Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Information Processing and Management 44 (2008) 770–780
INFORMATION PROCESSING & MANAGEMENT
www.elsevier.com/locate/infoproman

An *h*-index weighted by citation impact

Leo Egghe [a,b], Ronald Rousseau [a,b,c,*]

[a] *Hasselt University, Agoralaan, B-3590 Diepenbeek, Belgium*
[b] *Antwerp University, IBW, Universiteitsplein 1, B-2610 Wilrijk, Belgium*
[c] *KHBO (Association K.U.Leuven), Industrial Sciences and Technology, Zeedijk 101, B-8400 Oostende, Belgium*

Received 27 March 2007; received in revised form 10 May 2007; accepted 11 May 2007
Available online 26 June 2007

Abstract

An *h*-type index is proposed which depends on the obtained citations of articles belonging to the *h*-core. This weighted *h*-index, denoted as h_w, is presented in a continuous setting and in a discrete one. It is shown that in a continuous setting the new index enjoys many good properties. In the discrete setting some small deviations from the ideal may occur.

Keywords: *h*-Index; *h*-Type indices; Weighted *h*-index; Discrete and continuous approach; Power law model

1. Introduction

The *h*-index, also known as the Hirsch index, was introduced by Hirsch (2005) as an indicator for lifetime achievement. Considering a scientist's list of publications, ranked according to the number of citations received, the *h*-index is defined as the highest rank such that the first *h* publications received each at least *h* citations. Although this idea can be applied to many source–item relations we will mainly use the terminology of publications and citations as in Hirsch' original article.

For advantages and disadvantages of the *h*-index we refer to Hirsch (2005), Glänzel (2006) and Jin et al. (2007). In order to overcome some of these disadvantages scientists have proposed several 'Hirsch-type' indices with the intention of either replacing or complementing the original *h*-index. Among these we mention Egghe's *g*-index (2006a, 2006b), Kosmulski's $H^{(2)}$-index Jin's *A* and *AR*-indices (Jin, 2006, 2007) and the *R*-index (Jin, Liang, Rousseau, & Egghe, 2007).

We recall the definitions of the most interesting among these proposals. For the *g*-index as well as for the $H^{(2)}$-index one draws the same list as for the *h*-index. The *g*-index, on the one hand, is defined as the highest rank such that the cumulative sum of the number of citations received is larger than or equal to the square of this rank. Clearly $h \leqslant g$. The $H^{(2)}$-index, on the other hand, is *k* if *k* is the highest rank such that the first *k*

* Corresponding author. Address: KHBO (Association K.U.Leuven), Industrial Sciences and Technology, Zeedijk 101, B-8400 Oostende, Belgium.
E-mail addresses: leo.egghe@uhasselt.be (L. Egghe), ronald.rousseau@khbo.be (R. Rousseau).

0306-4573/$ - see front matter
doi:10.1016/j.ipm.2007.05.003

publications received each at least k^2 citations. The $H^{(2)}$-index will not be discussed further, as we do not think this is an interesting proposal. Jin's A-index achieves the same goal as the g-index, namely correcting for the fact that the original h-index does not take the exact number of citations of articles included in the h-core into account. This index is simply defined as the average number of citations received by the publications included in the Hirsch core. Mathematically, this is:

$$A = \frac{1}{h}\sum_{j=1}^{h} y_j \tag{1}$$

In formula (1) the numbers of citations (y_j) are ranked in decreasing order. Clearly $h \leqslant A$. The R-index, a correction on the A-index (Jin et al., 2007) is defined as:

$$R = \sqrt{\sum_{j=1}^{h} y_j} \tag{2}$$

The h-, A- and R-indices are related through the relation $R = \sqrt{A \cdot h}$. The AR-index, a refinement of the R-index which takes the age of the publications (denoted as a_j) into account, has been proposed by Jin (2007):

$$AR = \sqrt{\sum_{j=1}^{h} \frac{y_j}{a_j}} \tag{3}$$

The formulae shown in Eqs. (1)–(3) are those as defined for the discrete, practical case. They can also be defined in a general, continuous model. In this approach $\gamma(r)$ denotes the continuous rank-frequency function:

$$\gamma : [0, T] \to [1, +\infty] : r \to \gamma(r) \tag{4}$$

The function $\gamma(r)$ is by definition a decreasing, but not necessarily strictly decreasing, positive function, with $\gamma(T) = 1$ and $T > 1$. Assuming, as we do from now on, that $\int_0^T \gamma(s)\,\mathrm{d}s < +\infty$ the four previously mentioned Hirsch-type indices are defined in the continuous case as follows:

$$h \text{ is the unique solution of } r = \gamma(r) \tag{5}$$

$$g \text{ is the unique solution of } r^2 = \int_0^r \gamma(s)\,\mathrm{d}s \tag{6}$$

(assuming that $\int_0^T \gamma(s)\,\mathrm{d}s \leqslant T^2$); the g-index can also be characterized as the largest rank r such that

$$r^2 \leqslant \int_0^r \gamma(s)\,\mathrm{d}s$$

$$A = \frac{1}{h}\int_0^h \gamma(r)\,\mathrm{d}r \tag{7}$$

$$\text{and } R = \sqrt{\int_0^h \gamma(r)\,\mathrm{d}r} \tag{8}$$

In this article we focus on the fact that the h-index lacks sensitivity to performance changes and propose a citation-weighted h-index. As the construction of this new index is more elegant in the continuous case than in the discrete one, we first introduce it for the continuous model.

