Journal of Informetrics 6 (2012) 557–565

Contents lists available at SciVerse ScienceDirect

Journal of Informetrics

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

Modifying h-index by allocating credit of multi-authored papers whose author names rank based on contribution

Xuan Zhen Liu[a], Hui Fang[b,*]

[a] Library, Nanjing Medical University, Nanjing 210029, China
[b] School of Electronic Science and Engineering, State Key Laboratory of Analytical Chemistry for Life Science, Nanjing University, Nanjing 210093, China

ARTICLE INFO

Article history:
Received 23 February 2012
Received in revised form 4 May 2012
Accepted 7 May 2012

Keywords:
Citations
h-Index
Co-authorship
Author rank
Contribution evaluation

ABSTRACT

In the present work we introduce a modification of the h-index for multi-authored papers with contribution based author name ranking. The modified h-index is denoted by h_{mc}-index. It employs the framework of the h_m-index, which in turn is a straightforward modification of the Hirsch index, proposed by Schreiber. To retain the merit of requiring no additional rearrangement of papers in the h_m-index and in order to overcome its shortage of benefiting secondary authors at the expense of primary authors, h_{mc}-index uses combined credit allocation (CCA) to replace fractionalized counting in the h_m-index. The h_m-index is a special form of h_{mc}-index and fits for papers with equally important authors or alphabetically ordered authorship. There is a possibility of an author of lower contribution to the whole scientific community obtaining a higher h_{mc}-index. Rational h_{mc}-index, denoted by h_{mcr}-index, can avoid it. A fictitious example as a model case and two empirical cases are analyzed. The correlations of the h_{mcr}-index with the h-index and its several variants considering multiple co-authorship are inspected with 30 researchers' citation data. The results show that the h_{mcr}-index is more reasonable for authors with different contributions. A researcher playing more important roles in significant work will obtain higher h_{mcr}-index.

1. Introduction

The h-index proposed by Hirsch (2005) is regarded as an easily determinable (Opthof & Wilde, 2009; Schreiber, 2009a) and a robust (Egghe & Rousseau, 2006) measure of scientific research output. Besides, the h-index is a better indicator in predicting future scientific achievement of researchers (Bornmann & Daniel, 2005; Hirsch, 2007). It has attracted extensive attention (Bornmann, Mutz, Hug & Daniel, 2011).

Hirsch (2010, p. 742) pointed out: "Perhaps the most important shortcoming of the h-index is that it does not take into account in any way the number of co-authors of each paper". It gives all the citations, as full credit, of a paper evenly to every author of the paper. Thus, the credit of multi-authored papers is inflated (Chai, Hua, Rousseau, & Wan, 2008; Hagen, 2008). In addition, comparing scientific research output by using the h-index sometimes cannot discriminate researchers with the same h-index, as the h-index is an integer number (Batista, Campiteli, Kinouchi, & Martinez, 2006).

Several modified h-indices accounting for multi-authored papers have been proposed. Some of the modifications only take the number of authors into account. Batista et al. (2006) divide the h-index of a scientist by the average number of authors of the papers in h-core. Egghe (2008) and Schreiber (2008a, 2008b) count the credit of a paper fractionally according

* Corresponding author at: No. 22, Hankou Road, Nanjing 210093, China.
E-mail address: huifang_nju@yahoo.com.cn (H. Fang).

1751-1577/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.joi.2012.05.002

to the inverse number of authors. These works distribute the credit of a paper evenly to all its authors. They are fit for papers in some scientific subfields that observe the convention of alphabetically ordering authors, such as high energy physics. But not all the alphabetically ordered authorship papers mean that all authors contribute equally. Sometimes it is only a custom (Egghe, Rousseau, & Van Hooydonk, 2000).

However, in other fields with major papers throughout the scientific community, authors' names are ranked according to their roles (Hu, Rousseau, & Chen, 2010). Author rank is shifting from alphabetic ordering to a contribution-based ordering (He, Ding, & Yan, 2012). The common method is to rank the authors by the descending order of the authors' contributions (Hodge & Greenberg, 1981; Lukovits & Vinkler, 1995; Trueba & Guerrero, 2004; Van Hooydonk, 1997), except for the corresponding author(s). The *h-maj* presented by Hu et al. (2010) and the weighted h-index (w) presented by Zhang (2009) give all the citations to all the first and corresponding author(s). When any corresponding author is not the first author, *h-maj* and w inflate the fame of the paper. The w-index also exaggerates the credit of the multi-authored papers even when the papers are without a corresponding author or the only corresponding author is the first one.

Hagen (2008) put forward the Harmonic h-index. The Harmonic h-index first allocates citations of every paper of a researcher according to his/her rank in the paper and the paper's author number. Then the Harmonic h-index uses the allocated citation number to replace the original citation number of every paper of the researcher in determining h-index. With normalization in allocating citations, this index does not inflate credit of papers.

