Mathematical theory of the h- and g-index in case of fractional counting of authorship

Peer-reviewed author version

EGGHE, Leo (2008) Mathematical theory of the h- and g-index in case of fractional counting of authorship. In: JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY, 59(10). p. 1608-1616.

DOI: 10.1002/asi.20845
Handle: http://hdl.handle.net/1942/7916

Mathematical theory of the h- and g-index in case of fractional counting of authorship

by

L. Egghe

Universiteit Hasselt (UHasselt), Campus Diepenbeek, Agoralaan, B-3590 Diepenbeek, Belgium[1]
and
Universiteit Antwerpen (UA), Campus Drie Eiken, Universiteitsplein 1, B-2610 Wilrijk, Belgium
leo.egghe@uhasselt.be

ABSTRACT

This paper studies the h-index (Hirsch index) and the g-index of authors, in case one counts authorship of the cited articles in a fractional way. There are two ways to do this: or one counts the citations to these papers in a fractional way or one counts the ranks of the papers in a fractional way as credit for an author.

In both cases we define the fractional h- and g-indexes and we present inequalities (both upper and lower bounds) between these fractional h- and g-indexes and their corresponding unweighted values (also involving, of course, the co-authorship distribution). Wherever applicable, examples and counterexamples are provided.

---

[1] Permanent address
Key words and phrases: h-index, g-index, fractional counting
Acknowledgement: The author is grateful to Prof. Dr. R. Rousseau for suggesting to study fractional h- and g-indexes.

In a concrete example (the publication-citation list of the present author) we make explicite calculations of these fractional h- and g-indexes and we show that they are not very different from the unweighted ones.

I. Introduction

Only one and a half years ago (August 2005), J.E. Hirsch proposed his famous h-index (Hirsch (2005)): a scientist has h as h-index if h of his/her papers have at least h citations each and the other papers have no more than h citations each (version in Hirsch (2005)) and in the arXiv repository of 29 September 2005 (arXiv:physics/0508025 v5)[2]. A simpler, but equivalent, formulation of the definition of the h-index is as follows: rank the papers of a scientist in decreasing order of the number of citations they have received. Then this scientist has h-index h if $r = h$ is the highest rank such that the first h papers have at least h citations.

Since its definition (in physics) the h-index has attracted lots of attention: the growth of the number of papers on the h-index and related indexes is spectacular and it is nowadays virtually impossible to present a complete reference list: we limit ourselves to some early reactions on the h-index: Ball (2005), Bornmann and Daniel (2005), Braun, Glänzel and Schubert (2005), Egghe and Rousseau (2006), Glänzel (2006a,b), Popov (2005) and van Raan (2006).

It is seldom seen that a new indicator has such an impact on the scientific minds: the h-index is already accepted – at this early stage – as an indicator, presented in the Web of Science, whenever a ranking is retrieved according to received citations. The h-index is also applied to journals – see Braun, Glänzel and Schubert (2005) or to "topics" (replacing "scientist") – see Banks (2006). The advantages of the h-index are described in the above publications: it is a simple single number incorporating publication as well as citation data (hence comprising quantitative as well as qualitative or visibility aspects). Another advantage of the h-index is that it is insensitive to a set of uncited (or lowly cited) papers (which every scientist has due to

[2] This paper corrected the original formulation (arXiv:physics/0508025 v3) in the sense that "no more" replaced the word "fewer" (in arXiv, 17 August 2005) which was an erroneous formulation (h could not exist in this early definition).

several reasons, e.g. the publication of requested local articles for which even a reputed scientist cannot be blamed for).

In Egghe (2006a – see also 2006b,c) a disadvantage of the h-index is revealed: the h-index is also insensitive to one or several outstandingly highly cited papers. Indeed, although highly cited papers are important for the determination of the value h of the h-index, once such a highly cited paper is selected to belong to the top h papers, its actual number of citations (at any time) is not used anymore. Indeed, once a paper is selected to the top group, the h-index calculated in subsequent years is not influenced by this paper's received citations further on, even if this paper doubles or triples its number of citations.

In Egghe (2006a,b,c) we introduced the g-index in order to overcome the above mentioned disadvantage of the h-index, whilst keeping its advantages. Notice that, by the very definition of the h-index, the papers on rank $1,...,h$ each have at least h citations, hence these papers have, together, at least h^2 citations. But h is not necessarily the largest rank with this property. Therefore, the g-index is defined as the largest rank such that the first g papers have, together, at least g^2 citations. It is obvious that $g \geq h$ in all cases. By giving practical examples (on the h- and g-index of the Price medallists) we show the advantage of the g-index above the h-index and we also present (in Egghe and Rousseau (2006) and Egghe (2006a)) formulae for the h- and g-index in case we have a Lotkaian paper-citation information production process – see Egghe (2005) on Lotkaian IPPs.

