already widely used in many decision making tasks, like in evaluation of individual scientists, research grant allocation, or even production planning. It is known that the h-index is related to the discrete Sugeno integral and the Ky Fan metric introduced in the 1940s. The aim of this paper is to propose a few modifications of this index as well as other fuzzy integrals – also on bounded chains – that lead to better discrimination of some types of data that are to be aggregated. All of the suggested compensation methods try to retain the simplicity of the original measure.**

***Index Terms*—h-index, Sugeno integral, Ky Fan metric, Shilkret integral, decomposition integrals**

I. Introduction

EVEN though the original purpose of Hirsch's proposal in 2005 [1] was to compare the scientific outputs of prominent physicists, the usage of the famous h-index is not limited to the domain of scientometrics: it may be used whenever there is a need to combine quality and quantity of agents represented by non-negative numeric lists into a single number, compare, e.g., [2], [3].

Let $\mathcal{S}$ denote the set of decreasingly ordered non-negative integer sequences of any length, i.e., $\mathcal{S} = \{(x_1, \dots, x_n) : n \in \mathbb{N}, x_1, \dots, x_n \in \mathbb{N}_0,\ x_1 \geq \cdots \geq x_n\}$. In the classical scientometric setting, x_i denotes the number of citations received by the ith most cited paper of a scientist represented by the citation sequence $\mathbf{x}$.

The h-index is a function $\mathsf{H} : \mathcal{S} \to \mathbb{R}$ such that:

$$\mathsf{H}(x_1, \dots, x_n) = \begin{cases} \max\{h = 1, \dots, n : x_h \geq h\} & \text{if } x_1 \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Equivalently, $\mathsf{H}(x_1, \dots, x_n) = \bigvee_{i=1}^{n} x_i \wedge i = \max\{\min\{x_1, 1\}, \min\{x_2, 2\}, \dots, \min\{x_n, n\}\}$. Much is known about this data fusion tool, especially from the point of view of aggregation theory. For example, it may be shown that this aggregation function is minitive, maxitive, and modular, compare [4]. Torra and Narukawa in [5] showed that the h-index is a special case of the discrete Sugeno integral [6] (with respect to the counting measure), one that does not differentiate between papers with no citations and non-existing papers. Such a class of data fusion and decision support methods was already studied more than 40 years ago, especially in the domain of multicriteria decision making. Due to that, this bibliometric index may be conceived as a prominent application of fuzzy (non-additive) measures and integrals, see [7]. Notably, the history of this kind of fusion functions dates back to the 1940s, as it might be shown that the h-index is the Ky Fan metric [8] computed with respect to the $\mathbf{0}$ vector. It is worth emphasizing that the creation of impact indices to measure the performance of scientists, research groups, or whole institutes is not the only successful area of applications of fuzzy measures, integrals, and corresponding aggregation functions. For instance, Beliakov and James studied various ways to rank journals using such tools, see [9], [10], and also [11].

Even though the h-index as a performance measure is being widely criticized, one should face the fact that despite all of that it is often applied in practice. This index is used nowadays (directly or indirectly, as a supporting measure) in the evaluation of scientists in many countries – including young scientists. In the next section we discuss some flaws of the h index and other Sugeno integrals. The main aim of this research note is to present a few possible, practically useful ways to compensate their drawbacks, see Section III. Finally, in Section IV we discuss their possible extensions to different input data types, like elements in a bounded chain.

II. Defects

Let us define the following preordering relation $\trianglelefteq$ on $\mathcal{S}$. We write $(x_1, \dots, x_n) \trianglelefteq (y_1, \dots, y_m)$, whenever $n \leq m$ and $x_1 \leq y_1, \dots, x_n \leq y_n$. It is often assumed (see, e.g., [3] and further references therein) that any reasonable scientometric index K evaluating the performance of scientists should be increasing with respect to this preordering, i.e., K should be a homomorphism between $(\mathcal{S}, \trianglelefteq)$ and $(\mathbb{R}, \leq)$. In other words, $\mathbf{x} \trianglelefteq \mathbf{y}$ implies that $\mathsf{K}(\mathbf{x}) \leq \mathsf{K}(\mathbf{y})$. Of course, we see that this is the case of the h-index. Also note that the restriction of K to a set of sequences of equal lengths is increasing with respect to each variable.

For each n, let us define four lists like:

- $\mathbf{x}^{(n,1)} = (\underbrace{n, \dots, n}_{n \text{ times}})$, i.e., n papers with exactly n citations each,
- $\mathbf{x}^{(n,2)} = (\underbrace{n, \dots, n}_{n \text{ times}}, n, n, \dots)$, i.e., "infinitely many" papers with n citations each,

Radko Mesiar is with the Faculty of Civil Engineering, Department of Mathematics, Slovak University of Technology, 810 05 Bratislava, Slovakia and with the National Supercomputing Center IT4Innovations, Division University of Ostrava, IRAFM, 30.Dubna 22, 701 03 Ostrava 1, Czech Republic; Email: mesiar@math.sk.