2. Construction of a citation-weighted *h*-index: continuous case

Theorem 1. *Let h be the h-index of* $\gamma(r)$ *then the equation* $\frac{\int_0^t \gamma(r)\,\mathrm{d}r}{h} = \gamma(t)$ *has always a unique solution. This unique solution, denoted as* r_0, *is called the w-rank of the given rank-frequency distribution.*

Proof. Since $\gamma(r)$ is a rank-frequency function it is strictly positive. Define now,

$$k(t) = \gamma(t) - \frac{\int_0^t \gamma(r)\,\mathrm{d}r}{h}$$

The function $k(t)$ is continuous and strictly decreasing on $[0,T]$. Indeed: $k'(t) = \gamma'(t) - \frac{\gamma(t)}{h} < 0$. We further see that $k(0) > 0$ and $k(T) = \gamma(T) - \frac{\int_0^T \gamma(r)\,\mathrm{d}r}{h} \leqslant 1 - \frac{T}{h} < 0$, as $h < T$. By the intermediate value theorem the function k takes all values on the interval $[k(T), k(0)]$. Hence there exists at least one value $r_0 \in [0,T]$ such that $k(r_0) = 0$. Consequently, $\frac{\int_0^{r_0} \gamma(r)\,\mathrm{d}r}{h} = \gamma(r_0)$. Uniqueness follows from the facts that $\gamma(t)$ is non-increasing and that the function $\int_0^t \gamma(r)\,\mathrm{d}r$ is strictly increasing.

Clearly, $r_0 \leqslant h$, as $\frac{\int_0^t \gamma(r)\,\mathrm{d}r}{h}$ is increasing in t and takes a value which is at least equal to h in h, while $\gamma(t)$ is non-increasing and $\gamma(h) = h$ by definition. □

Definition (*the continuous citation-weighted* h-index). Let h be the h-index of $\gamma(r)$ and let r_0 be the unique solution of the equation

$$\frac{\int_0^t \gamma(r)\,\mathrm{d}r}{h} = \gamma(t) \tag{9}$$

Then the weighted h-index, denoted as h_w is defined as:

$$h_\mathrm{w} = \sqrt{\int_0^{r_0} \gamma(r)\,\mathrm{d}r} \tag{10}$$

$$\text{or, equivalently: } h_\mathrm{w} = \sqrt{h \cdot \gamma(r_0)} \tag{11}$$

3. Properties of the weighted h-index h_w

Theorem 2. *If the h-index of the continuous rank-frequency function $\gamma(r)$ is h and if the restriction of $\gamma(r)$ to $[0,h]$ is a constant function then $h_w = h$ and $\gamma(r)|_{[0,h]} = h$.*

Proof. We know already that $r_0 \leqslant h$. Let now $\gamma(r) = C$ on $[0,h]$. Then the defining equation $\frac{\int_0^{r_0} \gamma(r)\,\mathrm{d}r}{h} = \gamma(r_0)$ becomes: $\frac{\int_0^{r_0} C\,\mathrm{d}r}{h} = C$, with $r_0 = h$ as its unique solution. Then $h_\mathrm{w} = \sqrt{\int_0^h C\,\mathrm{d}r} = \sqrt{C \cdot h}$. By definition we know that $h = \gamma(h) = C$, hence $h_\mathrm{w} = \sqrt{C \cdot h} = \sqrt{h \cdot h} = h$.

When the number of citations is the same for each r, articles at each (continuous) rank are weighted equally, hence it is natural that $h_\mathrm{w} = h$. This simple relation is not true if the definition of the h_w-index does not contain a square root. Hence, this theorem explains why we introduced the extra mathematical operation of taking a square root. □

Theorem 3. *If the restriction of the continuous rank-frequency function $\gamma(r)$ to $[0,h]$ is not the constant function $\gamma(r) = h$, then the w-rank is strictly smaller than the h-index: $r_0 < h$ and $h \leqslant h_w$.*

Proof. Assume that $r_0 = h$. Then $\frac{\int_0^{r_0} \gamma(r)\,\mathrm{d}r}{h} = \frac{\int_0^{h} \gamma(r)\,\mathrm{d}r}{h} > \gamma(h) = \gamma(r_0)$, as γ is not the constant function $\gamma(r) = h$ on $[0,h]$. The inequality $\frac{\int_0^{r_0} \gamma(r)\,\mathrm{d}r}{h} > \gamma(r_0)$ contradicts Eq. (9), hence $r_0 < h$. From Eq. (9) we see that $h_\mathrm{w}^2 = h \cdot \gamma(r_0) \geqslant h \cdot \gamma(h) = h^2$. Hence: $h \leqslant h_\mathrm{w}$. □

Corollary 1. *If $\gamma(r)$ is strictly decreasing on $[0,h]$ then $h < h_w$.*

Corollary 2. *The w-rank of a continuous rank-frequency function $\gamma(r)$ with h-index equal to h is equal to h if and only if the restriction of $\gamma(r)$ to $[0,h]$ is the constant function h.*

We already know that if the restriction of $\gamma(r)$ to $[0, h]$ is the constant function h then $r_0 = h$ and $h = h_{\mathrm{w}}$. If now $r_0 = h$ then, by definition, $\int_0^{r_0} \gamma(r)\,\mathrm{d}r = \int_0^h \gamma(r)\,\mathrm{d}r = h \cdot \gamma(h) = h^2$. The second equality sign follows from the definition of r_0 $(=h)$. Hence the restriction of $\gamma(r)$ to $[0, h]$ is the constant function h and $h_{\mathrm{w}} = h$.