In informetrics, lots of attention has been given to co-authored papers, i.e. papers written by several authors. The natural question to be posed is: how are the credits of each author counted, e.g. does every author in a 3-authored paper gets a credit of 1 (total counting) or does every author gets a credit of $\frac{1}{3}$ (fractional counting) – see Egghe (2005), Egghe, Rousseau and Van Hooydonk (2000). In general, fractional counting is preferred since this does not increase the total weight of a single paper.

The problem raised in this paper (mentioned to me by R. Rousseau) is how to calculate h- and g-indexes for authors when we use a fractional crediting system. Rousseau suggests to give an author of an m-authored paper only a credit $\frac{c}{m}$ if the paper received c citations. This will be studied in this paper and will be called “fractional counting on citations”. It is clear that also another fractional crediting system is possible: “fractional counting on papers”: for each author in an m-authored paper, the paper only occupies a fractional rank of $\frac{1}{m}$. Also this fractional crediting system will be studied in this paper in the connection of the h- and g-index.

In the next section we define exactly both fractional crediting systems, give the definitions of the fractional h- and g-indexes in each case and we present concrete examples.

In the third section we present inequalities (both upper and lower bounds) between the fractional h- index and the unweighted (“classical”) h-index, also incorporating of course the co-authorship distribution, i.e. the function $\varphi(i) = \#$ authors in the paper on rank i. The same is done for the g-index and these inequalities are proved in both fractional crediting systems. Examples and (if applicable) counterexamples are presented. We also show that the proved inequalities are optimal (i.e. that they cannot be improved).

In the fourth section we calculate the non-fractional h- and g-index for the present author and compare these values with the fractional h- and g-index (again in both fractional crediting systems).

The paper closes, in the fifth section, by making concluding remarks, including the formulation of some open problems and topics for further research.

II. The h- and g-index for fractional authorship counts

II.1 Introduction

The calculation of the “classical” h- and g-index of an author (or a topic,…), denoted h and g respectively is based on a ranked list as in Table 1: the papers of the author are ranked in decreasing order of the number of citations they have received.

Table 1. General form of a ranked list of papers in decreasing order of the number of received citations

r (rank paper)	# (number of citations)
1	y_1
2	y_2
3	y_3
$\vdots$	$\vdots$
T	y_T

Here T denotes the total number of papers (under consideration). Note that $(y_1, y_2, ..., y_T)$ is decreasing.

Then the h-index is defined as the largest rank $r = h$ such that $y_h \geq h$. This definition is visualized as in Table 2 (which we will also use further on): $r = h$ is the largest rank such that $y_h \geq h$ since $y_{h+1} \leq h < h + 1$.

Table 2. Visualization of the calculation of the h-index

r	#
1	$y_1 \geq h$
2	$y_2 \geq h$
$\vdots$	$\vdots$
h-1	$y_{h-1} \geq h$
h	$y_h \geq h$
h+1	$y_{h+1} \leq h$
$\vdots$	$\vdots$
T	$y_T \leq h$

The g-index is defined as the largest rank $r = g$ such that $\sum_{i=1}^{g} y_i \geq g^2$. It is obvious that $g \geq h$ in all cases. Here we assume that $g \leq T$: in case $\sum_{i=1}^{T} y_i \geq T^2$ we can define $g = T$ or better (see Egghe (2006a)) we can add, in Table 2, fictitious articles with zero citations: we add enough of these "articles" so that $g \leq T$ (and even $g < T$) where we denote by T the new number of articles (including the fictitious ones).

So, with this possible extension, the calculation of the g-index is visualized as in Table 3:

$r = g$ is the largest rank such that $\sum_{i=1}^{g} y_i \geq g^2$ since $\sum_{i=1}^{g+1} y_i < (g+1)^2$.

Table 3. Visualization of the calculation of the g-index

r	r^2	#	$\sum$ #
1	1	y_1	$y_1 \geq 1$
2	4	y_2	$y_1 + y_2 \geq 4$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$g-1$	$(g-1)^2$	y_{g-1}	$\sum_{i=1}^{g-1} y_i \geq (g-1)^2$
g	g^2	y_g	$\sum_{i=1}^{g} y_i \geq g^2$
$g+1$	$(g+1)^2$	y_{g+1}	$\sum_{i=1}^{g+1} y_i < (g+1)^2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
T	T^2	y_T	$\sum_{i=1}^{T} y_i < T^2$

Let us illustrate the calculation of the h- and g-index on a very simple example (Table 4).