Marek Gagolewski is with the Systems Research Institute, Polish Academy of Sciences, ul. Newelska 6, 01-447 Warsaw, Poland and with the Faculty of Mathematics and Information Science, Warsaw University of Technology, ul. Koszykowa 75, 00-662 Warsaw, Poland; Email: gagolews@ibspan.waw.pl.

Manuscript received XX; revised YY.

- $\mathbf{x}^{(n,3)} = (\underbrace{\infty, \dots, \infty}_{n \text{ times}})$, i.e., n papers with "infinitely many" citations,
- $\mathbf{x}^{(n,4)} = (\underbrace{\infty, \dots, \infty}_{n \text{ times}}, n, n, \dots)$, i.e., "infinitely many" papers of which n have "infinitely many" citations and the rest have n citations each.

Note that for all $i \in \{1, 2, 3, 4\}$ it holds that $\mathsf{H}(\mathbf{x}^{(n,i)}) = n$. Moreover, we have the following result, see Fig. 1 for a graphical illustration.

Proposition 1. *Let $\mathbf{y} \in \mathcal{S}$. Then for any n it holds that $\mathsf{H}(\mathbf{y}) = n$ if and only if $\mathbf{x}^{(n,1)} \trianglelefteq \mathbf{y} \trianglelefteq \mathbf{x}^{(n,4)}$.*

Sketch of the proof. By the definition of the h-index, it is easily seen that $\mathbf{x}^{(n,1)}$ is the smallest (with respect to $\trianglelefteq$) vector such that $\mathsf{H}(\mathbf{x}^{(n,1)}) = n$ and that $\mathbf{x}^{(n,4)}$ is the largest one such that $\mathsf{H}(\mathbf{x}^{(n,4)}) = n$. □

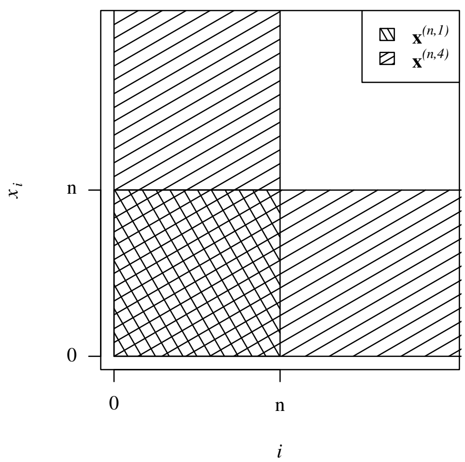

Figure 1. Illustration for Proposition 1.

In other words, the h-index is insensitive to a large number of papers with relatively small number of citations (compare $\mathbf{x}^{(n,1)}$ with $\mathbf{x}^{(n,2)}$ or $\mathbf{x}^{(n,4)}$), but also with respect to a large number of citations received by a small number of papers (compare $\mathbf{x}^{(n,1)}$ with $\mathbf{x}^{(n,3)}$ or $\mathbf{x}^{(n,4)}$). Additionally, please note that the addition of papers with 0 citations never causes the h-index to change its value.

Several approaches towards how to compensate some of the defects of the h-index without neglecting its "spirit" are known in the literature. For example, $h(2)$-index by Kosmulski [12], gives higher output values for scientists with a high number of citations. It is a performance measure $\mathsf{H}_2 : \mathcal{S} \to \mathbb{N}_0$ given by:

$$\mathsf{H}_2(x_1, \dots, x_n) = \begin{cases} \max\{h : x_h \geq h^2\} & \text{if } x_1 \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

In this case we have $\mathsf{H}(\mathbf{x}^{(n,1)}) = \mathsf{H}_2(\mathbf{x}^{(n,2)}) = \lfloor \sqrt{n} \rfloor$ and $\mathsf{H}(\mathbf{x}^{(n,3)}) = \mathsf{H}_2(\mathbf{x}^{(n,4)}) = n$. Note that $\mathsf{H}_2(x_1, \dots, x_n) = \lfloor \sqrt{\bigvee_{i=1}^{n} x_i \wedge i^2} \rfloor$. We see that this index is also a (monotonously transformed) Sugeno integral, this time with respect to a different monotone measure.