Corollary 3. *If the restriction of the continuous rank-frequency function $\gamma(r)$ to $[0, h]$ is not the constant function $\gamma(r) = h$, then the weighted h-index is always strictly smaller than the R-index*: $h_w < R$.

Proof. As $h_{\mathrm{w}} = \sqrt{\int_0^{r_0} \gamma(r)\,\mathrm{d}r}$ and $r_0 < h$ it follows that $h_{\mathrm{w}} < \sqrt{\int_0^h \gamma(r)\,\mathrm{d}r} = R$. $\square$

Corollary 4. *If the restriction of the continuous rank-frequency function $\gamma(r)$ to $[0, h]$ is not the constant function $\gamma(r) = h$, then the g-index is always strictly larger than the weighted h-index*: $g > h_w$.

Proof. We already know that $r_0 < h \leqslant h_{\mathrm{w}}$. Hence we also have: $\int_0^{h_{\mathrm{w}}} \gamma(r)\,\mathrm{d}r > \int_0^{r_0} \gamma(r)\,\mathrm{d}r = h_{\mathrm{w}}^2$. Now g is the largest value such that $r^2 \leqslant \int_0^r \gamma(s)\,\mathrm{d}s$. Hence $g > h_{\mathrm{w}}$. $\square$

4. Requirements for an *h*-type index in the continuous model

If $\gamma_1(r)$ and $\gamma_2(r)$ are two continuous rank-frequency functions defined on $[0, T]$ and if $\forall r \in [0, T]: \gamma_1(r) \leqslant \gamma_2(r)$ then an acceptable h-type measure M must satisfy the relation $M(\gamma_1(r)) \leqslant M(\gamma_2(r))$.

If there exists $\delta > 0$ such that the relations $\forall r \in]0, \delta]: \gamma_1(r) < \gamma_2(r)$ and $\forall r \in [0, T]: \gamma_1(r) \leqslant \gamma_2(r)$ imply that $M(\gamma_1(r)) < M(\gamma_2(r))$, then M is said to be a strong h-type measure.

Theorem 4. *The h-index itself is an acceptable h-type measure, but it is not a strong h-type measure.*

Proof. The h-index is the unique solution of the equation $r = \gamma(r)$. As the function $f(r) = r$ is increasing in r and the function $\gamma(r)$ is decreasing in r, it follows that the larger the values of $\gamma(r)$ the larger the h-index. This proves that the Hirsch index is an acceptable h-type index.

We next give an example to show that the h-index is not a strong h-type measure. Let $\gamma_1(r)$ and $\gamma_2(r)$ be continuous, decreasing functions, such that $\forall r \in [0, T]: \gamma_1(r) \leqslant \gamma_2(r)$. Moreover $\gamma_1(r)$ is equal to $20 - 3r$ on $[0, 5]$, and $\gamma_2(r)$ is equal to $15 - 2r$ on $[0, 5]$. Then $h(\gamma_1(r))$, the h-index of $\gamma_1(r)$, is equal to 5, and equal to $h(\gamma_2(r))$. This shows that the h-index is not a strong h-type measure. $\square$

Theorem 5. *The weighted h-index*, h_w, *is an acceptable h-type measure.*

Proof. Let $\gamma_1(r)$ and $\gamma_2(r)$ be two continuous rank-frequency functions defined on $[0, T]$, such that $\forall r \in [0, T]: \gamma_1(r) \leqslant \gamma_2(r)$. Then we know already that h_1, the h-index of $\gamma_1(r)$, is smaller than or equal to h_2, the h-index of $\gamma_2(r)$. Let r_j denote the w-rank of $\gamma_j(r)$, $j = 1, 2$.

If $r_1 \leqslant r_2$ then we have for $h_{2,\mathrm{w}}$, the weighted h-index of $\gamma_2(r)$:

$$\begin{aligned} h_{2,\mathrm{w}} &= \sqrt{\int_0^{r_2} \gamma_2(r)\,\mathrm{d}r} \\ &\geqslant \sqrt{\int_0^{r_1} \gamma_2(r)\,\mathrm{d}r} \geqslant (*)\sqrt{\int_0^{r_1} \gamma_1(r)\,\mathrm{d}r} = h_{1,\mathrm{w}} \end{aligned} \tag{10}$$

If $r_1 > r_2$ then:

$$\begin{aligned} h_{2,\mathrm{w}} &= \sqrt{h_2 \cdot \gamma_2(r_2)} \\ &\geqslant \sqrt{h_1 \cdot \gamma_2(r_2)} \geqslant \sqrt{h_1 \cdot \gamma_1(r_2)} \geqslant \sqrt{h_1 \cdot \gamma_1(r_1)} = h_{1,\mathrm{w}} \end{aligned} \tag{11}$$

This proves that the h_{w}-index is always an acceptable h-type measure.