Table 4. Illustration of the calculation of h and g

r	r^2	#	$\sum$ #
1	1	10	10
2	4	5	15
3	9	3	18
4	16	2	$20 > 4^2$
5	25	1	$21 < 5^2$
6	36	1	22

We immediately see that $h = 3$ and that $g = 4$.

Suppose now that we also have the information on the number of co-authors of each paper. For paper $i = 1,...,T$ we will, generally, denote by $\varphi(i)$ the number of co-authors of paper i. How can we use these numbers in order to calculate fractional h- and g-indexes, i.e. h- and g-indexes in which we consider an author of paper i (hence where there are $\varphi(i)$ co-authors in total) to be credited with a value $\frac{1}{\varphi(i)}$ for each i ? For more on fractional (and other) crediting systems we refer the reader to Egghe, Rousseau and Van Hooydonk (2000) or to Egghe (2005).

There are two different ways to do this: fractional count of the citations or fractional count of the papers: this will be done in the next two subsections.

II.2 Fractional h- and g-indexes using fractional citation counts

By fractional citation counts we consider the following variant of Table 1: see Table 5: i.e. all citation counts y_i are divided by $\varphi(i)$, the number of authors of paper i.

Table 5. Fractional citation counts

r	# (fraction of citations)
1	$y_1 / \varphi(1)$
2	$y_2 / \varphi(2)$
3	$y_3 / \varphi(3)$
M	M
T	$y_T / \varphi(T)$

Note that, with this operation, Table 5 does not necessarily give the papers in decreasing order of fractional citation counts. So we have to rearrange Table 5 in order to have a table in which the $\frac{y_i}{\varphi(i)}$ are decreasing. We suppose this has been executed.

We now define the fractional h-index, denoted by h_f, as the largest rank $r = h_f$ such that

$$\frac{y_{h_f}}{\varphi\left(y_{h_f}\right)} \geq h_f \tag{1}$$

We define the fractional g-index, denoted by g_f as the largest rank $r = g_f$ such that

$$\sum_{i=1}^{g_f} \frac{y_i}{\varphi(i)} \geq g_f^2 \tag{2}$$

So these definitions are exactly the same as in the non-weighted case but we use the new Table 5.

We illustrate this on the simple example in Table 4 where we add co-authorship information (the values $\varphi(i)$): see Table 6.

Table 6. Simple example with co-authorship information

r	# authors	# citations
1	2	10
2	1	5
3	3	3
4	1	2
5	1	1
6	1	1

Paper 1 has 10 citations and 2 co-authors, so an author of such a paper receives a fractional citation count of 5. The second paper has only this author and hence keeps its 5 citations. The same author in the third paper receives a score 1 since there are 3 citations and 3 authors. Hence this score is lower than in the fourth paper: this paper keeps its 2 citations since this author is the only one. Finally, for the same reason, the fifth and sixth paper keep their 1 citation. For decreasing order we have to rearrange the papers as in Table 7 (papers 3 and 4 are interchanged).

Table 7. Calculation of h_f and g_f

r	#	r^2	$\sum$ #
1	5	1	5
2	5	4	10
3	2	9	$12 > 3^2$
4	1	16	$13 < 4^2$
5	1	25	14
6	1	36	15

It is now clear that $h_f = 2$ and $g_f = 3$.

II.3 Fractional h- and g-indexes using fractional paper counts

In this fractional counting method we leave the citation scores unchanged but we change the paper scores into the fractional counting method. This is a very classical way of giving paper scores to an author in a multi-authored paper: an author of a paper that has m authors in total receives a fractional score of $\frac{1}{m}$. These new paper counts replace the entire rankings $r = 1,2,...,T$, i.e. 1 is replaced by $\frac{1}{\varphi(1)}$, 2 is replaced by $\frac{1}{\varphi(1)} + \frac{1}{\varphi(2)}$ and so on. Table 8 makes this very clear, being the variant of Table 1 in case of fractional paper counts.

Table 8. Fractional paper counts

r	# citations
$1/\varphi(1)$	y_1
$1/\varphi(1) + 1/\varphi(2)$	y_2
$1/\varphi(1) + 1/\varphi(2) + 1/\varphi(3)$	y_3
$\vdots$	$\vdots$
$\sum_{i=1}^{T} 1/\varphi(i)$	y_T

Note that with this operation, the order of the papers is not changed since the values of the citation scores is not changed (of course, the rank values are changed !).

