IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 4, 2020, accepted December 12, 2020, date of publication December 17, 2020, date of current version December 31, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3045435

Assessing the Scientific Impact of Individual Scholars With Multi-Scale H-Index

FENG MA[1] AND YUAN HUANG[2]
[1]College of Economics and Management, Anhui Agricultural University, Hefei 230036, China
[2]Campus Library, Hefei University of Technology, Hefei 230036, China

Corresponding author: Feng Ma (phoenixma@ahau.edu.cn)

This work was supported in part by the National Science Foundation of China under Contract 61702009, and in part by the Social Science Planning Program of Anhui Province under Contract AHSKQ2016D80.

ABSTRACT Quantitative evaluation on the individual's scientific impact is an important research topic in the academic evaluation fields. Following the pioneering milestone of h-index, many variants have been proposed and but still suffer some limitations. In this paper, we propose a novel multi-scale h-index (MH-index, or Ma-Huang index) to comprehensively measure the scientific impact of scholars based on the publication record. We investigate the index from the mathematical perspective of correlation function between citation count and paper number. We reveal that different definitions of the correlation function lead to different index schemes. By defining the correlation functions with gradually varied slope, we can obtain a series of index elements, each of which captures the citation record in different granularities. Our approach can be essentially interpreted as a generalized h-index in multiple scales, which is obtained by first generating a multi-scale representation of a scholar's paper citation record and then calculating the h-index score for each scale. Our approach can be extended to consider all publication citations and subtly emphasizes those highly cited publications. To validate the proposed indicators, an experimental study is conducted on 82 scholars in the scientometrics field and comparison is made with several existing indices. Our proposed MH-index is demonstrated to be more balanced and fine-grained for evaluating and comparing the scientific impact of individual scientists.

INDEX TERMS Multi-scale analysis, bibliometrics, information processing, mathematical model.

I. INTRODUCTION

In the scientific field, there is an important research problem, *i.e.*, how to quantify the cumulative scientific significance and broad impact of an individual scholar. A proper overall quantitative index will provide a useful indicator to evaluate and compare different individuals competing for the same resource, such as universities faculty application or advancement and grant award, when scientific achievement is a critical assessment criterion [1]. Generally, the scientific impact of an individual scholar is complex and related with many factors, such as the publication record, the citation record, the research project commitment, the involvement of editorial service, etc. To make it tractable and computationally feasible, only the publication and citation records of an individual are usually used to extract an overall indicator number for assessment. Following this paradigm, early criteria which are commonly used include total number of papers, total number of citations, average citations per paper, number of significant papers, etc. Although intuitive, those criteria suffer from the imbalance between productivity and paper importance or the involvement of arbitrary parameters.

To avoid the disadvantage of the above criteria, as a pioneering milestone, Hirsch [1] proposed the h-index to measures the scientific impact of an individual's research work by simultaneously considering the publication productivity and the article impact (citation count). The h-index is an easily computable indicator to characterize the scientific output of a researcher. The definition of h-index can also be extended to assess other subjects, such as groups of scholars [2], journals [3], [4], patents [5], countries [6] and web pages [7]. To date, the h-index has been taken as one of the most important indicators in both Google Scholar, Scopus and Web of Science. As a pioneering work, the h-index measures the scientific impact of an individual by counting the maximal number of core papers with citation times over a threshold equal to the core paper number.

The associate editor coordinating the review of this manuscript and approving it for publication was Chi-Hua Chen.

Although simple and useful, the h-index suffers a nontrivial limitation that no additional credit is given to those highly cited articles, whereas every citation counts on scientific evaluation [8]–[10] [11]–[15]. To address this weakness, many h-index variants are proposed, such as g-index [16], [17], AR-index [18], R-index [18], Tapered h-index [19], Hg-index [20], q-2 index [21], PR-index [22], ch-index [23], t-index [24], π-index [25], w-index [26], EM-index [27], *etc*. To give credit to excess citation count, some other index schemes are proposed, such as e-index [28], j-index [29], Hl-index [30], and EM-index [27].

Another limitation of h-index is that it relies on only those articles which have significant impact and high citation record, whereas every article contributes to research output and scientific impact of an individual. To consider the importance of all cited articles of an individual, several other indices are proposed, such as the multidimensional h-index [31], mock h-index [32], central interval index [33], two sided h-index [34], l-index [35]. Egghe and Rousseau proposed to weight the h-index by citation count [36]. To compare different index methods, comprehensive studies have been made by Bornmann et al [12], [13].