Recently Liu and Fang (2012) proposed CCA h-index, denoted as h_c. After rearranging the author order to let the corresponding author tie for the first rank, the normalized credit allocated proportion of the i-th author in an n-authored paper is:

$$p(i,n) = \frac{i^{-q}}{\sum_{j=1}^{n} j^{-q}} \tag{1}$$

The first author and corresponding author(s) are defined as the most important authors in (Liu & Fang, 2012) or MIA for short. If a researcher is one of the MIAs, his/her normalized credit allocated proportion in a paper is:

$$p(i,n) = \frac{\sum_{j=1}^{n_{\mathrm{MIA}}} j^{-q}}{n_{\mathrm{MIA}} \sum_{j=1}^{n} j^{-q}} \tag{1$'$}$$

where n_{MIA} is the number of MIAs of the paper. If the paper has c citations, the allocated citation to the i-th author is

$$c_c = c \times p(i,n) \tag{2}$$

Similar to the Harmonic h-index, h_c uses c_c to replace c of every paper of a researcher in determining the h_c-index.

If q in Eq. (1) or (1′) is 0, then c_c equals c/n, which is the fractional counting (Egghe, 2008). If $q = 1$, then c_c is the Harmonic counting (Hagen, 2008). In other words, the fractional counting and Harmonic counting are special forms of Eq. (2) with $q = 0$ and 1. Thus, Eq. (2) with q in [0,1] is between these two methods, named as combined credit allocation (CCA). Model-experiments show that h_c with $q = 2/3$ is more suitable for a wide range of author numbers (Liu & Fang, 2012) and this value of q will be used in the subsequent analysis.

Applying CCA to modify h-index accounting for multi-authored papers, one needs extra reordering of the papers after citations are allocated. In view of this point, the strategy in h_m (Schreiber, 2008a, 2008b) which fractionally counts papers (Egghe, 2008) is more intuitive and convenient. In this paper, we will modify h_m-index with the CCA method instead of fractionalized counting for contribution-based author ranked papers.

The rest of this paper is organized as follows. In the second part, we modify h_m-index with CCA, denoted by h_{mc}-index. Then, the rational h_{mc}-index, h_{mcr}-index, is proposed to eliminate a bug in the h_{mc}-index and make it more applicable. In the fourth part, two empirical examples are employed to compare h-, h_m-, h_c-, h_{mc}-indices and their rational variants. In the fifth part, we investigate the correlations of h_{mcr}-index with h-index and its variants incorporating co-authorship. Further discussions are given in the last part.

2. Modifying h_m-index for contribution based ranking paper

Usually, more important authors rank before less important ones in papers which list authors according to their contribution. The most common exception is that the corresponding author may be at any position in the author list, which is taken into considered in the Harmonic h-index (Hagen, 2008) and h_c-index (Liu & Fang, 2012). The first and corresponding author(s) are the most important authors (MIAs) of a paper. Up to now, the information which can be used in allocating credit of a paper to each author is the author number and author rank. Fractionalized counting adopted in h_m-index only makes use of the information of author number for multi-authored papers. In the following, we modify h_m-index with Eq. (1) or (1′).

Table 1
A fictitious example of a model dataset with 10 publications of an author, put into order according to the number of citations $c(r)$ for paper rank r. Those papers which contribute to the h-, the h_m-, the h_{mc}-, and the h_c-index are indicated by bold face in the first, fourth, fifth and seventh column, respectively.

r	$n(r)$	$rank(r)$	$r_{eff}(r)$	$p_{eff}(r)$	$c(r)$	$c_c(r)$
1	1	1	**1.00**	**1.00**	56	**56**
2	4	1	**1.25**	**1.40**	33	**13.2**
3	3	1	**1.58**	**1.87**	29	**13.63**
4	4	2	**1.83**	**2.12**	13	3.25
5	2	1	**2.33**	**2.73**	8	**4.91**
6	4	2	**2.58**	**2.98**	6	1.5
7	4	2	**2.83**	**3.23**	4	1
8	1	1	3.83	4.23	3	3
9	2	1	4.33	4.84	2	1.23
10	4	3	4.58	5.03	2	0.38

Schreiber (2008a, 2008b) proposed the effective rank concept in h_m-index scheme to replace the paper number in the original h-core. Let r be the rank that is attributed to a paper when the publication list of an author is sorted by the number of citations $c(r)$. For the top r papers, one obtains an effective rank:

$$r_{eff}(r) = \sum_{r'=1}^{r} \frac{1}{n(r')} \tag{3}$$

where $n(r')$ is the author number of the r'-th paper, $1/n(r')$ is the fractional credit the author obtained.

Our modification is using credit allocation in Eq. (1) or (1′) instead of $1/n(r')$ for papers of contribution based author ranking. Thus, authors playing different roles will gain different credit. In order to avoid confusion of the word "rank" in "effective rank" and "author rank", we use effective paper count to replace effective rank. For the top r papers, one obtains an effective paper count

$$p_{eff}(r) = \sum_{r'=1}^{r} p(rank(r'), n(r')) \tag{4}$$

where the function $p(i, n)$ is defined in Eq. (1) or (1′). $rank(r')$ is the author rank in the r'-th paper, and $n(r')$ is the author number of the r'-th paper.

Hirsch's index h is determined from

$$h = \max_{r}(r \le c(r)) \tag{5}$$

where $c(r)$ is the citation number of the r-th paper.