We now define the fractional h-index, denoted by h_F, as the largest rank $r = h_F$ such that

$$h_F = \sum_{i=1}^{k} \frac{1}{\varphi(i)} \le y_k \qquad (3)$$

Similarly we define the fractional g-index, denoted by g_F, as the largest rank $r = g_F$ such that

$$\left(\sum_{i=1}^{k} \frac{1}{\varphi(i)} \right)^2 \le \sum_{i=1}^{k} y_i \qquad (4)$$

where

$$g_F = \sum_{i=1}^{k} \frac{1}{\varphi(i)} \qquad (5)$$

Note that in this algorithm, evidently, the fractional h- and g-indexes, h_F and g_F, can be non-entire numbers, which is not an objection in itself.

We illustrate this on the same simple example in Table 6. Now fractional counting on papers, evidently, leads us to Table 9.

Table 9. Calculation of h_F and g_F

r	#	$\sum$ #
0.5	10	10
1.5	5	15
1.8333	3	18
2.8333	2	20
3.8333	1	$21 > (3.8333)^2$
4.8333	1	$22 < (4.8333)^2$

It is clear that now $h_F = 1.8333$ and $g_F = 3.8333$.

These simple definitions and examples illustrate the nature of the indexes h_f, g_f, h_F and g_F. In the next section, we present the mathematical theory of these indexes in relation with their non-weighted variants h and g and, of course, also featuring the values $\varphi(i)$ being the co-authorship distribution over the papers $i = 1,...,T$.

III. Mathematical theory of the indexes h_f, g_f, h_F and g_F

In the sequel (as we did before) we will denote by h and g the h-index and g-index, respectively, of the unweighted system as given by Table 1. We start by studying the fractional citation indexes h_f and g_f.

III.1 Mathematical theory of h_f and g_f

Let us denote, for every positive number $x \in \mathbb{R}^+$: $[x]$ = the greatest integer that is $\le x$. We have the following general result for h_f:

Theorem III.1.1: In all cases

$$\left[\frac{h}{\max\limits_{i=1,...,h} \varphi(i)}\right] \le h_f \le h \tag{6}$$

Proof: Let $r = 1,...,T$ denote the original ranks in the unweighted case: $y_1 \ge y_2 \ge ... \ge y_T$ and let π be this permutation of $\{1,...,T\}$ which gives the ranks in the fractional scoring system (fractional citation scores):

$$\frac{y_{\pi(1)}}{\varphi(\pi(1))} \geq \frac{y_{\pi(2)}}{\varphi(\pi(2))} \geq \ldots \geq \frac{y_{\pi(T)}}{\varphi(\pi(T))} \tag{7}$$

Consider the first h values

$$\left\{ \frac{y_{\pi(1)}}{\varphi(\pi(1))}, \ldots, \frac{y_{\pi(h)}}{\varphi(\pi(h))} \right\} \tag{8}$$

Suppose $\pi(j) \in \{1,\ldots,h\}$ for all $j = 1,\ldots,h.$ Then

$$\{y_{\pi(1)},\ldots,y_{\pi(h)}\} = \{y_1,\ldots,y_h\}$$

and hence, since $\varphi \geq 1$, $h_f \leq h$. If there is a $j \in \{1,\ldots,h\}$ such that $\pi(j) \notin \{1,\ldots,h\}$, then

$$\frac{y_{\pi(j)}}{\varphi(\pi(j))} \leq y_{\pi(j)} \leq y_i$$

for <u>all</u> $i = 1,\ldots,h$, since $\varphi \geq 1$, since $\pi(j) \notin \{1,\ldots,h\}$ and since the y_i are decreasing. In all these cases we have that

$$\frac{y_{\pi(j)}}{\varphi(\pi(j))} \leq \min_{i=1,\ldots,h} y_i$$

hence, necessarily, $h_f \leq h$ since $j \in \{1,\ldots,h\}$. This concludes the proof that $h_f \leq h$.

Let now

$$x = \frac{h}{\max_{i=1,\ldots,h} \varphi(i)} \tag{9}$$

For all $i = 1,\dots,[x] \leq h$ (since $\varphi \geq 1$) we have $y_i \geq h$ (definition of h) and hence

$$\frac{y_i}{\varphi(i)} \geq \frac{h}{\varphi(i)}$$

$$\geq \frac{h}{\max\limits_{j=1,\dots,h} \varphi(j)}$$

$$= x \geq [x] \geq i$$

Hence, for all $i = 1,\dots,[x]$ we have

$$\frac{y_i}{\varphi(i)} \geq i \qquad (10)$$

If these $[x]$ values $\frac{y_i}{\varphi(i)}$ constitute the $[x]$ largest values then we have proved that $h_f \geq [x]$ since h_f is the largest integer with property (10). If these first $[x]$ values $\frac{y_i}{\varphi(i)}$ do not constitute the $[x]$ largest values then some of these values are replaced by larger numbers $\frac{y_j}{\varphi(j)}$, $\exists j \in \{[x],\dots,T\}$, say replacing a smaller value $\frac{y_i}{\varphi(i)}\left(i \in \{1,\dots,[x]\}\right)$ but for which (10) is valid. Define the permutation $j = \pi(i)$ in this way, we hence have

$$\frac{y_{\pi(i)}}{\varphi(\pi(i))} \geq \frac{y_i}{\varphi(i)} \geq i \qquad (11)$$