In addition to the above two main defects: insensitive to highly cited papers and low cited papers, there are several other common shortcomings shared by the h-index and other citation based indicators: dependence on the specific research field, sensitivity to self-citations, ignorance of multi-authorship, and sensitivity to the scientific age of a scientist or the age of publication [13]. Besides the above issues, Waltman and Van Eck [37], [38] argue that h-index behaves in a counterintuitive way and may lead to inconsistencies in the way in which scientists are ranked. Further, they discuss several indicators that can be used as an alternative to the h-index and make emphasis on highly cited publications. The definition of h-index is regarded to suffer some arbitrariness [38], [39], since the two involved and compared quantities, *i.e.*, publications and citations, are of different physical meaning [40]. The geometric relation of the Lorenz curve and the h-index is discussed, which provides a new geometric interpretation of the h-index [41].

Considering the difficulty to evaluate the scientific impact of an individual with a single indicator, some other researchers resort to the combination of two or multiple indicators [42]. Although such an idea seems interesting and promising, special concern shall be taken on the combination or aggregation scheme. Besides, a new criterion for the h-index is proposed in the form of the citation distribution among publications [43].

In this paper, we investigate the index from the mathematical perspective of correlation function between citation count and paper number. We reveal that different definitions of the correlation function lead to different existing index schemes. By defining the correlation functions with gradually varied slope, we can obtain a series of index elements, each of which captures the citation record in different granularities. Our approach can be essentially interpreted as a generalized h-index in multiple scales, which is obtained by first generating a multi-scale representation of a scholar's paper citation record and then calculating the h-index score for each scale. In this way, a multi-dimensional h-index score vector is obtained. To obtain a single number for quantitative evaluation and comparison, a final index score is calculated by summarizing those multi-scale h-index scores in an adaptive weighing manner. Our approach can be extended to consider all publication citations and subtly emphasizes those important papers with high citation count. The proposed multi-scale h-index is more comprehensive, which not only considers those highly cited articles, but also counts those outside the h-core article set.

The rest of this paper is organized as follows. In Section II, we make a survey on some preliminary work including h-index, e-index, multidimensional h-index, iteratively weighted h-index, and EM-index, and discuss their advantage together with limitations. In Section III, we elaborate our novel multi-scale h-index and its properties. After that, we conduct experimental study in Section IV. Last but not least, we conclude this work in Section V.

II. RELATED WORK

This work is focused on the quantitative indicator of scientific impact based on an individual's record of article publication and citation. In the following, several related indices are surveyed and analyzed.

As the milestone index to measure an individual's scientific impact, h-index [1] is defined based on the citation of an individual's publication. The h-index of a scholar is h if h of his or her articles have at least h citations each while the citation of each of the rest papers is no larger than h. Geometrically, given the citation curve of all papers of an individual scholar in descending order, its intersection with the 45 degree line $y = x$ gives h, as illustrated in Fig. 1(a).

The h-index only considers those high-cited articles but ignores those with less-cited ones. To address this problem, the multi-dimensional h-index is proposed. Given the citation record of a scientist, his or her multi-dimensional h-index $(h_1, h_2, h_3, \cdots, h_k)$ is generated as follows. All the involved N articles are sorted by the citation count in descending order. Then, h_1 is obtained as the traditional h-index. After that, those top h_1 most-cited articles are removed and we compute h_2 as the h-index of the remaining $(N - h_1)$ articles. Similarity, h_{k+1} is obtained by computing the h-index on the remaining $(N - h_1 - h_2 - \cdots - h_k)$ most cited articles, which are obtained by checking the intersection of the line $y = x - \sum_{i=1}^{k} h_i$ with the article citation curve, as illustrated in Fig. 1 (b). Such process is iterated until all articles with citation count no less than 1 are checked. Such idea is also taken in the CSS (Characteristic Scores and Scales) based index scheme [44]. Since multi-dimensional h-index takes a vector form and is inconvenient for comparison, a global scalar value is defined