It is worth noting that many authors considered other settings than "h^2" on the right side of the equation defining the $h(2)$-index, e.g., "αh", $\alpha > 0$, or "h^β", $\beta \geq 1$, compare [13]. In particular, the family of c-indices by Bras-Amorós, Domingo-Ferrer, and Torra [14], is given by $\mathsf{C}_\alpha(\mathbf{x}) = \max\{\min\{\alpha 1, x_1\}, \dots, \{\alpha n, x_n\}\}$ for some $\alpha > 0$. All of them lead to monotonically transformed Sugeno integrals, perhaps with respect to monotone measures different than the counting one. Nevertheless, such types of modification schemes do not solve the main problem associated with this fuzzy integral: no matter the choice of a monotone measure, vectors like the aforementioned $\mathbf{x}^{(n,1)}, \dots, \mathbf{x}^{(n,4)}$ can always be found. For instance, in the case of the $h(2)$-index, we have:

- $\mathbf{x}'^{(n,1)} = (\underbrace{n^2, \dots, n^2}_{n \text{ times}})$,
- $\mathbf{x}'^{(n,2)} = (\underbrace{n^2, \dots, n^2}_{n \text{ times}}, n^2, n^2, \dots)$,
- $\mathbf{x}'^{(n,3)} = (\underbrace{\infty, \dots, \infty}_{n \text{ times}})$,
- $\mathbf{x}'^{(n,4)} = (\underbrace{\infty, \dots, \infty}_{n \text{ times}}, n^2, n^2, \dots)$.

Therefore, in the section to follow, we propose some new performance measures which are directly based on the h-index and compensate some of its drawbacks. It is worth noting that all these proposals can be also applied to all the other Sugeno integral-based indices.

III. Some compensation methods

Perhaps the simplest method to distinguish researchers with the same h-index is to consider the number of their papers with non-zero citations or the maximal number of citations as a secondary, supplementary measure (which leads to a lexicographic-like ordering on a set of chosen indices). However, in such a case we do not distinguish between $\mathbf{x}^{(n,2)}$ and $\mathbf{x}^{(n,4)}$ or between $\mathbf{x}^{(n,3)}$ and $\mathbf{x}^{(n,4)}$, respectively. In the following subsections we propose some more sophisticated methods either refining the original h-index (case A) or modifying it to compensate some defects of h (cases B, C, D, E).

A. Counting increments

Consider a scientist characterized by a sequence $\mathbf{x} \in \mathcal{S}$ and assume that $\mathsf{H}(\mathbf{x}) = n$. Each added new paper with 0 citations as well as a new citation to a paper of the considered agent shall be called an **increment** from now on. Then we can introduce a new impact function, n-reverse of h-index, denoted H^n, defined as the minimal number of increments applied to $\mathbf{x}$ such the modified list $\mathbf{x}^*$ leads to $\mathsf{H}(\mathbf{x}^*) = n + 1$. Clearly, the lower H^n, the better is the performance of the considered scientist.

Based on Proposition 1, it is evident that $1 \leq \mathsf{H}^n(\mathbf{x}) \leq 2n + 2$. The upper bound is obtained by adding 1 citation to each of the existing n papers in $\mathbf{x}^{(n,1)}$, then adding a new paper, followed by adding $n + 1$ citations to it.

Therefore, we may consider a secondary h-index $\mathsf{H}'(\mathbf{x}) = 2n + 2 - \mathsf{H}^n(\mathbf{x})$ ranging from 0 to $2n + 1$. Observe that

Please cite this paper as: Mesiar R., Gągolewski M., H-index and other Sugeno integrals: Some defects and their compensation, *IEEE Transactions on Fuzzy Systems* **24**(6), 2016, pp. 1668–1672, doi:10.1109/TFUZZ.2016.2516579.

$\mathsf{H}'(\mathbf{x}^{(n,1)}) = 0$, $\mathsf{H}'(\mathbf{x}^{(n,2)}) = n+1$, $\mathsf{H}'(\mathbf{x}^{(n,3)}) = n$, and $\mathsf{H}(\mathbf{x}^{(n,4)}) = 2n+1$. Thus, although the h-index does not distinguish records $\mathbf{x}^{(n,1)}, \dots, \mathbf{x}^{(n,4)}$ from each other, based on H' we see that $\mathbf{x}^{(n,4)}$ has the best performance while $\mathbf{x}^{(n,1)}$ has the worst one, which fully corresponds to our intuition.

Note that H' is also increasing with respect to the ordering $\trianglelefteq$, i.e., if $\mathsf{H}(\mathbf{x}) = \mathsf{H}(\mathbf{y})$ and $\mathbf{x} \trianglelefteq \mathbf{y}$, then $\mathsf{H}'(\mathbf{x}) \le \mathsf{H}'(\mathbf{y})$.