This can be done for every $i = 1,\dots,[x]$ making the $\frac{y_{\pi(i)}}{\varphi(\pi(i))}, i = 1,\dots,[x]$ decreasing. Since h_f is the largest integer with property (11) we hence have $h_f \geq [x]$. □

A similar result will be proved for g_f. First we need a Lemma;

Lemma III.1.2:

$$\sum_{i=1}^{a} y_i \geq ga \tag{12}$$

for every $a \in \mathbb{N}, a \leq g$.

Proof: By definition of the g-index (unweighted) we have

$$\sum_{i=1}^{g} y_i \geq g^2 \tag{13}$$

Since $a = \left(\frac{a}{g}\right)g$ and since $y_1,...,y_T$ decreases we have by (13) that

$$\sum_{i=1}^{a} y_i \geq \left(\frac{a}{g}\right)g^2 = ag. \qquad \square$$

Theorem III.1.3: In all cases

$$\left[\frac{g}{\max\limits_{i=1,...,g} \varphi(i)}\right] \leq g_f \leq g \tag{14}$$

Proof:

The proof that $g_f \leq g$ can be read on the lines of the proof that $h_f \leq h$ in Theorem III.1.1.

Denote by

$$a = \left[\frac{g}{\max\limits_{i=1,...,g} \varphi(i)}\right] \tag{15}$$

If $a = 0$, we remark that (14) (left hand inequality) is trivial. Hence we suppose $a > 0$ hence, since a is an integer, $a \geq 1$.

Then $a \in \mathbb{N}$, $a \leq g$ (trivially) and hence we can apply Lemma III.1.2 yielding

$$\sum_{i=1}^{\left[\frac{g}{\max\limits_{i=1,\dots,g}\varphi(i)}\right]} y_i \geq \left[\frac{g}{\max\limits_{i=1,\dots,g}\varphi(i)}\right] g$$

Hence

$$\sum_{i=1}^{\left[\frac{g}{\max\limits_{i=1,\dots,g}\varphi(i)}\right]} \frac{y_i}{\varphi(i)} \geq \left[\frac{g}{\max\limits_{i=1,\dots,g}\varphi(i)}\right] \frac{g}{\max\limits_{i=1,\dots,g}\varphi(i)} \tag{16}$$

since

$$g \geq \left[\frac{g}{\max\limits_{i=1,\dots,g}\varphi(i)}\right]$$

Hence (16) gives

$$\sum_{i=1}^{\left[\frac{g}{\max\limits_{i=1,\dots,g}\varphi(i)}\right]} \frac{y_i}{\varphi(i)} \geq \left(\left[\frac{g}{\max\limits_{i=1,\dots,g}\varphi(i)}\right]\right)^2 \tag{17}$$

Let now π be this permutation of $\{1,\dots,T\}$ yielding decreasing values $\frac{y_{\pi(i)}}{\varphi(\pi(i))}, i = 1,\dots,T$.

Then, obviously

$$\sum_{i=1}^{\left[\frac{g}{\max\limits_{i=1,\dots,g}\varphi(i)}\right]} \frac{y_{\pi(i)}}{\varphi(\pi(i))} \geq \sum_{i=1}^{\left[\frac{g}{\max\limits_{i=1,\dots,g}\varphi(i)}\right]} \frac{y_i}{\varphi(i)} \geq \left(\left[\frac{g}{\max\limits_{i=1,\dots,g}\varphi(i)}\right]\right)^2 \tag{18}$$

by (17). Since the $\frac{y_{\pi(i)}}{\varphi(\pi(i))}$ are decreasing and since g_f is the largest integer with this property we have that

$$g_f \geq \left[\frac{g}{\max\limits_{i=1,\dots,g}\varphi(i)}\right]$$

completing the proof. □

We continue with the mathematical study of the indexes h_F and g_F.