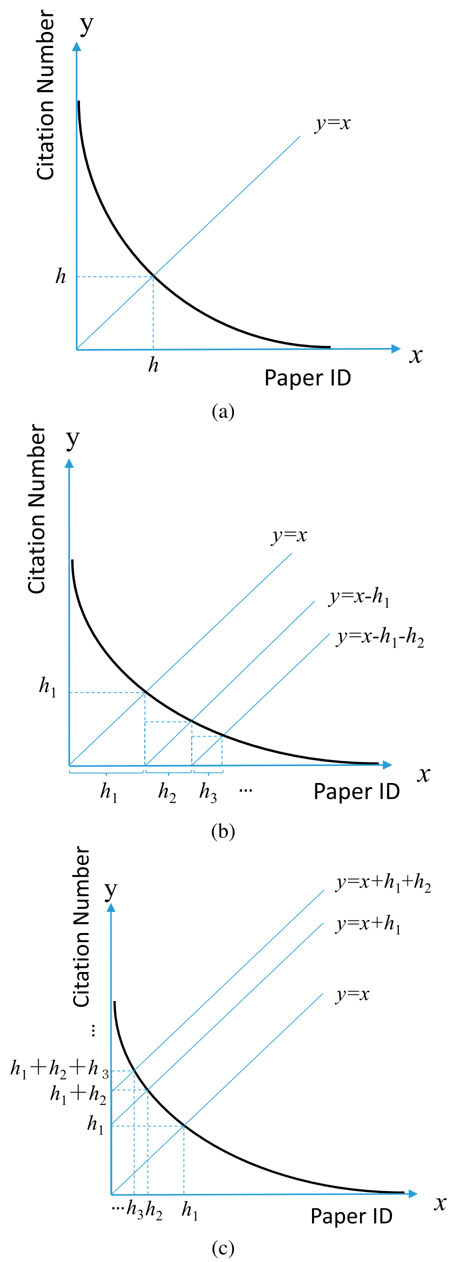


FIGURE 1. **Schematic illustration of (a) h-index, (b) multi-dimensional h-index, and (c) EM-index. The e-index is defined based on h-index with excess citation count considered, while the iteratively weighted h-index is a global representation of the multi-dimensional h-index.**

by the iteratively weighted h-index [45] as follows,

$$\mathrm{iw}(h) = \sum_{k=1}^{K} \frac{h_k}{k}. \tag{1}$$

To discriminate those highly-cited articles, some variants of h-index are proposed. Zhang [28] proposed e-index to consider the importance of the excess citation count over h obtained by h-index. Specifically, given the sorted article citations $(c_1, c_2, \cdots)$ in descending order of a scientist, the e-index is defined as squared root of the sum over the excess citation count of h-core articles as follows,

$$c = \sqrt{\sum_{k=1}^{h}(c_k - h)}. \tag{2}$$

The e-index suffers an issue that once the citation counts of those top h articles are all equal to h, the obtained e-index is equal to 0, which is unreasonable.

Another recent variant is EM-index [27], which adopts the concept of multi-dimensional h-index. Concretely speaking, for a scholar with article citations $(c_1, c_2, \cdots)$ ranked in descending order, EM-index first generates a multi-dimensional index vector $(h_1, h_2, h_3, \cdots, h_k)$. The first component h_1 is exactly the h-index value. The $(K+1)$-th component h_{k+1} is obtained by computing the h-index for the excess citation counts of the h-core papers in the k-th round, which is obtained by checking the intersection of the line $y = x + \sum_{i=1}^{k} h_i$ with the article citation curve, as illustrated in Fig. 1 (c). Finally, the EM-index is defined as the squared root of the sum over the all components in the index vector.

The above index schemes only consider the highly cited articles without proper weighting. Besides, they fail to explore all articles with at least one citation. To make a comprehensive evaluation, both the top and the tail citation counts should be considered in an adaptive weighting manner. Towards this goal, [34] proposed a novel two-sided h-index. Bihari and Tripathi [27] further extended their EM-index to EM'-index by considering the citation counts of all articles after subtracting the h-index from the h-core articles.

III. OUR APPROACH: MULTI-SCALE H-INDEX

To evaluate an individual's scientific impact based on the publication and citation records, the general principle beneath the existing schemes, such as h-index, e-index, iterative weighted h-index and EM-index, is to correlate the publication count with the citation count under some constraints. Following such a paradigm, an intuitive idea is to define a measure as: the index value of an individual is y, if at most the y articles have citation counts over x while the citation counts of the remaining articles are all no larger than x. In this perspective, a function $f(\cdot)$ is defined to correlate y and x as follows,

$$y = f(x). \tag{3}$$

With different definitions of the correlation function, we can obtain different index schemes, as discussed in the above section. In h-index, the correlation function is an identity function $f(x) = x$, which is inherited by many other variants. Although such a function is simple and demonstrates effective in many cases, it is not flexible to well capture the fine-grained characteristics of the citation curve. Based on such a motivation, we propose to define the correlation function as follows,

$$y = f_k(x) = 2^k x, \tag{4}$$

where k is an integer and 2^k indicates the slope. It is notable that, when k is equal to 0, the above correlation function degenerates to that of the classic h-index. By checking the intersection point of the line function $y = f_k(x)$ with the citation count curve, we can obtain the corresponding index component h_k. Through varying the value of k, a multi-dimensional index vector $(\cdots, h_{-2}, h_{-1}, h_0, h_1, h_2, \cdots)$ is generated, as illustrated in Fig. 2. Each component $h_k > 0$ and h_0 is equal to the classic h-index value. With a larger k, the corresponding component h_k considers articles with much higher citations.

Another perspective to interpret our scheme is the multi-scale representation of the descending citation curve. In the above discussion, we fix the citation curve $y = g(x)$ but change the correlation functions with varying values of k. Alternatively, we can also fix all the correlation function $y = f_k(x)$ as the identity function $f(x) = x$, but transform the citation curve $y = g(x)$ into multi-scale representations: $y = g_k(x) = 2^{-k}g(x)$. As a result, in the k-th scale, after the transformation of the citation curve, we can calculate the component h_k as the classic h-index on the transformed citation curve. In this sense, we call our scheme as *multi-scale h-index*. Our idea essentially follows the spirit of the scale space theory [46], which has been widely explored in the machine vision community of IEEE [47]–[49].