The h_m-index is defined as (Schreiber, 2008a)

$$h_m = \max_{r}(r_{eff}(r) \le c(r)) \tag{6}$$

in correspondence with Eq. (5).

Similarly, we define h_{mc}-index as

$$h_{mc} = \max_{r}(p_{eff}(r) \le c(r)) \tag{7}$$

For convenience, we define r_h as the maximum r that $r_{eff}(r) \le c(r)$ for h_m-index or $r_{eff}(r) \le c(r)$ for h_{mc}-index.

Table 1 shows a fictitious example as a model case to demonstrate the determination of h_{mc}-index and its behavior in contrast to the h_m-index and h_c-index. The papers are listed in descending order by their number of citations in the sixth column. Obviously $h = 6$. Counting papers fractionally according to the inverse of the author number yields the effective rank r_{eff} given in the fourth column. In the seventh row, the r_{eff} is beneath the number of citations, while in the eighth row, r_{eff} is beyond the number of citations. Thus, r_h is 7 for the h_m-index. Consequently $h_m = 2.83$. One more paper has entered into the h_m-core due to the fractionalized counting. Counting the citations fractionally according to Eq. (2) as in the last column of Table 1, the fourth paper dropped out of the h_c-core, so that the h_c-index is obtained as $h_c = 4$ after rearranging the papers according to the values in the last column.

The effective paper count p_{eff} in the fifth column is calculated as the non-equal fractionally allocation to each author according to Eq. (1). For the h_{mc}-index, r_h is also 7. Thus, $h_{mc} = 3.23$.

The researcher with the papers in Table 1 played a relatively important role in these papers. Except in the last paper, his normalized credit allocated proportion, as defined in Eq. (1), of each paper is higher than the inverse of author number, so his $p_{eff}(r)$ is higher than $r_{eff}(r)$. It is reasonable that his h_{mc} is higher than h_m.

But in some cases, the results of definitions in Eqs. (6) and (7) are inconsistent. If the citation number of the eighth paper is 4, not 3 as in Table 1 and other values remain unchanged, the h_m rises to 3.83, whereas h_{mc} is still 3.23 because one more paper enters h_m-core and the h_{mc}-core does not increase, see Case 1 in Table 2. Or, if the author number of the eighth paper

Table 2
The h_{mr} and h_{mcr}-indices for different cases of the eighth paper in Table 1. Those papers which contribute to the h_m-, and the h_{mc}-index are indicated by bold face in the fifth and sixth column, respectively.

Case	r	$n(r)$	$rank(r)$	$r_{eff}(r)$	$p_{eff}(r)$	$c(r)$	h_{mr}	h_{mcr}
1	7	4	2	2.83	**3.23**	4	3.86	4
	8	1	1	**3.83**	4.23	4		
	9	2	1	4.33	4.84	2		
2	7	4	2	2.83	3.23	4	3.46	3.87
	8	2	1	**3.33**	**3.84**	4		
	9	2	1	3.83	4.45	2		

increases to 2 and its citation number is 4, the h_m and h_{mc} increase to 3.33 and 3.84, see Case 2 in Table 2. In this case, h_{mc} being higher than h_m is reasonable. But this h_{mc} is illogically higher than that when the author number of the eighth paper is 1, because the author did more work when he did it alone. The h_m-index may also encounter this problem. For example, if the author number of the eighth paper is changed to 10 in Table 1, the h_m-index will increase to 2.93, higher than that when the author did it alone. To avoid these unreasonable results, we adopt the idea of the rational variant of h-index for the h_m and h_{mc}-indices.

3. Rational variant of *h*-indices

For general cases, each paper of a researcher obtains a score when his/her papers are ranked in descending order by their number of citations. Suppose $s(r)$ is the score of the r-th paper. In h-index and h_c-index, $s(r)=r$. In h_m-index, $s(r)=r_{eff}(r)$. And in h_{mc}-index, $s(r)=p_{eff}(r)$.

Fig. 1 shows the method to determine rational h-indices.

Definition of h_{xr}: Let point $A_r(s(r), c(r))$ denote score and number of citations (or number of allocated citations) of the r-th paper. And let $y=c(x)$ denote its piecewise linear interpolation. Then the rational h_x-index, h_{xr}, is the abscissa of the intersection of the function $c(x)$ and the angle bisector $y=x$.

In the above definition, h_x represents h-, h_m-, h_c- or h_{mc}-index.

Now we will determine the formula for h_{xr}. In h-index or h_c-index, r_h equals h or h_c. As $h_x=s(r_h)\le h_{xr}$ and $c(r_h+1)\le s(r_h+1)$, this intersection is situated on the line segment connecting $(s(r_h), c(r_h))$ and $(s(r_h+1), c(r_h+1))$. Consequently: $s(r_h)\le h_{xr}<s(r_h+1)$.