III.2 Mathematical theory of h_F and g_F

Theorem III.2.1: In all cases

$$\sum_{i=1}^{h} \frac{1}{\varphi(i)} \leq h_F \leq h \tag{19}$$

Proof:

$$\sum_{i=1}^{h} \frac{1}{\varphi(i)} \leq h \leq y_h \tag{20}$$

since $\varphi \geq 1$ and by definition of h. Since in this case of fractional paper counts, the values $y_1, \dots, y_T$ are kept the same (see Table 10), the order of $y_1, \dots, y_T$ is unchanged (decreasing order). Since we have (20) and since h_F is the largest number such that

$$h_F = \sum_{i=1}^{k} \frac{1}{\varphi(i)} \le y_k \tag{21}$$

(see (3)), we have that $k \ge h$, hence

$$h_F = \sum_{i=1}^{k} \frac{1}{\varphi(i)} \ge \sum_{i=1}^{h} \frac{1}{\varphi(i)}$$

proving the left inequality in (19). By (21)

$$h_F \le y_k \tag{22}$$

and since $k \ge h$ we have, by definition of h:

$$y_k \le h \tag{23}$$

(22) and (23) yield $h_F \le h$, hence completing this proof. □

For g_F only one inequality is valid.

<u>Theorem III.2.2</u>: In all cases

$$\sum_{i=1}^{g} \frac{1}{\varphi(i)} \le g_F \tag{24}$$

<u>Proof</u>:

$$\left(\sum_{i=1}^{g} \frac{1}{\varphi(i)} \right)^2 \le g^2 \le \sum_{i=1}^{g} y_i \tag{25}$$

since $\varphi \ge 1$ and by definition of g. Now g_F is the largest number

$$g_F = \sum_{i=1}^{k} \frac{1}{\varphi(i)} \tag{26}$$

for which

$$\left(\sum_{i=1}^{k} \frac{1}{\varphi(i)} \right)^2 \le \sum_{i=1}^{k} y_i \tag{27}$$

(see (4) and (5)). So, by (25), (26) and (27) we have

$$g_F \ge \sum_{i=1}^{g} \frac{1}{\varphi(i)}. \qquad \square$$

Table 4 and Table 9 already yielded an example of $g_F < g$. Let us now present an example of $g_F > g$.

<u>Example III.2.3</u>: This is an example in which all papers have 5 authors

Table 10. Example of $g_F > g$

r	# authors	# citations	$\sum$ # cit.
1	5	10	10
2	5	5	15
3	5	3	18
4	5	2	$20 > 4^2$
5	5	1	$21 < 5^2$
6	5	1	22
$\vdots$	$\vdots$	$\vdots$	$\vdots$

From rank 5 on, all papers have 1 citation and have 5 authors. It is clear that $h = 3$ and $g = 4$. For the sake of completeness and illustration of the inequalities (19) we also calculate h_F : see Table 11.

Table 11. Calculation of h_F

r (fractional)	# cit.
0.2	10
0.4	5
0.6	3
0.8	2
1.0	1
1.2	1

We clearly have $h_F = 1 < h$ and also

$$h_F \geq \sum_{i=1}^{h} \frac{1}{\varphi(i)}$$

is satisfied since

$$0.6 = \sum_{i=1}^{h} \frac{1}{\varphi(i)} < h_F.$$

Now we calculate g_F using Table 12. It now follows that $g_F = 7.2 > g$. Note that

$$g_F \geq \sum_{i=1}^{g} \frac{1}{\varphi(i)}$$

is satisfied since

$$\sum_{i=1}^{g} \frac{1}{\varphi(i)} = 0.8 < g_F.$$

Table 12. Calculation of g_F

r (fractional)	$\sum$ # cit.
0.2	10
0.4	15
0.6	18
0.8	20
1.0	21
1.2	22
1.4	23
$\vdots$	$\vdots$
7.0	51
7.2	$52>(7.2)^2$
7.4	$53<(7.4)^2$

Remark: The above theory can also be applied to fractional country scores (instead of fractional author scores) cf. Egghe, Rousseau and Van Hooydonk (2000). Here $\varphi(i) \geq 1$ for all i is still valid but the values are not always entire (as in the case of number of authors). An example will illustrate this: suppose a paper i has 5 authors and that country c appears 3 times (i.e. 3 of these 5 authors are of country c). Then the fractional score $\frac{1}{\varphi(i)}$, used all over this paper (in the fractional citation scores as well as in the fractional paper scores) equals

$$\frac{1}{\varphi(i)} = \frac{3}{5}$$

hence $\varphi(i) = \frac{5}{3} \geq 1$. All the proved results are valid for these fractional h- and g-indexes of a country. The same can be said about research groups etc.