We define two versions of multi-scale h-index. As the basic multi-scale h-index (MH-index) scheme, we impose that k is non-negative, as illustrated in Fig. 2(a). Based on that, we further extend it by keeping the full component vector with all possible k and call it the extended multi-scale h-index scheme, which is abbreviated as MH(ext)-index, as illustrated in Fig. 2(b).

The above generated index vector is not convenient for evaluation and comparison. To summarize a global number from the multi-dimensional index vector, the multi-scale h-index score is defined as follows,

$$\mathrm{MH} = \sqrt{\sum_k w_k h_k}, \tag{5}$$

where w_k is a weighting scalar to discriminate the component h_k in different scales. Considering the above multi-scale formulation, w_k is defined as α^k. The selection of value range for α is based on two considerations. Firstly, α shall be larger than 1, such that more credit should be assigned to more highly cited article. Secondly, it is unreasonable to set α as a very large value, since it will introduce undue weight to those highly cited article. To this end, it is suggested $\alpha < 2$. In our experiments, we set $\alpha = \sqrt{2}$. It is notable that the weighting parameter α^k for $y_k(k < 0)$ is decreasing exponentially and the value of $y_k(k < 0)$ is also smaller than $y_k(k > 0)$. Therefore, the citation numbers below the conventional h-index threshold level only make a very small contribution to four final extended multi-scale h-index value. In other words, the extended multi-scale h-index value is dominated by the number of citations on and above the conventional h-index threshold level (i.e., $h_0, h_1, h_2, \cdots$).

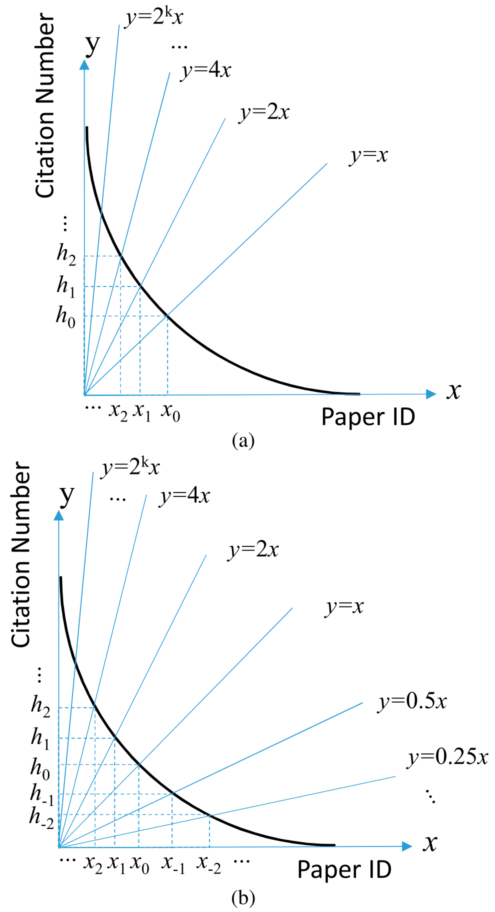


FIGURE 2. **Schematic illustration of our proposed (a) multi-scale h-index and (b) the extended multi-scale h-index.**

To illustrate the advantage of our proposed multi-scale h-index, we conduct comparison with h-index, iterative-weighted h-index (iw(h)-index) and EM-index on four synthetic examples in Table 1. Intuitively, the author with citation count A has less scientific impact than the author with B. However, both h-index and iterative-weighted h-index get the same index values for both authors and fail to discriminate them. In contrast, our multi-scale h-index distinguish A from B with notable difference. On the other hand, it is obvious that the scientific impact of A is less than that of C but greater than that of D. However, the EM-index assigns them with identical index values while our multi-scale h-index clearly differentiate them with the correct impact rank.

IV. EXPERIMENTS AND DISCUSSION

To comprehensively evaluate our proposed multi-scale h-index scheme, following [27], we collect the citation data of 82 scholars in the scientometrics and bibliometrics fields from Google Scholar. Those scholars are in the same research community, such that there is no domain gap in their publication and citation records for assessment and comparison. The total citation counts of those scholars range from 390 to

TABLE 1. Comparison of different index schemes on synthetic citation counts of scientists. To show the calculation details, let take the citation record A as an example. For iw(h)-index, we first get a 2D index sequence (4, 1) as illustrated in Fig. 1(b), then calculate $iw(h) = \frac{4}{1} + \frac{1}{2} = 4.5$. For EM-index, we first get a 5D index sequence (4, 2, 2, 1, 1) as illustrated in Fig. 1(c), then compute $EM(h) = \sqrt{4+2+2+1+1} = 3.16$. For MH-index, we first get a 4D index sequence (4, 2, 2, 1) as illustrated in Fig. 2(a) together with the corresponding weighting factor ($\sqrt{2}^0, \sqrt{2}^1, \sqrt{2}^2, \sqrt{2}^3$), then calculate MH(h) $=\sqrt{4 \cdot \sqrt{2}^0 + 2 \cdot \sqrt{2}^1 + 2 \cdot \sqrt{2}^2 + 1 \cdot \sqrt{2}^3} = 3.70$. The index value of MH(ext)-index can be obtained similarly.