For a given group of scientists, we can thus easily introduce a refined ranking, based primarily on the h-index, and secondarily (if there are ties) on the secondary h-index H'. Intuitively, this index measures how "easy" it is to increase an agent's h-index by 1.

Example 1. Let us study a data set which consists of two citation records for both of which we have $h = 13$: $\mathbf{x} = (55, 39, 39, 31, 30, 27, 25, 22, 20, 18, 18, 17, 15, 12, 12, 11, 8, 8, 7, 7, 3, 3, 2, 2, 1, 1, 0, 0, 0)$ and $\mathbf{y} = (284, 82, 60, 39, 31, 26, 26, 24, 22, 22, 16, 15, 13, 12, 12, 12, 7, 5, 5, 5, 5, 5, 4, 4, 3, 3, 2, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0)$. Our secondary h-index yields $26 = \mathsf{H}'(\mathbf{x}) > \mathsf{H}'(\mathbf{y}) = 25$.

B. Lower 2-h-index

To compensate the lack of consideration of a significant number of papers with small number of citations, we propose the usage of the sum of the standard h-index and the shifted h-index, $\mathsf{H}(\mathbf{x}|_{\mathsf{H}(\mathbf{x})})$, where $\mathbf{x}|_m = (x_{m+1}, x_{m+2}, \dots, x_n)$ for any $0 \le m \le n$. This new impact function will be called the lower 2-h-index, $\mathsf{L2H}(\mathbf{x}) = \mathsf{H}(\mathbf{x}) + \mathsf{H}(\mathbf{x}|_{\mathsf{H}(\mathbf{x})})$.

Proposition 2. *The lower 2-h-index is a function increasing with respect to $\trianglelefteq$.*

For the prototypical citation sequences we have $\mathsf{L2H}(\mathbf{x}^{(n,1)}) = n$, $\mathsf{L2H}(\mathbf{x}^{(n,2)}) = 2n$, $\mathsf{L2H}(\mathbf{x}^{(n,3)}) = n$, and $\mathsf{L2H}(\mathbf{x}^{(n,4)}) = 2n$. Also please note that $\mathsf{H}(\mathbf{x}|_{\mathsf{H}(\mathbf{x})})$ alone may be used as a secondary h-index.

Example 2. Let us go back to the data discussed in Example 1. We have $\mathsf{L2H}(\mathbf{x}) = 13 + \mathsf{H}(12, 12, 11, 8, 8, 7, 7, \dots) = 13 + 7 = 20$ and $\mathsf{L2H}(\mathbf{y}) = 13 + \mathsf{H}(12, 12, 12, 7, 5, \dots) = 13 + 5 = 18$.

C. Upper 2-h-index

To compensate the lack of consideration of a large number of citations for some of the papers, we propose the use the sum of the standard h-index and the truncated h-index, $\mathsf{H}(\mathbf{x}|^{\mathsf{H}(\mathbf{x})})$, where $\mathbf{x}|^m = (\max\{x_1 - m, 0\}, \max\{x_2 - m, 0\}, \dots)$. The new scientometric index from now on shall be called the upper 2-h-index, $\mathsf{U2H}(\mathbf{x}) = \mathsf{H}(\mathbf{x}) + \mathsf{H}(\mathbf{x}|^{\mathsf{H}(\mathbf{x})})$.

Proposition 3. *The upper 2-h-index is monotone with respect to $\trianglelefteq$.*

For instance, we have $\mathsf{U2H}(\mathbf{x}^{(n,1)}) = n$, $\mathsf{U2H}(\mathbf{x}^{(n,2)}) = n$, $\mathsf{U2H}(\mathbf{x}^{(n,3)}) = 2n$, and $\mathsf{U2H}(\mathbf{x}^{(n,4)}) = 2n$.

Example 3. In Example 1, we have $\mathsf{U2H}(\mathbf{x}) = 13 + 8 = 21$ and $\mathsf{U2H}(\mathbf{y}) = 13 + 9 = 22$.

D. 3-h-index

To compensate both defects discussed above we propose the arithmetic mean of the lower and the upper 2-h-indices, calling the new scientific index the 3-h-index, i.e., $\mathsf{3H}(\mathbf{x}) = (\mathsf{L2H}(\mathbf{x}) + \mathsf{U2H}(\mathbf{x}))/2 = \mathsf{H}(\mathbf{x}) + (\mathsf{H}(\mathbf{x}|_{\mathsf{H}(\mathbf{x})}) + \mathsf{H}(\mathbf{x}|_{\mathsf{H}(\mathbf{x})}))/2$.