III.3 The inequalities, proved above, are optimal

It is easy to show that all inequalities proved above (i.e. (6), (14), (19) and (24)) are optimal and hence cannot be improved. This is clear for the right hand sides of the inequalities by taking any example where all papers are written by one author. Then $h_f = h_F = h$ and $g_f = g_F = g$.

For the left hand sides of the inequalities, we present the example in Table 13 (fictitious example).

Table 13. Fictitious example

r	# authors	# citations
1	5	5
2	5	5
3	5	5
4	5	5
5	5	5

Here $\varphi(i) = 5$ for all $i = 1,2,3,4,5$. Clearly $h = 5$. Now we have that

$h_f = h_F = 1 = \left[\frac{h}{\max_{i=1,\ldots,h} \varphi(i)}\right] = \sum_{i=1}^{h} \frac{1}{\varphi(i)}$, showing that the left hand sides of (6) and (19) can be reached. Also $g = 5$ and it is easy to see that $g_f = g_F = 1 = \left[\frac{g}{\max_{i=1,\ldots,g} \varphi(i)}\right] = \sum_{i=1}^{g} \frac{1}{\varphi(i)}$, showing that the left hand sides of (14) and (24) can be reached.

So, this example shows that the proved inequalities cannot be improved and they support the mathematical theory.

We close this paper by presenting and discussing the h, g, h_f, g_f, h_F and g_F values of the present author (on January 8, 2007). This will show that the extreme differences between the fractional and non-fractional h- and g-indices, obtained in the above example, are not true in practise.

IV. The indexes h, g, h_f, g_f, h_F and g_F for L. Egghe

A search in the Web of Science on January 8, 2007 yielded Table 13 for the present author from which h and g (unweighted) can be determined (we stop at $y_T = 8$ since we will not need more articles as will become clear further on). We added the number of authors of each paper for further use (first column).

Table 14. Calculation of the h- and g-index of L. Egghe
(as of January 8, 2007)

# aut.	r	r^2	# cit.	å #cit.
1	1	1	53	53
1	2	4	42	95
1	3	9	40	135
2	4	16	36	171
2	5	25	22	193
2	6	36	19	212
2	7	49	17	229
1	8	64	17	246
2	9	81	16	262
1	10	100	16	278
2	11	121	15	293
1	12	144	15	308
1	13	169	14	322
1	14	196	14	336
3	15	225	13	349
2	16	256	13	362
1	17	289	13	375
1	18	324	12	387
1	19	361	12	399
1	20	400	11	410 > 400
2	21	441	11	421 < 441
1	22	484	11	432
2	23	529	10	442
1	24	576	10	452
3	25	625	10	462
2	26	676	9	471
1	27	729	9	480
2	28	784	8	488
1	29	841	8	496
1	30	900	8	504

It is clear from Table 13 that $h = 14$ and $g = 20$ for the present author.

The fractional citation scores are presented in Table 14, sorted on the new ranks.

Table 15. Calculation of h_f and g_f

r (old)	r (new)	fract.cit.	å fract.cit.
1	1	53	53
2	2	42	95
3	3	40	135
4	4	18	153
8	5	17	170
10	6	16	186
12	7	15	201
13	8	14	215
14	9	14	229
17	10	13	242
18	11	12	254
19	12	12	266
5	13	11	277
20	14	11	288
21	15	11	299
6	16	9.5	308.5
7	17	8.5	317
9	18	8	$325 > (18)^2$
11	19	7.5	$332.5 < (19)^2$
16	20	6.5	339
15	21	4.333…	343.333…

The second column presents the new rankings according to the fractional citation scores (third column). Note that the fourth column presents the cumulative fractional citation scores according to the second column. We find $h_f = 12$ and $g_f = 18$.

Finally we calculate the h- and g-index h_F and g_F for fractional paper counts. Note that now the order of the papers does not change (but their rank values do !) – see Table 15.

Table 16. Calculation of h_F and g_F

r (fract.)	# cit.	å #cit.
1	53	53
2	42	95
3	40	135
3.5	36	171
4	22	193
4.5	19	212
5	17	229
6	17	246
6.5	16	262
7.5	16	278
8	15	293
9	15	308
10	14	322
11	14	336
11.3333	13	349
11.8333	13	362
12.8333	13	375
13.8333	12	387
14.8333	12	399
15.8333	11	410
16.3333	11	421
17.3333	11	432
17.8333	10	442
18.8333	10	452
19.1666	10	462
19.6666	9	471
20.6666	9	480
21.1666	8	488
22.1666	8	$496 > (22.1666)^2$
23.1666	8	$504 < (23.1666)^2$

From Table 15 it is clear that $h_F = 12.8333$ and that $g_F = 22.1666$.