Citation Counts	h-index	iw(h)-index	EM-index	MH-index	MH(ext)-index
A = (10, 8, 5, 5, 4, 0, 0)	4	4.50	3.16	3.70	5.07
B = (100, 8, 5, 5, 4, 0, 0)	4	4.50	10	5.60	6.59
C = (10, 10, 10, 10, 10, 10, 10, 10)	8	8	3.16	4.68	5.25
D = (10, 2, 1, 1, 1, 1, 0)	2	3.28	3.16	2.87	4.35

128,909, while the publication article number varies from 27 to 598. This dataset is diverse, which contains scholars with both high productivity and high citation count, scholars with high productivity and less citation count, and scholars with average productivity and average citation count. The total publication and citation together with the h-index of the 82 authors are illustrated in Fig. 3.

We compare our proposed MH-index with h-index, iterative-weighted h-index, EM-index, e-index, and raw total citation counts. The results of each index scheme on the 82 scholars are listed in Table 2. In Table 3, we show the rank of the authors based on the index scores calculated with each method. Our proposed MH-index provides a more fine-grained representation on the publication and citation records. Distinguished from h-index, multi-dimensional h-index and iteratively weighted h-index, our approach gives credit to those highly cited articles. On the other hand, our MH-index is different from EM-index in that, we calculate the significance index factor in a multi-scale paradigm and assigns more weight to the significance factor from articles of larger citations. Moreover, compared with e-index and raw total citation counts which both suffer the risk of overemphasizing the extremely highly cited articles, our MH-index adaptively weights those articles with moderate scale. The above properties make our MH-index produce more consistent and better ranking results on the scholar's scientific impact. More case study results are discussed below.

For instance, for the scholars Linton C. Freeman and Stan Wasserman, their h-index values are 49 and 45, respectively, while their iw(h)-index values are 66.47 and 61.41, respectively. As a result, they are ranked to no. 23 and no. 29 by h-index, and no. 33 and no. 38 by iw(h)-index. In fact, compared with the top-ranking authors, the publication article number of Linton C. Freeman (109 articles) and Stan Wasserman (106 articles) is relatively few. However, both of them have quite a few very highly cited articles with remarkable total citation count (35424 and 39101). Their top 1 citation counts are 27,680 and 11,329, respectively. Those extremely high citations are ignored by h-index and iteratively weighted h-index. In contrast, our MH-index makes a comprehensive consideration of this fact and rank them to no. 2 and no. 4, respectively. Similar results are observed by e-index and raw total citation counts. In other words, our MH-index scheme

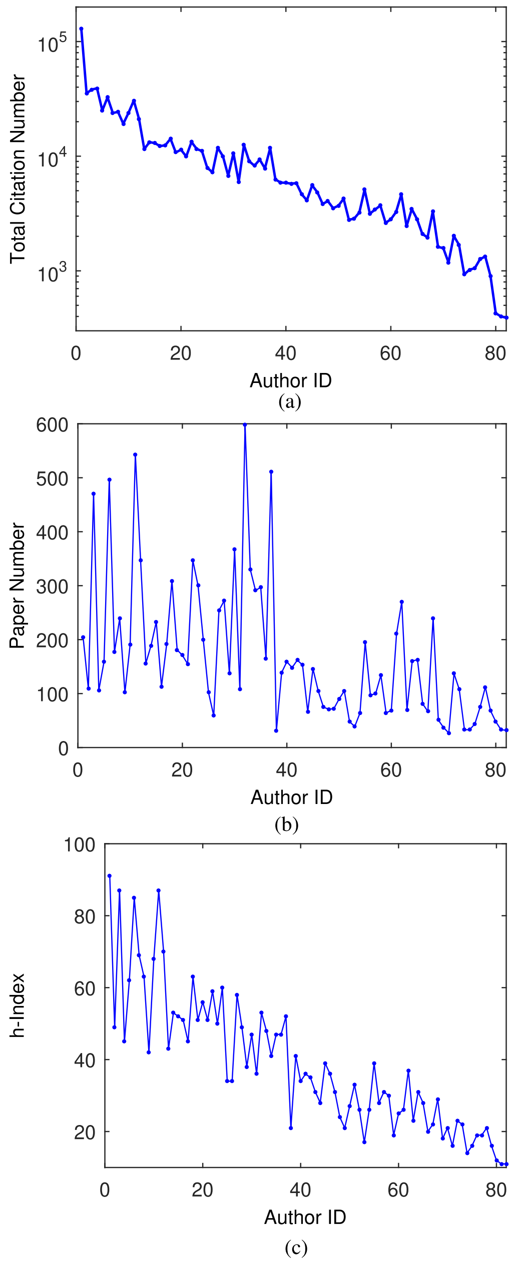


FIGURE 3. Illustration of (a) the total citation number, (b) paper number and (c) h-index of the studied 82 authors. The authors are sorted based on their MH-index scores in descending order. The data collection date from Google Scholar is April 26, 2017.