We investigate now whether the h_w-index is a strong h-type measure. If there exists $\delta > 0$ such that $\forall r \in]0, \delta] : \gamma_1(r) < \gamma_2(r)$ then the inequality indicated with (*) in the proof above becomes a strict inequality. This takes care of the case $r_1 \leqslant r_2$.

If $r_1 > r_2$ and $h_2 > h_1$ then

$$\begin{aligned} h_{2,w} &= \sqrt{h_2 \cdot \gamma_2(r_2)} \\ &\geqslant \sqrt{h_2 \cdot \gamma_2(r_1)} \geqslant \sqrt{h_2 \cdot \gamma_1(r_1)} > \sqrt{h_1 \cdot \gamma_1(r_1)} = h_{1,w} \end{aligned} \tag{11}$$

If $h_1 = h_2$ $(=h)$ and if one of the inequalities below, indicated with (**), is strict, i.e. $\gamma_1(r_1) < \gamma_1(r_2)$, or $\gamma_2(r_1) < \gamma_2(r_2)$, or $\gamma_1(r_1) < \gamma_2(r_1)$, or $\gamma_1(r_2) < \gamma_2(r_2)$ then $h_{2,w} > h_{1,w}$:

$$h_{2,w} = \sqrt{h \cdot \gamma_2(r_2)} \geqslant (**) \sqrt{h \cdot \gamma_2(r_1)} \geqslant (**) \sqrt{h \cdot \gamma_1(r_1)} = h_{1,w}$$
$$h_{2,w} = \sqrt{h \cdot \gamma_2(r_2)} \geqslant (**) \sqrt{h \cdot \gamma_1(r_2)} \geqslant (**) \sqrt{h \cdot \gamma_1(r_1)} = h_{1,w}$$

This shows that it is only possible that $h_{2,w} = h_{1,w}$ under the condition that there exists $\delta > 0$ such that $\forall r \in]0, \delta] : \gamma_1(r) < \gamma_2(r)$, if $h_1 = h_2$; $\gamma_1(r)$ and $\gamma_2(r)$ are constant on an interval situated in $]0, h]$; and moreover equal on this interval. An example of such a case is the following situation:

$$\gamma_1(r) = \begin{cases} -2x + 16 & r \in [0, 4] \\ 8 & r \in]4, 10] \\ < 8 & r > 10 \end{cases}$$
$$\gamma_2(r) = \begin{cases} -4x + 24 & r \in [0, 4] \\ 8 & r > 4 \end{cases}$$

Indeed: $\gamma_1(r) < \gamma_2(r)$ on $[0, 4[$; $h_1 = h_2 = 8$; $\gamma_1(r) = \gamma_2(r)$ on $[4, 8]$, $r_1 = 6$, $r_2 = 4$ and $h_{2,w} = h_{1,w} = 8$. These functions also illustrate the result obtained in Theorem 3: if the rank-frequency function is not constant then it is possible that $h = h_w$ but then the w-rank must be strictly smaller than h. □

We conclude this investigation by stating the following result.

Theorem 6. *The weighted h-index, h_w, is a strong h-type measure in the set of strictly decreasing rank-frequency functions.*

Note that this result is not true for the h-index. We repeat that even in the set of all decreasing rank-frequency functions the required strict inequality $h_{2,w} > h_{1,w}$ is often – but not always! – true.

5. The h_w-index in the power law model

We recall the following defining relations for the rank-frequency and size-frequency functions in the power law (Lotkaian) model (Egghe, 2005).

The size-frequency relation is given by the function $\Phi(x)$:

$$\Phi : [1, +\infty] \to [0, C] : x \to \Phi(x) = \frac{C}{x^\alpha} \tag{12}$$

In Eq. (12) C is a strictly positive constant, and $\alpha > 2$. The restriction to the case $\alpha > 2$ is necessary for the convergence of the integrals used further on. The corresponding, this means: mathematically equivalent, rank-frequency relation is given by the function Γ:

$$\Gamma : [0, T] \to [1, +\infty] : r \to \Gamma(r) = \frac{B}{r^\beta} \tag{13}$$

with $B > 0$ and $0 < \beta < 1$. In order to define Γ in 0, as in the previous sections, we put $\Gamma(0) = +\infty$. We further note that this function Γ is strictly decreasing, hence all properties studied for the general continuous rank-frequency function are valid with strict inequalities. The parameters α and β are related through the equation:

$$\beta = \frac{1}{\alpha - 1} \tag{14}$$

The other variables satisfy the following equalities:

$$B = \left(\frac{C}{\alpha - 1}\right)^{\frac{1}{\alpha-1}} \tag{15}$$

and

$$T = \frac{C}{\alpha - 1} \tag{16}$$

Theorem 7. *If* $\alpha > 2$ *then*

$$h_{\mathrm{w}} = \left(\frac{\alpha - 1}{\alpha - 2}\right)^{\frac{1}{2(\alpha-1)}} T^{\frac{1}{\alpha}} = \left(\frac{\alpha - 1}{\alpha - 2}\right)^{\frac{1}{2(\alpha-1)}} h \tag{17}$$