The line connecting $(s(r_h), c(r_h))$ and $(s(r_h+1), c(r_h+1))$ has the following equation:

$$y = c(r_h) + \frac{c(r_h+1) - c(r_h)}{s(r_h+1) - s(r_h)}(x - s(r_h)) \tag{8}$$

Setting $y=x$ determines the intersection and then by definition $h_{xr}=x$.

$$h_{xr} = \frac{c(r_h)s(r_h+1) - c(r_h+1)s(r_h)}{s(r_h+1) - s(r_h) + c(r_h) - c(r_h+1)} \tag{9}$$

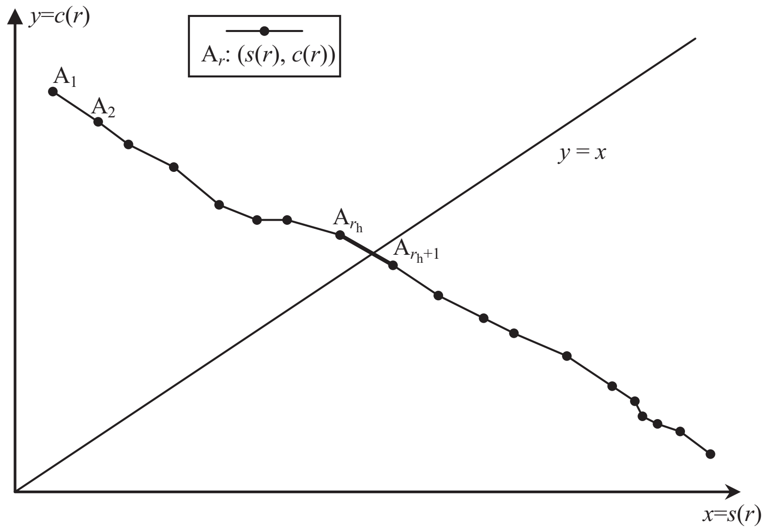

Fig. 1. Determination of rational variant of h-indices. Point A_r represents the score, $s(r)$ and citation number, $c(r)$ of r-th paper of a researcher.

Table 3
The citation records for the 32 most cited papers of Z. Those papers which contribute to the h-, the h_m-, the h_{mc}-, and the h_c-index are indicated by bold face in the first, fifth, sixth, and eighth column, respectively.

r	$n(r)$	$n_{MIA}(r)$	$rank(r)$	$r_{eff}(r)$	$p_{eff}(r)$	$c(r)$	$c_c(r)$
1	5	2	5	**0.20**	**0.12**	50	6.00
2	5	2	1	**0.40**	**0.41**	48	**13.73**
3	7	2	1	**0.54**	**0.64**	43	**10.23**
4	3	2	1	**0.88**	**1.03**	43	**16.60**
5	4	2	1	**1.13**	**1.36**	42	**13.65**
6	7	2	1	**1.27**	**1.59**	29	6.90
7	5	1	5	**1.47**	**1.71**	24	2.88
8	5	2	4	**1.67**	**1.85**	22	3.06
9	5	2	1	**1.87**	**2.14**	22	6.29
10	2	2	1	**2.37**	**2.64**	20	**10.00**
11	2	2	1	**2.87**	**3.14**	20	**10.00**
12	5	1	3	**3.07**	**3.31**	19	3.21
13	4	2	3	**3.32**	**3.50**	17	3.26
14	4	2	1	**3.57**	**3.82**	17	5.53
15	7	1	1	**3.71**	**4.12**	16	4.67
16	7	2	3	**3.85**	**4.26**	16	2.25
17	5	1	1	**4.05**	**4.61**	15	5.26
18	4	2	4	**4.30**	**4.77**	14	2.22
19	7	2	1	**4.45**	**5.00**	13	3.09
20	4	1	1	**4.70**	**5.40**	12	4.79
21	8	2	1	**4.82**	**5.62**	12	2.66
22	6	2	1	**4.99**	**5.88**	12	3.10
23	9	2	1	**5.10**	**6.09**	11	2.29
24	3	2	1	**5.43**	**6.48**	11	4.25
25	3	1	1	**5.77**	**6.95**	11	5.21
26	3	2	1	**6.10**	**7.34**	9	3.48
27	7	2	6	**6.24**	**7.43**	9	0.80
28	6	1	6	**6.41**	**7.52**	9	0.86
29	5	2	5	**6.61**	**7.64**	8	0.96
30	8	2	1	**6.73**	**7.86**	8	1.77
31	6	2	1	**6.90**	8.12	8	2.07
32	3	2	1	7.23	8.51	7	2.70

For h_r and h_{cr}-indices, h_{xr} can be simplified to:

$$h_{xr} = \frac{c(h)(h+1) - c(h+1)h}{1 + c(h) - c(h+1)} \quad (9')$$

which is same as Eq. (2) of (Chai et al., 2008) and Eq. (6) of (Guns & Rousseau, 2009). In addition, the equivalent form of Eqs. (9) and (9′) have been used in the interpolated h_m-index (Schreiber, 2009b) and h-index (Schreiber, 2008c). Similarly, the interpolation method has also been applied in g-index (Guns & Rousseau, 2009) or fractional g-index (Schreiber, 2009c, 2010).

From Fig. 1 and Eq. (9) it is easy to see that $h_{xr} = c(r_h)$ in two cases. One is $c(r_h) = s(r_h)$, the other is $c(r_h + 1) = c(r_h)$. The first case is that point A_{r_h} is on the line of $y = x$. The second case is that the line segment between A_{r_h} and A_{r_h+1} is horizontal.