TABLE 2. Comparison of different index schemes on synthetic citation counts of scientists. In the table, "raw" means the naive index scheme of simple total citation count.

Author Name	h-index	iw(h)-index	EM-index	e-index	raw	**MH-index**	**MH(ext)-index**
Mark Newman	91	125.16	125.10	342.79	128909	44.97	48.42
Loet Leydesdorff	87	154.57	77.96	148.93	38143	29.96	37.19
C. Lee Giles	85	148.48	49.03	132.44	32604	28.17	35.65
Richard S.J. Tol	87	158.20	32.71	116.68	30572	26.44	35.01
Linton C. Freeman	49	66.47	106.39	179.29	35424	30.60	33.18
Hirsch J.E.	63	104.98	80.99	129.69	24312	27.00	32.16
Stan Wasserman	45	61.41	166.32	190.52	39101	29.47	32.01
Steve Lawrence	62	84.62	49.08	141.13	25025	28.41	31.99
Matthew O. Jackson	69	97.77	59.69	130.07	23845	27.54	31.93
Adamantios Diamantopoulos	68	98.70	63.46	130.37	23907	26.84	31.40
Enrique Herrera-Viedma	70	110.00	36.81	114.13	21142	25.72	31.38
Santo Fortunato	42	61.48	73.16	128.12	19170	26.96	29.80
Wolfgang Gla¨nzel	63	107.45	24.74	79.49	14189	21.92	28.49
Claes Wohlin	52	89.23	64.13	86.57	12996	22.66	28.02
Gerhard Woeginger	59	107.55	41.06	73.03	13299	20.69	27.96
Albert Zomaya	53	108.90	20.83	62.25	12533	18.95	27.87
Michael Jennions	53	85.64	39.28	90.39	13276	22.70	27.62
Ben R Martin	43	69.83	46.55	90.37	11529	22.85	26.89
Yannis Manolopoulos	50	91.17	45.14	76.72	11576	20.62	26.87
Guang-Hong Yang	52	108.31	22.47	57.39	11764	17.95	26.78
Roberto Todeschini	45	75.68	58.61	92.36	12441	21.96	26.61
Andras Schubert	51	78.85	49.28	80.36	10789	21.86	26.43
Mauno Vihinen	58	101.55	21.70	68.53	11850	19.59	26.39
Anthony F.J. van Raan	56	85.77	25.98	78.59	11381	21.30	26.32
Ronald Rousseau	47	93.46	30.46	66.78	10604	19.03	26.22
Andrew D. Jackson	60	93.24	28.02	70.96	11070	20.56	26.21
Anne-Wil Harzing	51	70.77	24.74	92.64	12260	22.49	26.13
Blaise Cronin	49	88.30	30.36	68.09	9983	19.22	25.60
Josep Domingo-Ferrer	48	90.11	23.17	59.62	9053	18.75	25.45
Lutz Bornmann	47	89.20	25.90	61.92	9347	18.62	25.39
Henk F. Moed	51	75.52	36.70	76.84	9940	20.76	25.24
Vicenç Torra	41	76.11	29.27	65.38	8248	18.74	24.63
Mark Fine	47	78.05	27.77	59.62	7830	18.51	23.91
Johan Bollen	34	51.97	50.20	77.78	7936	20.48	23.65
Herbert Van de Sompel	38	59.37	25.53	65.37	6757	19.08	23.10
Judit Bar-Ilan	41	63.98	23.19	55.08	5891	17.87	22.38
Michael S. Rosenberg	34	42.86	34.45	76.30	7227	20.05	22.29
Cagan H. Sekercioglu	36	53.16	23.45	63.52	5966	18.96	22.29