Proof. In this model $\int_0^{r_0} \Gamma(r)\,\mathrm{d}r = \int_0^{r_0} \frac{B}{r^\beta}\,\mathrm{d}r = \frac{B}{1-\beta} r_0^{1-\beta}$ (18). By definition (9) and Eq. (14) we find: $\frac{1}{h} \cdot \frac{B \cdot r_0^{1-\beta}}{1-\beta} = \Gamma(r_0) = \frac{B}{r_0^\beta}$, and hence: $r_0 = (1-\beta)h$. Then, by (11): $h_{\mathrm{w}} = \sqrt{h \cdot \frac{B}{r_0^\beta}} = \sqrt{h \cdot \frac{B}{(1-\beta)^\beta h^\beta}}$. As $1 - \beta = \frac{\alpha-2}{\alpha-1}$ and $h = T^{1/\alpha}$ (Egghe & Rousseau, 2006), we obtain:

$$h_{\mathrm{w}} = \sqrt{T^{\frac{\alpha-2}{\alpha(\alpha-1)}} \cdot T^{\frac{1}{\alpha-1}} \cdot \left(\frac{\alpha - 1}{\alpha - 2}\right)^{\frac{1}{\alpha-1}}} = \left(\frac{\alpha - 1}{\alpha - 2}\right)^{\frac{1}{2(\alpha-1)}} \cdot T^{\frac{1}{\alpha}} = \left(\frac{\alpha - 1}{\alpha - 2}\right)^{\frac{1}{2(\alpha-1)}} \cdot h$$

This proves Theorem 7. □

Theorem 7 confirms the results obtained for the general continuous model. Indeed: as $\frac{\alpha-1}{\alpha-2} > 1$, for $\alpha > 2$, it follows that $h_{\mathrm{w}} > h$. Moreover, as $g = \left(\frac{\alpha-1}{\alpha-2}\right)^{\frac{\alpha-1}{\alpha}} \cdot h$ (Egghe, 2006c), the relations $\frac{\alpha-1}{\alpha} > \frac{1}{2(\alpha-1)}$ and $\frac{\alpha-1}{\alpha-2} > 1$ for $\alpha > 2$, imply that $g > h_{\mathrm{w}}$.

In the subsequent sections we will discuss the weighted *h*-index in the discrete case.

6. Construction of a citation-weighted *h*-index: discrete case

Table 1 visualizes the construction of the *h*-index. We will refer to such a table as an *h*-table. The left-hand column indicates the rank of a scientist's articles according to the number of received citations. The right-hand column shows the actual number of citations received by each of these articles; *h* denotes the *h*-index.

Such discrete data points $(j, y_j)_j$ can be connected leading to a polygonal line (a set of connected line segments). This polygonal line can be considered the graph of a continuous, decreasing function. In most practical cases this graph is not strictly decreasing. Anyway, the continuous theory explained above may be applied to this function, leading to a weighted h_{w}-index that has all the properties mentioned above. Yet, this h_{w}-index, similar to the real-valued *h*-index (Rousseau, in press) is not easy to calculate. One might prefer an approach which is more suited to the discrete case. We present such an approach, but unfortunately, we have to pay a

Table 1
Method for calculating the *h*-index

Rank (r)	Number of citations
1	$y_1 \geqslant h$
2	$y_2 \geqslant h$
3	$y_3 \geqslant h$
...	...
$h-1$	$y_{h-1} \geqslant h$
h	$y_h \geqslant h$
$h+1$	$y_{h+1} \leqslant h$
...	...

Table 2
Table for the determination of the discrete citation-weighted h-index (h_w)

Weighted rank (r_w)	Number of citations
$r_w(1) = \frac{y_1}{h} \geqslant 1$	$y_1 \geqslant h$
$r_w(2) = \frac{y_1}{h} + \frac{y_2}{h} \geqslant 2$	$y_2 \geqslant h$
...	...
$r_w(j) = \frac{\sum_{i=1}^{j} y_i}{h} \geqslant j$	$y_j \geqslant h$
...	...
$r_w(h) = \frac{\sum_{i=1}^{h} y_i}{h} \geqslant h$	$y_h \geqslant h$
$r_w(h+1) = \frac{\sum_{i=1}^{h+1} y_i}{h}$	$y_{h+1} \leqslant h$
...	...

price for this. Indeed, some nice properties are not always true anymore in this discrete approach. Yet, these exceptions are just marginal effects, probably never, or extremely rarely happening in real applications.

If y_j denotes the number of citations received by article j, then $y_j = \left(\frac{y_j}{h}\right)h$. Hence, if $h \neq 0$ we may say that y_j's h-weight is $\frac{y_j}{h}$. This weight is now used to change the ranks of Table 1, leading to Table 2.

A table such as Table 2 will be referred to as an h_w-table.