Proposition 4. *The 3-h-index is also a function increasing with respect to $\trianglelefteq$.*

We have: $\mathsf{3H}(\mathbf{x}^{(n,1)}) = n$, $\mathsf{3H}(\mathbf{x}^{(n,2)}) = 3n/2$, $\mathsf{3H}(\mathbf{x}^{(n,3)}) = 3n/2$, and $\mathsf{3H}(\mathbf{x}^{(n,4)}) = 2n$.

Remark 5. If the above scientometric index was defined as $\mathsf{3H}'(\mathbf{x}) = \mathsf{H}(\mathbf{x}) + \mathsf{H}(\mathbf{x}|_{\mathsf{H}(\mathbf{x})}) + \mathsf{H}(\mathbf{x}|_{\mathsf{H}(\mathbf{x})})$, we would not get a function increasing with respect to $\trianglelefteq$. For example, consider $\mathbf{x} = (3, 2, 1) \trianglelefteq \mathbf{y} = (3, 3, 3)$. We have $\mathsf{3H}'(\mathbf{x}) = 2 + 1 + 1 = 4 > \mathsf{3H}'(\mathbf{y}) = 3 + 0 + 0 = 3$.

Example 4. In Example 1 we have $\mathsf{3H}(\mathbf{x}) = 13 + (7+8)/2 = 20.5$ and $\mathsf{3H}(\mathbf{y}) = 13 + (5+9)/2 = 20$.

Example 5. Let us take $\mathbf{u} = (7, 5, 4, 3, 3, 3, 1, 0)$. Then $\mathsf{H}(\mathbf{u}) = 3$, $\mathsf{L2H}(\mathbf{u}) = 3 + 3 = 6$, $\mathsf{U2H}(\mathbf{u}) = 3 + 2 = 5$, $\mathsf{3H}(\mathbf{u}) = 3 + (2+3)/2 = 5.5$. This situation is illustrated in Figure 2.

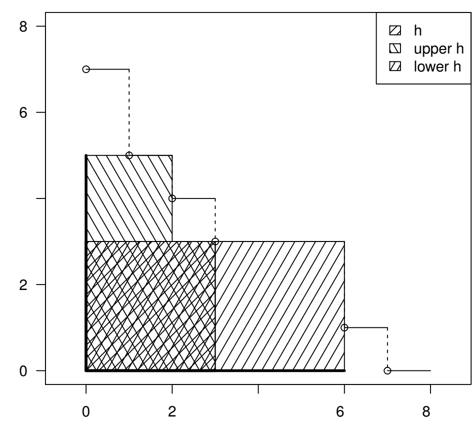

Figure 2. Illustration for Example 5.

Remark 6. Similar compensation schemes can be proposed for other scientometric indices which ignore possible important information in the citation vectors too. This is not only the case of the $h(2)$-index and c-indices, but also, e.g., the w-index [15]:

$$\mathsf{W}(\mathbf{x}) = \max\{w : x_i \ge w - i + 1 \text{ for all } i \le w\},$$

which can be expressed as a Sugeno integral too, but this time computed on appropriately transformed input vectors, see [7] for more details.

Additionally, let us consider the Kosmulski MaxProd index [16],

$$\mathsf{MP}(x_1, \dots, x_n) = \max\{i x_i : i = 1, 2, \dots, n\},$$

which can be interpreted as the area of the greatest rectangle that can be fit under the citation curve (in the case of the Hirsch index this was the greatest square). As noted in [7], the MaxProd index is exactly the Shilkret [17] fuzzy integral with respect to the counting measure. To some extent, this type of performance index also suffers from the aforementioned drawbacks. Before introducing its compensation method in the spirit of the upper and lower h-indices, we shall note that this time the maximum can be attained at the same time for different i. For instance, the modified MaxProd-like indices may be given by:

$$\begin{aligned} \mathsf{L2MP}(\mathbf{x}) &= \max\left\{ix_i + \mathsf{MP}(\mathbf{x}|_i) : i = 1,2,\dots,n\right\}, \\ \mathsf{U2MP}(\mathbf{x}) &= \max\left\{ix_i + \mathsf{MP}(\mathbf{x}|^{x_i}) : i = 1,2,\dots,n\right\}, \\ \mathsf{3MP}(\mathbf{x}) &= \max\left\{ix_i + \mathsf{MP}(\mathbf{x}|_i) + \mathsf{MP}(\mathbf{x}|^{x_i}) : \right. \\ &\qquad \left. i = 1,2,\dots,n\right\}, \end{aligned}$$