Table 2 shows two aforementioned cases that induce confused h_m and h_{mc}-indices. When the citation number of the eighth paper is changed to 4 and other values remain unchanged as in Table 1, h_{mcr} is 4 and is higher than h_{mr} which is 3.86, despite the fact that one more paper enters h_{mr}-core and the h_{mcr}-core does not increase. When the author number of the eighth paper increases to 2 and its citation number is 4, the h_{mr} and h_{mcr} are 3.46 and 3.87. In this case, h_{mr} and h_{mcr} are all lower than when the author number of the eighth paper is 1. It is accordant to the fact that the single author does higher proportion of work than any one of multi-authors.

4. Empirical examples

To demonstrate the determination of the h_{mcr}-index for an empirical dataset, we use real data from the Science Citation Index provided by Thomson Scientific in the WoS. Table 3 lists the 32 most cited publications of a professor at Nanjing University, denoted by Z. The 16th top paper has 16 citations, so Z's h and h_r-indices are both 16. The 30th top paper has p_{eff} of 7.86 and 8 citations, and those of the 31th top paper are 8.12 and 8 respectively. Thus the r_h for h_{mc}-index is 30, and it implies that h_{mc}-core has 30 papers extending from h-core of 16 papers. According to definition in Eq. (7), h_{mc} is 7.86. Because the citation numbers of r_h-th and $r_h + 1$-th are both 8, h_{mcr} is 8.

For comparison, r_{eff} for h_m-index is also calculated from Table 3. In 20 of the 30 top publications, Z is MIA. As consequence of his important roles in these papers, p_{eff} is higher than r_{eff} since the second paper. Thus, one more paper is in h_m-core than in h_{mc}-core, and r_h for h_m-core is 31. h_m and h_{mr}-indices are determined as 6.9 and 7.17 from the data in Table 3. Z's h_{mcr} is higher than h_{mr} because of his important roles in his publications.

Table 4
The citation records for the 17 most cited papers of W. Those papers which contribute to the h-, the h_m-, the h_{mc}, and the h_c-index are indicated by bold face in the first, fifth, sixth, and eighth column, respectively.

r	$n(r)$	$n_{MIA}(r)$	$rank(r)$	$r_{eff}(r)$	$p_{eff}(r)$	$c(r)$	$c_c(r)$
1	4	1	1	**0.25**	**0.40**	34	**13.56**
2	1	1	1	**1.25**	**1.40**	29	**29.00**
3	2	1	1	**1.75**	**2.01**	20	**12.27**
4	2	1	1	**2.25**	**2.63**	17	**10.43**
5	1	1	1	**3.25**	**3.63**	17	**17.00**
6	2	2	1	**3.75**	**4.13**	14	7.00
7	2	1	1	**4.25**	**4.74**	13	**7.98**
8	2	2	1	**4.75**	**5.24**	12	6.00
9	2	1	1	**5.25**	**5.85**	12	**7.36**
10	2	2	1	**5.75**	**6.35**	11	5.50
11	3	1	1	**6.08**	**6.83**	11	5.21
12	4	2	1	**6.33**	**7.15**	10	3.25
13	3	1	1	**6.67**	**7.63**	9	4.26
14	3	1	1	**7.00**	8.10	8	3.79
15	3	1	1	**7.33**	8.57	8	3.79
16	3	1	1	**7.67**	9.05	8	3.79
17	2	1	1	8.17	9.66	7	4.29

In the last column of Table 3, the allocated citation, c_c, of each paper is counted by Eq. (2). Rearranging the papers according to the values in the last column, only 6 papers satisfy that $r \le c_c(r)$ so that Z's CCA h-index, h_c, is 6. $c_c(h_c)$ and $c_c(h_c+1)$ are 10 and 6.9 respectively. Applying Eq. (9′), h_{cr} is obtained as 6.98. The 6 papers in h_c-core contain relatively more citations and important roles of Z played in the paper. The other 10 papers which are in the h-core and ruled out of h_c-core can be classified into three cases. The first case is that the paper has lower citations, such as the 12–16th papers in the h-core. The second case is that Z plays fewer roles in the paper. The most outstanding example is the first paper, as it has many more citations than other papers, but Z ranks at the last position in the 5 coauthors. The 7th or 8th papers are also in this category. The last case is that the paper has relatively more authors. Z is one of the two MIAs in the 6th and 9th papers which have more citations than the 10th and 11th papers, but the 10th and 11th papers have only two authors while the 6th and 9th papers have 7 and 5 authors respectively.

Table 4 lists the 17 most cited publications of another professor at Nanjing University, denoted by W. The 11th top paper has 11 citations, so W's h and h_r-indices are both 11. The 13th top paper has p_{eff} of 7.63 and 9 citations, and those of the 14th top paper are 8.10 and 8 respectively. Thus the r_h for h_{mc}-index is 13. According to definition in Eqs. (7) and (9), h_{mc} and h_{mcr} are 7.63 and 8.08 respectively.