Definition (*the discrete citation-weighted* h-index: h_w). This new index is defined as:

$$h_w = \sqrt{\sum_{i=1}^{r_0} y_i} \tag{18}$$

where r_0 is the largest row index j such that

$$r_w(j) \leqslant y_j \tag{19}$$

If h is zero then also h_w is set equal to zero. From Eq. (19) it follows that $y_{r_0} > 0$.

Note that $r_w(r_0)$ is not necessarily a natural number. By its definition it is a positive rational number. Of course, r_0 is a natural number. The citations y_j, may or may not be natural numbers, depending, for instance, on the decision of counting citations as whole numbers or as fractions (in case of multi-authorship). The first r_0 rows of the h_w-table are referred to as the essential part of the h_w-table, as the other rows play no role in the determination of the h_w-index. Of course, we do need the first h (or $h+1$) rows of the h-table otherwise we could not determine h, which we need in the calculation of the h_w-index. The h_w-index is built upon the h-index.

An example

Consider the h-table (Table 3) and its corresponding h_w-table (Table 4).

As the h-index is 4 the corresponding h_w-table is given by Table 4.

Table 4 shows that $r_w(3) = 6.25$ is the largest weighted rank satisfying inequality (19). Hence, $r_0 = 3$ and this scientist's h_w-index is equal to $\sqrt{10+8+7} = 5$. At first, this construction may seem unnecessary compli-

Table 3
An h-table for a scientist with h-index 4

Rank (r)	Number of citations
1	10
2	8
3	7
4	4
5	3
...	...

Table 4
h_w-table corresponding to Table 3

Weighted rank (r_w)	Number of citations
$r_w(1) = \frac{10}{4} = 2.5$	10
$r_w(2) = \frac{18}{4} = 4.5$	8
$r_w(3) = \frac{25}{4} = 6.25$	7
$r_w(4) = \frac{29}{4} = 7.25$	4
$r_w(5) = \frac{32}{4} = 8$	3
...	...

cated, but it will become clear that this is not really the case. Advantages and properties of this particular construction are shown in the following sections.

7. Properties of the h_w-index

It will be shown that, with a slight exception, the h_w-index is not smaller than the h-index. Equality occurs when all articles receive the same number of citations equal to h (this follows from Theorem 8).

Theorem 8. *If the number of citations of the first h articles is the same, say equal to $k \geqslant h$, then $h_w = \sqrt{hk}$.*

Proof. With these assumptions we have the h_w-table (Table 5).

We see from Table 5 that $r_w(h) = k$ and $r_w(h+1) > k$ (unless $y_{h+1} = 0$). As $k \geqslant h$ this shows that $r_w(h+1) > y_{h+1}$, hence $r_0 = h$. If $y_{h+1} = 0$ then $r_w(h+1) = k > 0 = y_{h+1}$. So also in this case $r_0 = h$. By Eq. (18) we find that $h_w = \sqrt{\sum_{j=1}^{h} k} = \sqrt{hk}$. This proves Theorem 8. □

Theorem 8 gives an example where the number of rows in the essential h_w-table is equal to h. Note further that if $k = h$, $h_w = h$. In this case the h-table coincides completely with the corresponding h_w-table. This theorem also shows that there is no theoretical upper limit to the h_w-index.

Theorem 9. *The rank r_0 is never larger than h:*

$$r_0 \leqslant h \tag{20}$$

Proof. Assume that $r_0 > h$. Then we know by the definition of h that $h \geqslant y_{r_0}$. Then:

$$r_w(r_0) = \frac{\sum_{j=1}^{r_0} y_j}{h} > \frac{\sum_{j=1}^{h} y_j}{h} \geqslant h \geqslant y_{r_0}. \tag{21}$$

where the first strict inequality is due to the fact that $y_{r_0} > 0$. By Eq. (19) inequality (21) is impossible. Hence the assumption that $r_0 > h$ leads to a contradiction. This proves Theorem 9.

Table 5
h_w-table corresponding the case where the citations of the first h articles are the same

Weighted rank (r_w)	Number of citations
$r_w(1) = \frac{k}{h}$	$k \geqslant h$
$r_w(2) = \frac{2k}{h}$	$k \geqslant h$
...	...
$r_w(h) = \frac{\sum_{i=1}^{h} k}{h} = k$	$k \geqslant h$
$r_w(h+1) = \frac{\sum_{i=1}^{h+1} y_i}{h} = k + \frac{y_{h+1}}{h}$	$y_{h+1} \leqslant h$
...	...

Theorem 9 proves that the essential part of the h_w-table associated with a given h-table is never longer than the h-table. Theorem 8 shows that the tables can, however, coincide. By its construction the h_w-index takes the number of citations into account, which is the main reason for its introduction. Yet its calculation does not need more rows than the h-index, hence the precision problem (Jin et al., 2007), i.e., the fact that it is rather difficult to collect all data necessary for the determination of the h-index, is exactly the same for the determination of the h_w-index as for the h-index. Contrary to the construction of the g-index the precision problem does not worsen.

Consequently, it follows from Theorem 9 that

$$h_w \leqslant R \tag{22}$$

Indeed, as $r_0 \leqslant h$ it follows that $h_w = \sqrt{\sum_{i=1}^{r_0} y_i} \leqslant \sqrt{\sum_{i=1}^{h} y_i} = R$. If all y_j, $(j = 1, \ldots, h)$ are equal to h, then $h_w = R = h$.