For the data set in Example 5 we have $\mathsf{MP}(\mathbf{u}) = 18$ ($i = 6$), $\mathsf{L2MP}(\mathbf{u}) = 22$ (1), $\mathsf{U2MP}(\mathbf{u}) = 22$ (6), and $\mathsf{3MP}(\mathbf{u}) = 24$ (2). On the other hand, in Example 1, we have $\mathsf{MP}(\mathbf{x}) = 204$ ($i = 12$), $\mathsf{L2MP}(\mathbf{x}) = 274$ (7), $\mathsf{U2MP}(\mathbf{x}) = 274$ (16), and $\mathsf{3MP}(\mathbf{x}) = 326$ (12), as well as $\mathsf{MP}(\mathbf{y}) = 284$ ($i = 1$), $\mathsf{L2MP}(\mathbf{y}) = 482$ (1), $\mathsf{U2MP}(\mathbf{y}) = 482$ (10), and $\mathsf{3MP}(\mathbf{y}) = 560$ (3). Please note that these new indices (MaxProd- and Hirsch-based ones) are in fact instances of fuzzy decomposition integrals, see, e.g., [18], [19].

E. Assisted h-indices

To avoid neglecting uncited (or lowly cited) papers, we propose to consider a fixed award $\alpha \in \mathbb{N}$ consisting of the number of added citations to each published paper, and then applying the standard h-index. This new index, denoted with H_α, is given by $\mathsf{H}_\alpha(\mathbf{x}) = \mathsf{H}(x_1 + \alpha, x_2 + \alpha, \dots)$. Note that this approach distinguishes between uncited papers and non-existing papers (for very large α we have that $\mathsf{H}_\alpha(x_1, \dots, x_n) = n$), which is not the case of the original h-index. Thus, it may be of potential interest when assessing young scientists at the beginning of their careers (if bibliometric quantification is really necessary in their case) who have not had time for writing a significant number of contributions yet.

Symmetrically, we may also consider giving an award for having a highly cited paper in one's literary output. With this we approach for some $\beta \in \mathbb{N}$ a new index $\mathsf{H}^\beta(\mathbf{x}) = \mathsf{H}(\underbrace{x_1, \dots, x_1}_{\beta \text{ times}}, x_1, x_2, \dots, x_n)$. Observe that for large β it holds that $\mathsf{H}^\beta(\mathbf{x}) = x_1$.

Proposition 7. H_α *and* H^β *are increasing with respect to* $\trianglelefteq$ *for any* $\alpha, \beta \in \mathbb{N}$.

For example, we have: $\mathsf{H}_\alpha(\mathbf{x}^{(n,1)}) = n$, $\mathsf{H}_\alpha(\mathbf{x}^{(n,2)}) = n+\alpha$, $\mathsf{H}_\alpha(\mathbf{x}^{(n,3)}) = \max\{n, \alpha\}$, and $\mathsf{H}_\alpha(\mathbf{x}^{(n,4)}) = n+\alpha$. Moreover, $\mathsf{H}^\beta(\mathbf{x}^{(n,1)}) = n$, $\mathsf{H}^\beta(\mathbf{x}^{(n,2)}) = n$, $\mathsf{H}^\beta(\mathbf{x}^{(n,3)}) = n + \beta$, and $\mathsf{H}^\beta(\mathbf{x}^{(n,4)}) = n + \beta$. Note that H_α and H^β may be combined so that a scientometric index H_α^β is obtained.

Example 6. Let us go back to the two citation sequences in Example 1. Tables I and II give the values of $\mathsf{H} = \mathsf{H}_0^0$, $\mathsf{H}_\alpha = \mathsf{H}_\alpha^0$, $\mathsf{H}^\beta = \mathsf{H}_0^\beta$, and H_α^β for the two sequences of concern. In most of the cases, the agent represented by the numeric list $\mathbf{x}$ is indicated at least as influential as the one represented by $\mathbf{y}$.

Nevertheless, please notice that with a "creative" selection of the α and β parameters one may – to some degree – manipulate the obtained scientometric rankings. In fact, this is the drawback of all the scientometric impact measures, see also [20], [21]. We already noted that even the original Hirsch's proposal has many generalizations (compare the notion of, e.g., the c-index [14]; in the Sugeno integral-framework this corresponds to a choice of the discrete fuzzy measure). A relevant parameter selection depends on many factors, including the nature of and citation patterns in the given research field. It should be done well in advance, before actually assessing a group of researchers, to avoid any accusations of being biased. First of all, one may rely on experts' knowledge and/or intuition here. Alternatively, various methods for learning (fitting) scientometric indices from empirical data may be utilized, compare, e.g., [9], provided that one has access to a reference training data set (input vectors together with desired outputs).

Apart from adding artificial citations and papers, one may also try to construct an h-index-based performance measure via a redistribution of citations in a few first highly-cited papers to the other ones. For instance, if we assume that one (other choices are possible too) top-cited paper's citations may be spread over several entities so as to guarantee that the resulting h-index is as high as possible, this new measure computed on a tuple $(72, 9, 5, 3)$ yields a result identical to $\mathsf{H}(9, 9, 9, 9, 9, 9, 9, 9, 9, 5, 3)$, that is, equal to 9.