W is MIA in all the 17 papers in Table 4. Thus his p_{eff} is higher than r_{eff}. Three more papers are in h_m-core than in h_{mc}-core, and r_h for h_m-index is 16. h_m and h_{mr}-indices are determined as 7.67 and 7.78 from the data in Table 4. W's data show that rational h_m- and h_{mc}-indices amend the bug of h_m and h_{mc}-indices. The fact that h_{mcr} is higher than h_{mr} is consistent with W being the MIA in these papers, but his h_{mc} is lower than h_m.

Seven of the nine most cited papers enter h_c-core, as in the last column of Table 4. Thus, the CCA h-index, h_c is 7. $c_c(h_c)$ and $c_c(h_c+1)$ are 7.36 and 7 respectively. Applying Eq. (9′), h_{cr} is obtained as 7.26. The sixth and eighth papers have slightly more citations than the ninth paper, and they all have two authors. But W is the only one MIA in the ninth paper and earns a more normalized credit allocated proportion in this paper than in the sixth and eighth papers in which the two authors are equally important. Therefore, it is understandable that the sixth and eighth papers are excluded from h_c-core.

Z's papers have more citations than W's, so the former has a higher h-index. But most of W's papers have fewer authors than those of Z's papers. This indicates that W has more contribution proportionally in his papers than Z. The final result is that W has slightly higher h_{cr}-, h_{mr}-, and h_{mcr}-indices, though his h-index is obviously lower than Z's.

5. Correlation of h_{mcr}-index with h-index and its variants

Bornmann et al. (2011) recently investigated the correlations between h-index and its 37 variants. These variants modify h-index in different aspects, such as field dependence, self-citations, multi-authorship, career length, the age of publications and citation intensity. In our opinion, a mature h-index should have capability to remedy all the defects in the original h-index. Before h-index reaches to that level, it needs to be ascertained which aspects should be taken into account in modifying h-index and how to modify it. The present work only discusses the modification of h-index in consideration of the multi-authorship and we are not proposing a new index. We denote it by h_{mcr}-index in order to discriminate it from the original h-index and its other modifications. We think all other bibliometric indices of researchers based on publications, such as total number of papers, total number of citations, citations per paper, number of "significant papers", etc., should incorporate author number and author rank.

In this part, we inspect the correlations of h_{mcr}-index with the interpolated h-index, h_r (Schreiber, 2008c) and its variants considering multiple co-authorship using publication information of 30 professors and associate professors from Nanjing University and Nanjing Medicine University. In Table 5, h_{mr} is the interpolated h_m-index (Schreiber, 2009b). h_I is proposed

Table 5
Hirsch indices without and with taking multiple co-authorship into account. The last two columns show the orders in which the datasets appear after the list is sorted according to the h_{mr}-index and h_{mcr}-index.

Dataset	h_{r}	h_{mr}	h_I	*h-maj*	h_{first}	Weighted h	h_{mcr}	$O(h_{\mathrm{mr}})$	$O(h_{\mathrm{mcr}})$
R1	28.25	13.00	6.17	24	2	24	14.66	1	1
R2	26.00	12.01	5.16	24	5	22	13.68	3	2
R3	25.00	10.39	4.88	23	9	22	12.85	5	3
R4	24.50	6.42	2.18	22	5	22	9.00	13	7
R5	23.35	9.00	2.65	15	1	17	9.19	6	6
R6	21.50	5.38	1.98	12	12	14	6.27	24	20
R7	19.80	6.78	2.93	16	3	12	7.72	12	12
R8	17.67	6.13	2.75	15	6	15	8.00	16	10
R9	17.50	12.20	3.71	14	9	16	12.17	2	4
R10	16.00	7.18	3.32	12	7	12	8.00	9	10
R11	16.00	6.00	2.42	12	7	12	6.83	18	15
R12	15.00	5.85	2.50	11	11	12	7.39	20	13
R13	14.50	5.65	3.32	7	7	9	5.44	21	26
R14	14.00	5.94	2.68	8	2	7	5.55	19	25
R15	14.00	5.59	1.72	8	3	7	5.24	23	27
R16	13.80	7.00	3.13	8	3	10	6.54	10	17
R17	13.67	7.00	2.04	9	7	12	7.01	10	14
R18	13.00	9.00	4.97	11	10	12	9.00	6	7
R19	13.00	6.37	2.91	7	5	8	6.54	14	17
R20	12.67	4.46	2.53	9	5	9	5.15	28	28
R21	12.50	6.13	2.88	8	4	9	5.99	16	22
R22	11.67	5.62	1.92	11	11	11	6.53	22	19
R23	11.67	5.01	1.59	9	4	7	5.84	25	23
R24	11.50	4.77	2.69	11	7	9	5.83	27	24
R25	11.33	4.97	1.83	9	3	9	6.03	26	21
R26	11.00	10.75	8.64	10	10	11	10.58	4	5
R27	11.00	7.78	5.26	11	9	11	8.07	8	9
R28	10.00	6.37	2.70	9	5	9	6.58	14	16
R29	10.00	4.00	1.79	4	4	6	3.94	29	29
R30	7.67	2.39	0.96	7	7	7	2.85	30	30

by Batista et al. (2006) trying to quantify an individual's scientific research output valid across disciplines. *h-maj* includes only the papers in which the research is the first or corresponding author (Hu et al., 2010). h_{first} is the h index of first authored papers (Opthof & Wilde, 2009). Weighted h gives full credit for the first and last authorship, half credit for the second authorship, and a quarter credit for other authors (Lee, Kraus, & Couldwell, 2009).