Next we will prove an exact lower bound for the discrete h_w-index.

Theorem 10

$$h_w > \sqrt{h(h-1)} \tag{23}$$

Proof. We consider two cases.

If $r_0 = h$, then $y_{r_0} \geqslant r_w(r_0) = r_w(h) = \frac{\sum_{j=1}^{h} y_j}{h} \geqslant \frac{h^2}{h} = h$ (by Eq. (19)). By Eq. (18) we see that $h_w = \sqrt{\sum_{j=1}^{h} y_j} \geqslant h$. Clearly (23) is then also true.

If $r_0 \neq h$ then $r_0 < h$ (by Eq. (20)). Then $r_0 + 1 \leqslant h$ and hence $y_{r_0+1} \geqslant y_h \geqslant h$ (the y_j are ranked in decreasing order). By the definition of r_0 we know that

$y_{r_0} \geqslant \frac{\sum_{j=1}^{r_0} y_j}{h}$ and $y_{r_0+1} < \frac{\sum_{j=1}^{r_0+1} y_j}{h}$. This inequality leads to: $y_{r_0+1} < \frac{\sum_{j=1}^{r_0+1} y_j}{h} = \frac{\sum_{j=1}^{r_0} y_j + y_{r_0+1}}{h}$.

Hence: $\frac{\sum_{j=1}^{r_0} y_j}{h} > y_{r_0+1}\left(1 - \frac{1}{h}\right) \geqslant h\left(\frac{h-1}{h}\right) = h - 1$. By the definition of the h_w-index it follows that $h_w > \sqrt{h(h-1)}$.

This proves the theorem. □

Corollary. *The ceiling function of h_w, i.e. the smallest integer larger than or equal to h_w, is at least equal to h:*

$$ceil(h_w) \geqslant h \tag{24}$$

Proof. As $(h-1)^2 < (h-1)h < h^2$, we see that $ceil\left(\sqrt{h(h-1)}\right) = h$. Then Eq. (24) follows immediately from Eq. (23). □

We would have preferred that, also in the discrete case, $h_w \geqslant h$ (and not the ceiling function of h_w). This inequality is, however, not true in general. Yet, it is always true in the first case considered in the theorem ($r_0 = h$). The fact that $h_w > \sqrt{h(h-1)}$ and consequently *ceil*$(h_w) \geqslant h$, and not $h_w \geqslant h$ should be considered as a discrete aberration.

We next give an example where $h_w <$ h, but, of course $h_w > \sqrt{h(h-1)}$ and *ceil*$(h_w) = h$. Consider the following h-table (Table 6).

The h-index of Table 6 is 3, which happens also to be its g-index. The corresponding h_w-table is shown in Table 7.

Table 6
Example: h-table

Rank (r)	Number of citations
1	4
2	3
3	3
4	1
...	...

Table 7
h_w-table corresponding to Table 6

Weighted rank (r_w)	Number of citations
$r_w(1) = \frac{4}{3} = 1.333$	4
$r_w(2) = \frac{7}{3} = 2.333$	3
$r_w(3) = \frac{10}{3} = 3.333$	3
...	...

From Table 7 we see that $r_0 = 2$, and $h_w = \sqrt{4+3} \approx 2.645 < 3$. Yet $\sqrt{7} \approx 2.645 > \sqrt{3.2} (= \sqrt{h \cdot (h-1)}) = \sqrt{6} \approx 2.449$. We also have that $ceil(2.645) = 3 = h$.

The requirement $h_w \geqslant h$ is a logical requirement as we want to take the number of citations into account (at least to some extent). As the first h articles are weighted by the factor $\frac{y_j}{h} \geqslant 1$ it is natural to expect that $h_w \geqslant h$. In the continuous model this is always the case (Theorem 3). In the discrete case exceptions are possible, as shown above, but these are really exceptional cases which should be considered as discrete aberrations. We are convinced that the definition of the discrete h_w is structurally correct.

The following result shows the relation between the g-index and the discrete h_w-index. Recall that the floor function of a real number r is the largest integer smaller than or equal to r.

Theorem 11. *The g-index always satisfies the inequality*: $g \geqslant floor(h_w)$.

Proof. If $h_w \geqslant h$ then $floor(h_w) \geqslant h$. Then $\sum_{j=1}^{floor(h_w)} y_j \geqslant \sum_{j=1}^{h} y_j \geqslant h_w^2 \geqslant (floor(h_w))^2$ (using Eq. (22)). As g is the largest natural number such that $\sum_{j=1}^{g} y_j \geqslant g^2$, we see that $g \geqslant floor(h_w)$. If $h_w < h$, then clearly $floor(h_w) < g$ as $h \leqslant g$. In this case it is even true that $ceil(h_w) \leqslant g$. □

8. A general requirement for discrete *h*-type indices

In this section we propose a general requirement for discrete h-type indices. This requirement is the discrete analogue of the first one presented in the continuous case. Under the name discrete h-type indices we mean an index that uses an h-table for its calculation. Such a general measure will be denoted as HIR.