IV. Discussion

Described in this paper were some modifications of the h-index which, contrary to many other proposals in the literature, are of the same "spirit" as the original measure. All of them have a clear, intuitive interpretation and are very easy to compute.

Surely, the introduced h-index compensation methods are not only limited to the discrete Sugeno integral with respect to the counting measure. As we noted, in a similar manner one may easily modify Sugeno integrals based on different monotone measures. What is crucial indeed is the fact that the indicated flaws are observed when applying not solely the "simplest" of the Sugeno integrals, but all of them.

One should be aware of the fact that the number of citations is not the only way to measure the performance of an agent. In the scientometric context itself there are many other factors that can be taken into account in $\mathbf{x} \in \mathcal{S}$, too. For instance, this is the case of the influence of time. An immediate approach is to replace $(x_1, \dots, x_n)$ with the average per-year citation counts $(p_1, \dots, p_n)$. What is important, the introduced measures (as well as the original h-index, in this context called the time-normalized h) are still applicable in such a setting.

Please that the scope of this paper not only is on the field of bibliometrics, but to all the application areas where the Sugeno integral with respect to symmetric discrete measures are used. We proposed some general compensation methods for this fuzzy integral that may be useful, e.g., when computing the false discovery rate (FDR) in the multiple significance testing problem, compare [22], too.

It is worth noting that the first (increments' counting) and the last (assisted h) groups of compensation methods presented in this paper can be reformulated in such a way that they are valid if we aggregate elements on an arbitrary lower-bounded discrete chain. First of all, this is because of the fact that the Sugeno integral is in fact defined solely using join and meet operations and hence are particular weighted lattice polynomial functions. Then, if we have a linearly ordered set $A = (a_1, a_2, \dots)$ with $a_i \prec a_{i+1}$ for any i, one may define, e.g., $a_i + \alpha$ to be equal to $a_{i+\alpha}$ etc.

Acknowledgments

Radko Mesiar would like to acknowledge the support by the grant APVV-14-0013 and by the European Regional Development Fund in the IT4 Innovations Center of Excellence Project (CZ.1.05/1.1.00/02.0070). Marek Gagolewski's study was supported by the National Science Center, Poland, research project 2014/13/D/HS4/01700. The authors would like to thank the anonymous referees for interesting and useful remarks which significantly improved the quality of the manuscript.

Table I
VALUES OF $\mathsf{H}_\alpha^\beta(\mathbf{x})$ FOR DIFFERENT α, β; INPUT DATA ARE FROM EXAMPLE 1. EMPHASIZED ARE CASES IN WHICH $\mathsf{H}_\alpha^\beta(\mathbf{x}) > \mathsf{H}_\alpha^\beta(\mathbf{y})$.

α \ β	0	1	2	3	4	5	6	7	8	9	10
0	13	**14**	**15**	15	**16**	**17**	**17**	**18**	18	18	19
1	13	14	**15**	**16**	16	**17**	**18**	**18**	**19**	19	19
2	14	14	15	**16**	**17**	17	**18**	**19**	**19**	**20**	20
3	15	15	15	16	**17**	**18**	18	**19**	**20**	**20**	**21**
4	15	16	16	16	17	**18**	**19**	19	**20**	**21**	**21**
5	16	16	17	17	17	18	**19**	**20**	20	**21**	**22**
6	16	17	17	18	18	18	19	**20**	**21**	21	**22**
7	16	17	18	18	19	19	19	20	**21**	**22**	22
8	16	17	18	19	19	20	20	20	21	**22**	**23**
9	**17**	17	18	19	20	20	21	21	21	22	**23**
10	**18**	**18**	18	19	20	21	21	22	22	22	23

Table II
VALUES OF $\mathsf{H}_\alpha^\beta(\mathbf{y})$ FOR DIFFERENT α, β; INPUT DATA ARE FROM EXAMPLE 1. EMPHASIZED ARE CASES IN WHICH $\mathsf{H}_\alpha^\beta(\mathbf{y}) > \mathsf{H}_\alpha^\beta(\mathbf{x})$.