Table 6 lists Pearson's correlation coefficients between h_{mcr} and h_{r}, h_{mr}, h_I, *h-maj*, h_{first}, weighted h from Table 5. h_{mcr} correlates with h_{mr} the most. It is intelligible because h_{mcr} adopts the framework of h_{mr}. The only difference between the two modifications is that h_{mcr} incorporates author rank of every paper. The last two columns show the orders in which the researchers appear after the list is sorted according to the h_{mr} and h_{mcr}. Only 6 of the 30 researchers have same place in the two lists. For example, researcher R4 is a leader of a large group and he is often the MIA and his papers usually have 10–20 authors. He has the lowest ratio of $h_{\mathrm{mr}}/h_{\mathrm{r}}$. As h_{mcr} gives him more credit of his papers than h_{mr}, he has the highest ratio of $h_{\mathrm{mcr}}/h_{\mathrm{mr}}$. Accordingly, his rank based on h_{mcr} is 7 raised from rank 13 in h_{mr}-sorted list. h_{mcr} employs the information of author's contribution, avoiding underrating important authors, especially in larger groups. Our modification is promising because co-author numbers keep rising (Sekercioglu, 2008). On the other hand, the role of bibliometric indices is to compare scientific research output among researchers. Researchers have different ranks in h_{mr}- and h_{mcr}-sorted lists, and h_{mcr} applies more useful information representing author's role. Therefore h_{mcr} is an essential amelioration of h_{mr}. Similarly, the correlation coefficient between h-index and g-index is as high as 0.975 (Schreiber, 2008c). This does not encumber that g-index is an effective modification of the h-index, as g-index can measure both the actual scientific productivity and the scientific impact of a scientist, whereas the h-index measures mostly the quantity dimension (Schreiber, Malesios, & Psarakis, 2012).

6. Further discussion and summary

For the papers whose author names are listed based on author's contribution, the present investigation has compared different ways to take the number of authors and author rank into account in the determination of Hirsch-type indices. The

Table 6
Correlation between h_{mcr} and h_{r}, h_{mr}, h_I, *h-maj*, h_{first}, weighted h from Table 5.

	h_{r}	h_{mr}	h_I	*h-maj*	h_{first}	Weighted h
Correlation coefficient with h_{mcr}	0.749	0.956	0.728	0.845	0.091	0.871

modification of h-index into h_m-index (Schreiber, 2008a, 2008b) is convenient because it does not require rearrangement of the papers. But in addition to the fact that it cannot distinguish different contribution among the authors, h_m-index may be undesirable small for important authors when the number of authors increases (Liu & Fang, 2012). The h_c-index proposed by Liu and Fang (2012) makes use of the information of author rank which is not embodied in the h_m-index. But the h_c-index needs rearrangement of the papers. To combine the advantages of h_m- and h_c-indices, we present h_{mc}-index here. The h_{mc}-index adopts the framework of the h_m-index and replaces the effective number of publications, $r_{eff}(r)$, which is the summation of the inverse of author number of the r most cited papers with effective paper count, $p_{eff}(r)$, which in turn is the summation of normalized credit allocated proportion of the author obtained in each of the r most cited papers.

In the actual usage of the h_{mc}-index, one may encounter a problem. If a researcher plays fewer roles in his/her $r_h + 1$-th paper, he/she may obtain higher h_{mc}-index. The same problem also exists in the h_m-index. To overcome this problem, we employ the method of rational h-index to ameliorate h_x-index (h_x is the general term for h-, h_m-, h_c- and h_{mc}-index). The rational h_x-index, denoted as h_{xr}, is derived in Eqs. (9) and (9′). It takes the data of the $r_h + 1$-th paper into consideration to eliminate the inconsistency in the h_x-index, and makes the Hirsch-type indices more robust. This has been shown in the above examples.

Different from h_c-core which comprises only the papers of higher citation number and more important contributions from the researcher, h_{mc}-core includes the papers of higher citation number whether the researcher is important or not. A highly cited paper dedicates less to the p_{eff} of an author who contributes less to the paper, therefore influences his/her h_{mc}-index less. The h_{mc}-core containing such paper reflects the researcher's integrative experience in important work. Therefore, h_{mcr}-index provides more information of a researcher than h_c-index. Obviously, a researcher with more contributions in important work will gain a higher h_{mcr}-index.

Relative to the original h-index, the additional information of papers used by h_m-index is author number. Similarly, h_{mc}-index and h_{mcr}-index add author rank to the information used by h_m-index. The newly modified index utilizes more information to increase its rationality, but at the cost of more operations to process data. However, the powerful ability of the computer today enables software to handle these tasks quite easily.

Acknowledgements

We are grateful to Wendy Powell of Valleyford, Washington, USA, for editing the English of our manuscript. The authors thank the anonymous reviewers for useful suggestions improving this contribution.