Requirement. If two rankings $Y = (y_j)_j$ and $Z = (z_j)_j$ are given and for each $j = 1, 2, \ldots, y_j \leqslant z_j$ then $\mathrm{HIR}(Y) \leqslant \mathrm{HIR}(Z)$.

It is clear that the h-index, the g-index, the H^2-index and the R-index satisfy this requirement. The A-index on the other hand does not. Indeed: if $Y = (10, 1)$ and $Z = (10, 2)$ then $h(Y) = 1$ and $A(Y) = 10$; $h(Z) = 2$ and $A(Z) = 6$.

Unfortunately, the discrete h_w-index does not satisfy this requirement either. Indeed: let $Y = (8, 4, 3)$ and let $Z = (9, 4, 3)$. Then $h_Y = h_Z = 3$; $r_0(Y) = 2$, while $r_0(Z) = 1$, hence $h_w(Y) = \sqrt{8+4} \approx 3.46$, while $h_w(Z) = \sqrt{9} \approx 3.00$. This shows that, strictly speaking, the discrete h_w is not an acceptable h-type measure. Yet, we consider this as a discrete aberration, due to the fact that the discrete process is only an approximation of the better, continuous process.

Table 8
h-Type indices for five small European countries (2000 publications)

	h	h_w	R	A	AR	max
Iceland	36	49.41	58.08	93.69	22.42	412
Estonia	32	39.34	46.57	67.78	17.98	270
Lithuania	31	37.03	42.52	58.32	16.41	197
Latvia	23	29.02	34.79	52.61	13.43	197
Luxembourg	20	23.45	27.18	36.95	10.49	98
Var.	0.23	0.28	0.28	0.34	0.28	0.50

Table 9
Data related to the calculation of the h_w-indices

	h_w	r_0	$r_w(r_0)$	y_{r_0}
Iceland	49.41	17	67.81	68
Estonia	39.34	17	48.38	49
Lithuania	37.03	19	44.23	45
Latvia	29.02	9	36.61	37
Luxembourg	23.45	12	27.50	28

9. Examples

An as illustration of the calculation of the h_w-index we calculated the h, h_w, R and A index for the publications in 2000 of five small European countries: Iceland, Estonia, Lithuania, Latvia and Luxembourg. Citation data were collected on March 15, 2007 from the Web of Science. Table 8 shows the resulting values for different h-type indices. The last column gives the number of citations of the most cited article. The last line of Table 7 shows the coefficient of variation of the five values. This is an indication of the spread of corresponding index values. Table 9 gives some specific values for the calculations of the h_w-values. As all these articles have the same age, namely 6.71 year (on average), the value of their AR-index is equal to the R-index divided by $\sqrt{6.71} \approx 2.59$.

Clearly, the discrete aberrations that, in theory might have occurred do not happen in these practical cases.

10. Conclusion

A citation-weighted h-index, denoted as h_w, is introduced. It is shown that in a continuous setting this new index satisfies the relations: $h \leqslant h_w < g$; $h_w < R$, if the rank-frequency function is not constant on $[0, h]$. The continuous weighted h-index is an acceptable h-type measure, and even strongly acceptable on the set of strictly decreasing rank-frequency functions. The exact relation between h and h_w is determined in the power law model. Generally, the discrete version enjoys similar good properties, but because of its discrete nature some exceptions may exist.

Acknowledgement

The work presented by Ronald Rousseau in this paper was supported by the National Natural Science Foundation of China by grant no. 70673019.

References

Egghe, L. (2005). *Power laws in the information production process: Lotkaian informetrics*. Oxford(UK): Elsevier.
Egghe, L. (2006a). How to improve the *h*-index. *The Scientist, 20*(3), 14.
Egghe, L. (2006b). An improvement of the *H*-index: The *G*-index. *ISSI Newsletter, 2*(1), 8–9.
Egghe, L. (2006c). Theory and practice of the *g*-index. *Scientometrics, 69*(1), 131–152.
Egghe, L., & Rousseau, R. (2006). An informetric model for the Hirsch index. *Scientometrics, 69*(1), 121–129.
Glänzel, W. (2006). On the opportunities and limitations of the *H*-index. *Science Focus, 1*(1), 10–11 (in Chinese). English version available at E-LIS: code 9535.
Hirsch, J. E. (2005) An index to quantify an individual's scientific research output. In *Proceedings of the National Academy of Sciences of the USA, 102*(46), 16569–16572.
Jin, BH. (2006). H-index: an evaluation indicator proposed by scientist. *Science Focus, 1*(1), 8–9 (in Chinese).
Jin, BH. (2007). The *AR*-index: complementing the *h*-index. *ISSI Newsletter, 3*(1), p. 6.
Jin, BH., Liang, L., Rousseau, R., & Egghe, L. (2007). The *R*- and *AR*-indices: complementing the *h*-index. *Chinese Science Bulletin, 52*(6), 855–863.
Kosmulski, M. (2006). A new Hirsch-type index saves time and works equally well as the original *h*-index. *ISSI Newsletter, 2*(3), 4–6.
Rousseau, R. (in press). Simple models and the corresponding *h*- and *g*-index. In *Research yearbook on science studies and management of science and technology in China*, issue 2006/2007.