α \ β	0	1	2	3	4	5	6	7	8	9	10
0	13	13	14	15	15	16	16	17	18	**19**	**20**
1	13	14	14	15	16	16	17	17	18	19	**20**
2	14	14	15	15	16	17	17	18	18	19	20
3	15	15	15	16	16	17	18	18	19	19	20
4	**16**	16	16	16	17	17	18	19	19	20	20
5	16	**17**	17	17	17	18	18	19	20	20	21
6	16	17	**18**	18	18	18	19	19	20	21	21
7	16	17	18	**19**	19	19	19	20	20	21	22
8	16	17	18	19	**20**	20	20	20	21	21	22
9	16	17	18	19	20	**21**	21	21	21	22	22
10	17	17	18	19	20	21	**22**	22	22	22	23

References

[1] J. E. Hirsch, "An index to quantify individual's scientific research output," **Proceedings of the National Academy of Sciences**, vol. 102, no. 46, pp. 16 569–16 572, 2005.

[2] F. Franceschini and D. A. Maisano, "The Hirsch index in manufacturing and quality engineering," **Quality and Reliability Engineering International**, vol. 25, pp. 987–995, 2009.

[3] A. Cena and M. Gagolewski, "OM3: Ordered maxitive, minitive, and modular aggregation operators – Axiomatic and probabilistic properties in an arity-monotonic setting," **Fuzzy Sets and Systems**, vol. 264, pp. 138–159, 2015.

[4] M. Gagolewski, "On the relationship between symmetric maxitive, minitive, and modular aggregation operators," **Information Sciences**, vol. 221, pp. 170–180, 2013.

[5] V. Torra and Y. Narukawa, "The h-index and the number of citations: Two fuzzy integrals," **IEEE Transactions on Fuzzy Systems**, vol. 16, no. 3, pp. 795–797, 2008.

[6] M. Sugeno, "Theory of fuzzy integrals and its applications," Ph.D. dissertation, Tokyo Institute of Technology, 1974.

[7] M. Gagolewski and R. Mesiar, "Monotone measures and universal integrals in a uniform framework for the scientific impact assessment problem," **Information Sciences**, vol. 263, pp. 166–174, 2014.

[8] K. Fan, "Entfernung zweier zufälligen Größen und die Konvergenz nach Wahrscheinlichkeit," **Mathematische Zeitschrift**, vol. 49, pp. 681–683, 1943.

[9] G. Beliakov and S. James, "Citation-based journal ranks: The use of fuzzy measures," **Fuzzy Sets and Systems**, vol. 167, pp. 101–119, 2011.

[10] ——, "Using linear programming for weights identification of generalized Bonferroni means in R," **Lecture Notes in Computer Science**, vol. 7647, pp. 35–44, 2012.

[11] G. Beliakov, H. Bustince, and T. Calvo, **A Practical Guide to Averaging Functions**. Springer, 2016.

[12] M. Kosmulski, "A new Hirsch-type index saves time and works equally well as the original h-index," **ISSI Newsletter**, vol. 2, no. 3, pp. 4–6, 2006.

[13] S. Alonso, F. J. Cabrerizo, E. Herrera-Viedma, and F. Herrera, "h-index: A review focused on its variants, computation and standardization for different scientific fields," **Journal of Informetrics**, vol. 3, pp. 273–289, 2009.

[14] M. Bras-Amorós, J. Domingo-Ferrer, and V. Torra, "A bibliometric index based on the collaboration distance between cited and citing authors," **Journal of Informetrics**, vol. 5, no. 2, pp. 248–264, 2011.

[15] G. J. Woeginger, "An axiomatic characterization of the Hirsch-index," **Mathematical Social Sciences**, vol. 56, no. 2, pp. 224–232, 2008.

[16] M. Kosmulski, "MAXPROD — A new index for assessment of the scientific output of an individual, and a comparison with the h-index," **Cybermetrics**, vol. 11, no. 1, p. 5, 2007.

[17] N. Shilkret, "Maxitive measure and integration," **Indagationes Mathematicæ**, vol. 33, pp. 109–116, 1971.

[18] R. Mesiar and A. Stupňanová, "Decomposition integrals," **International Journal of Approximate Reasoning**, vol. 54, no. 8, pp. 1252–1259, 2013.

[19] Y. Even and E. Lehrer, "Decomposition-integral: Unifying Choquet and the concave integrals," **Economic Theory**, vol. 56, no. 1, pp. 33–58, 2014.

[20] M. Gagolewski, "Scientific impact assessment cannot be fair," **Journal of Informetrics**, vol. 7, no. 4, pp. 792–802, 2013.

[21] M. Gagolewski and R. Mesiar, "Aggregating different paper quality measures with a generalized h-index," **Journal of Informetrics**, vol. 6, no. 4, pp. 566–579, 2012.

[22] Y. Benjamini and Y. Hochberg, "Controlling False Discovery Rate: A practical and powerful approach to multiple testing," **Journal of the Royal Statistical Society. Series B**, vol. 57, no. 1, pp. 289–300, 1995.