V. Conclusions and suggestions for further research

In this paper, the h- and g-indexes of authors are extended to their "fractional" versions. This is done in two different ways. One method considers fractional citation counts where, for each m-authored paper a citation count of y is divided by m. The corresponding h- and g-index is denoted by h_f and g_f. Another method leaves the citation counts unchanged but replaces the ranks by the fractional paper count: a paper with m authors adds a fractional rank of $\frac{1}{m}$. The corresponding h- and g-index is denoted by h_F and g_F.

We present the mathematical theory of h_f, g_f, h_F and g_F in function of h and g (the non-fractional h and g index) and of the distribution $\varphi(i)$ of the number of authors in paper $i = 1,\dots,T$ (where there are T papers) and i is determined by the number of citations y_i, in decreasing order $y_1,\dots,y_T$.

We prove that, in all cases,

$$\left[\frac{h}{\max\limits_{i=1,\dots,h}\varphi(i)}\right] \le h_f \le h$$

($[x]$ denotes, for every positive number x, the largest entire number that is $\le x$). Similarly, but with a different proof, we show that, in all cases,

$$\left[\frac{g}{\max\limits_{i=1,\dots,g}\varphi(i)}\right] \le g_f \le g$$

Further we show that, in all cases,

$$\sum_{i=1}^{h}\frac{1}{\varphi(i)} \le h_F \le h$$

and

$$\sum_{i=1}^{g}\frac{1}{\varphi(i)} \le g_F$$

while we show by example that $g_F < g$ as well as $g_F > g$ are possible. We also show by example that these inequalities cannot be improved: all extreme values can be reached. We cannot perform better (by e.g. presenting functional equalities between these fractional indexes and their non-fractional counterparts h and g) without involving the citation scores y_i, $i = 1,...,T$. We leave it as an open problem to work out exact formulae for h_f, g_f, h_F and g_F in function of h, g, $\varphi(i)$ and y_i $(i = 1,...,T)$ but this is even an open problem for h and g in function of y_i, $i = 1,...,T$.

We remark that the indexes h_f, g_f, h_F and g_F and their theory can also be applied to the case of fractional country scores, where the $\varphi(i)$ remain ³ 1 but can have non-entire, rational values.

We close the paper by calculating h, g, h_f, g_f, h_F and g_F for this author and note that the corresponding h- and g-values are not far away from each other, hereby showing that the extreme cases are not true in practise.

We encourage to apply these indexes h, g, h_f, g_f, h_F and g_F to authors (journals, countries, institutes, topics, …) of several fields and to draw conclusions concerning their comparisons.

References

Ball, P. (2005). Index aims for fair ranking of scientists. Nature 436, 900.

Banks, M.G. (2006). An extension of the Hirsch index: indexing scientific topics and compounds. Scientometrics 69(1), 161-168.

Bornmann, L. and Daniel, H.-D. (2005). Does the h-index for ranking of scientists really work ? Scientometrics 65, 391-392.

Braun, T., Glänzel, W. and Schubert, A. (2005). A Hirsch-type index for journals. The Scientist 19(22), 8, November 2005.

Egghe, L. (2005). Power Laws in the Information Production Process: Lotkaian Informetrics. Elsevier, Oxford, UK, 2005.

Egghe, L. (2006a). Theory and practise of the g-index. Scientometrics 69(1), 131-152.

Egghe, L. (2006b). How to improve the h-index. The Scientist 20(3), 14, March 2006.

Egghe, L. (2006c). An improvement of the h-index: the g-index. ISSI Newsletter 2(1), 8-9.

Egghe, L. and Rousseau, R. (2006). An informetric model for the Hirsch index. Scientometrics 69(1), 121-129.

Egghe, L., Rousseau, R. and Van Hooydonk (2000). Methods for accrediting publications to authors or countries: consequences for evaluation studies. Journal of the American Society for Information Science 51(2), 145-157.

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

Glänzel, W. (2006b). On the opportunities and limitations of the H-index. Science Focus 1(1), 10-11 (in Chinese).

Hirsch, J.E. (2005). An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of the United States of America 102, 16569-16572. Original arXiv paper: arXiv:physics/0508025 v3, later corrected in: arXiv:physics/0508025 v5.

Popov, S.B. (2005). A parameter to quantify dynamics of a researcher's scientific activity. arXiv:physics/0508113.

Rousseau, R. (2007). The influence of missing publications on the Hirsch index. Journal of Informetrics 1(1), 2-7.

van Raan, A.F.J. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. Scientometrics 67(1), 491-502.