Marek Kosmulski	34	58.78	24.64	58.72	5853	17.78	22.26
Daniel HD	36	61.77	25.90	56.38	5769	17.28	21.92
Serge Galam	35	60.57	19.13	57.58	5817	17.25	21.88
Weiguo (Patrick) Fan	39	64.89	18.68	51.81	5569	16.81	21.66
Kène Henkens	39	67.51	14.97	44.68	5112	15.56	21.10
Roger Brumback	37	70.53	12.88	34.97	4640	14.10	20.94
Jayant Vaidya	31	51.50	23.79	53.84	4660	17.01	20.94
Benny Lautrup	36	54.41	21.38	52.23	4842	16.78	20.67
Aric Hagberg	28	41.21	33.56	53.21	4121	16.85	19.87
Maria Bordons	33	51.97	18.17	48.33	4256	15.93	19.87
Nees Jan van Eck	31	45.14	23.56	48.11	3835	16.56	19.84
Christoph Bartneck	30	51.62	21.05	43.58	3711	14.89	19.41
Heidi Winklhofer	21	25.25	63.46	75.70	6270	17.92	19.37
Dimitrios Katsaros	27	42.19	27.91	49.42	3692	15.95	19.20
Carlos Pecharroman	31	49.88	17.46	40.93	3422	15.08	19.11
Fiorenzo Franceschini	31	54.87	15.10	37.40	3458	13.89	19.00
Sergio Alonso	24	34.10	19.39	57.05	4054	16.49	18.92
Peter Jacso	26	48.94	20.83	39.91	3258	14.17	18.89
Berwin Turlach	21	35.35	27.13	51.54	3523	16.05	18.87
Gangan Prathap	29	56.53	14.46	32.28	3306	12.95	18.85
Paul Wouters	28	44.28	17.03	42.25	3155	15.14	18.69
John Irvine	26	37.26	25.59	47.04	3233	15.65	18.34
Hendrik P. van Dalen	28	50.22	12.37	33.21	2824	13.29	18.15
Ludo Waltman	26	34.32	23.56	43.58	2793	15.73	18.08
Sune Lehmann	17	24.20	35.11	49.18	2848	15.66	17.46
Duncan Lindsey	25	36.50	21.79	43.31	2813	14.41	17.37
Ruediger Mutz	23	35.50	19.52	39.71	2452	13.95	17.00
Francisco Javier Cabrerizo	19	29.72	19.39	44.77	2622	14.57	16.97
Jörn Altmann	23	43.54	13.78	26.59	2018	11.55	16.53
Morten Schmidt	20	33.28	18.60	36.54	2099	13.24	16.40
Rodrigo Costas	22	33.54	17.46	34.25	1956	13.24	16.25
Miguel A Garcıa-Pérez	22	41.40	14.73	21.86	1686	10.80	15.56
Clint D. Kelly	18	27.85	17.35	32.42	1618	12.72	15.27
Raj Kumar Pan	21	27.45	19.49	31.16	1577	12.53	14.93
Birger Larsen	21	38.92	8.12	16.76	1336	9.76	14.57
Domenico A. Maisano	19	32.85	15.10	22.83	1263	10.36	14.21
Yu-Hsin Liu	16	20.45	17.55	29.10	1182	12.09	13.86
Luca Mastrogiacomo	16	28.48	9.75	18.03	901	9.50	13.24
Andras Telcs	19	26.47	10.63	23.26	1054	10.44	13.12
Alireza Abbasi	14	20.03	13.78	25.08	931	10.76	12.88
Nils T. Hagen	16	21.97	10.72	25.36	1023	10.46	12.81
Ash Mohammad Abbas	12	20.49	8.60	11.75	427	7.97	10.83
Raf Guns	11	17.72	8.54	13.08	400	7.90	10.44
Fred Y. Ye	11	16.89	7.28	13.71	390	7.70	9.95