References

Batista, P. D., Campiteli, M. G., Kinouchi, O., & Martinez, A. S. (2006). Is it possible to compare researchers with different scientific interests? *Scientometrics*, *68*(1), 179–189.
Bornmann, L., & Daniel, H. D. (2005). Does the *h*-index for ranking of scientists really work? *Scientometrics*, *65*(3), 391–392.
Bornmann, L., Mutz, R., Hug, S. E., & Daniel, H. D. (2011). A multilevel meta-analysis of studies reporting correlations between the *h* index and 37 different *h* index variants. *Journal of Informetrics*, *5*(3), 346–359.
Chai, J. C., Hua, P. H., Rousseau, R., & Wan, J. K. (2008). The adapted pure *h*-index. In H. Kretschmer, & F. Havemann (Eds.), *Proceedings of WIS 2008* Berlin.
Egghe, L. (2008). Mathematical theory of the *h*- and *g*-index in case of fractional counting of authorship. *Journal of the American Society for Information Science and Technology*, *59*(10), 1608–1616.
Egghe, L., & Rousseau, R. (2006). An informetric model for the Hirsch-index. *Scientometrics*, *69*(1), 121–129.
Egghe, L., Rousseau, R., & Van Hooydonk, G. (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.
Guns, R., & Rousseau, R. (2009). Real and rational variants of the *h*-index and the *g*-index. *Journal of Informetrics*, *3*(1), 64–71.
Hagen, N. T. (2008). Harmonic allocation of authorship credit: Source-level correction of bibliometric bias assures accurate publication and citation analysis. *PLoS One*, *3*(12), e4021.
He, B., Ding, Y., & Yan, E. (2012). Mining patterns of author orders in scientific publications. *Journal of Informetrics*, *6*(3), 359–367.
Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, *102*(46), 16569–16572.
Hirsch, J. E. (2007). Does the *h* index have predictive power? *Proceedings of the National Academy of Sciences*, *104*(49), 19193–19198.
Hirsch, J. E. (2010). An index to quantify an individual's scientific research output that takes into account the effect of multiple coauthorship. *Scientometrics*, *85*(3), 741–754.
Hodge, S. E., & Greenberg, D. A. (1981). Publication credit. *Science*, *213*(4511), 950.
Hu, X., Rousseau, R., & Chen, J. (2010). In those fields where multiple authorship is the rule, the *h*-index should be supplemented by role-based *h*-indices. *Journal of Information Science*, *36*(1), 73–85.
Lee, J., Kraus, K. L., & Couldwell, W. T. (2009). Use of the *h* index in neurosurgery clinical article. *Journal of Neurosurgery*, *111*(2), 387–392.
Liu, X. Z., & Fang, H. (2012). Fairly sharing the credit of multi-authored papers and its application in the modification of *h*-index and *g*-index. *Scientometrics*, *91*(1), 37–49.
Lukovits, I., & Vinkler, P. (1995). Correct credit distribution: A model for sharing credit among coauthors. *Social Indicators Research*, *36*(1), 91–98.
Opthof, T., & Wilde, A. A. M. (2009). The Hirsch-index: A simple, new tool for the assessment of scientific output of individual scientists. *Netherlands Heart Journal*, *17*(4), 145–154.
Schreiber, M. (2008a). To share the fame in a fair way, h_m modifies *h* for multi-authored manuscripts. *New Journal of Physics*, *10*, 040201.
Schreiber, M. (2008b). A modification of the *h*-index: The h_m-index accounts for multi-authored manuscripts. *Journal of Informetrics*, *2*(3), 211–216.
Schreiber, M. (2008c). An empirical investigation of the *g*-index for 26 Physicists in comparison with the *h*-index, the *A*-index, and the *R*-index. *Journal of the American Society for Information Science and Technology*, *59*(9), 1513–1522.
Schreiber, M. (2009a). A case study of the modified Hirsch index h_m accounting for multiple coauthors. *Journal of the American Society for Information Science and Technology*, *60*(6), 1274–1282.
Schreiber, M. (2009b). The influence of self-citation corrections and the fractionalised counting of multi-authored manuscripts on the Hirsch index. *Annals of Physics (Berlin)*, *18*(9), 607–621.
Schreiber, M. (2009c). Fractionalized counting of publications for the g-index. *Journal of the American Society for Information Science and Technology*, *60*(10), 2145–2150.

Schreiber, M. (2010). How to modify the g-index for multi-authored manuscripts. *Journal of Informetrics*, *4*(1), 42–54.

Schreiber, M., Malesios, C. C., & Psarakis, S. (2012). Exploratory factor analysis for the Hirsch index, 17 h-type variants, and some traditional bibliometric indicators. *Journal of Informetrics*, *6*(3), 347–358.

Sekercioglu, C. H. (2008). Quantifying coauthor contributions. *Science*, *322*(5900), 371.

Trueba, F. J., & Guerrero, H. (2004). A robust formula to credit authors for their publications. *Scientometrics*, *60*(2), 181–204.

Van Hooydonk, G. (1997). Fractional counting of multiauthored publications: Consequences for the impact of authors. *Journal of the American Society for Information Science*, *48*(10), 944–945.

Zhang, C. T. (2009). A proposal for calculating weighted citations based on author rank. *EMBO Reports*, *10*(5), 416–417.