TABLE 3. Comparative result of the author rank based on h-index, iteratively weighted h-index, EM-index, e-index, raw (total citation count), MH-index and the extended MH-index.

Author Name	h-index	iw(h) index	EM-index	e-index	raw	**MH-index**	**MH(ext)-index**
Mark Newman	1	4	2	1	1	1	1
Loet Leydesdorff	2	2	5	4	3	3	2
C. Lee Giles	4	3	15	6	5	6	3
Richard S.J. Tol	3	1	25	11	6	11	4
Linton C. Freeman	23	33	3	3	4	2	5
Hirsch J.E.	8	10	4	9	8	8	6
Stan Wasserman	29	38	1	2	2	4	7
Steve Lawrence	10	23	14	5	7	5	8
Matthew O. Jackson	6	13	10	8	10	7	9
Adamantios Diamantopoulos	7	12	8	7	9	10	10
Enrique Herrera-Viedma	5	5	20	12	11	12	11
Santo Fortunato	32	37	6	10	12	9	12
Wolfgang Gla¨nzel	9	9	38	19	13	18	13
Claes Wohlin	17	18	7	17	16	15	14
Gerhard Woeginger	12	8	18	26	14	22	15
Albert Zomaya	15	6	52	34	17	32	16
Michael Jennions	16	22	19	15	15	14	17
Ben R Martin	31	31	16	16	23	13	18
Yannis Manolopoulos	22	16	17	23	22	23	19
Guang-Hong Yang	18	7	47	40	21	37	20
Roberto Todeschini	30	27	11	14	18	17	21
Andras Schubert	19	24	13	18	26	19	22
Mauno Vihinen	13	11	49	28	20	27	23
Anthony F.J. van Raan	14	21	33	20	24	20	24
Ronald Rousseau	26	14	26	30	27	30	25
Andrew D. Jackson	11	15	29	27	25	24	26
Anne-Wil Harzing	20	29	39	13	19	16	27
Blaise Cronin	24	20	27	29	28	28	28
Josep Domingo-Ferrer	25	17	46	36	31	33	29
Lutz Bornmann	27	19	34	35	30	35	30
Henk F. Moed	21	28	21	22	29	21	31
Vicenç Torra	33	26	28	31	32	34	32
Mark Fine	28	25	31	37	34	36	33
Johan Bollen	43	46	12	21	33	25	34
Herbert Van de Sompel	37	40	37	32	36	29	35
Judit Bar-Ilan	34	35	45	43	39	39	36
Michael S. Rosenberg	44	56	23	24	35	26	37
Cagan H. Sekercioglu	39	45	44	33	38	31	38
Marek Kosmulski	45	41	40	38	40	40	39
Daniel HD	40	36	35	42	42	41	40
Serge Galam	42	39	58	39	41	42	41
Weiguo (Patrick) Fan	35	34	59	47	43	45	42
Kène Henkens	36	32	69	55	44	55	43
Roger Brumback	38	30	74	65	47	62	44
Jayant Vaidya	47	49	41	44	46	43	45
Benny Lautrup	41	44	50	46	45	46	46
Aric Hagberg	53	59	24	45	49	44	47
Maria Bordons	46	47	61	51	48	51	48
Nees Jan van Eck	48	53	42	52	51	47	49
Christoph Bartneck	51	48	51	56	52	58	50
Heidi Winklhofer	66	75	9	25	37	38	51
Dimitrios Katsaros	56	57	30	49	53	50	52
Carlos Pecharroman	49	51	63	60	56	57	53
Fiorenzo Franceschini	50	43	67	63	55	64	54
Sergio Alonso	61	66	56	41	50	48	55
Peter Jacso	57	52	53	61	58	61	56
Berwin Turlach	67	64	32	48	54	49	57
Gangan Prathap	52	42	71	69	57	68	58
Paul Wouters	54	54	66	59	60	56	59
John Irvine	58	61	36	53	59	54	60
Hendrik P. van Dalen	55	50	75	67	62	65	61
Ludo Waltman	59	65	43	57	64	52	62
Sune Lehmann	75	76	22	50	61	53	63
Duncan Lindsey	60	62	48	58	63	60	64
Ruediger Mutz	62	63	54	62	66	63	65
Francisco Javier Cabrerizo	71	70	57	54	65	59	66
Jörn Altmann	63	55	72	72	68	72	67
Morten Schmidt	70	68	60	64	67	66	68
Rodrigo Costas	64	67	64	66	69	67	69
Miguel A Garcıa-Pérez	65	58	70	77	70	73	70
Clint D. Kelly	74	72	65	68	71	69	71
Raj Kumar Pan	68	73	55	70	72	70	72
Birger Larsen	69	60	81	79	73	78	73
Domenico A. Maisano	72	69	68	76	74	77	74
Yu-Hsin Liu	76	79	62	71	75	71	75
Luca Mastrogiacomo	77	71	78	78	79	79	76
Andras Telcs	73	74	77	75	76	76	77
Alireza Abbasi	79	80	73	74	78	74	78
Nils T. Hagen	78	77	76	73	77	75	79
Ash Mohammad Abbas	80	78	79	82	80	80	80
Raf Guns	81	81	80	81	81	81	81
Fred Y. Ye	82	82	82	80	82	82	82

pays special attention and make more emphasize on those highly cited publication.

Another interesting example is Heidi Winklhofer, whose h-index is 22 but EM-index is 63.46. As a result, the scholar is ranked to as high as no. 9 by EM-index, which is unreasonable considering the relatively less publication article number (31) and total citation (6270). That is due to the fact that EM-index extremely favors the case like Heidi Winklhofer, whose top 1 citation is 4027 but the following one takes a cliff-like drop to 433. Our MH-index well address such cases and rank Heidi Winklhofer to no. 38.

In Fig. 4, we study four scholars, i.e., Santo Fortunato, Ben R Martin, Vicenc Torra and Judit Bar-Ilan, with h-index values of 42, 43, 41 and 41, respectively, which are very close. However, their citation curves are dramatically different. Our MH-index well discriminates them with multi-dimensional impact factor in multi-scale representations. The component level analysis on the multi-dimensional index vector for the four scholars is illustrated in Fig. 4. The component numbers of Santo Fortunato and Ben R Martin are 13 and 12, respectively, which indicates their top-1 citations are above $2^{12} = 4096$ and $2^{11} = 2014$, respectively. Comparing Vicenc Torra and Judit Bar-Ilan, it is observed from the component curves that the former has more highly-cited publication articles while the latter publishes more articles with citation between the average and the top ones. As a result, the scientific impact of Vicenc Torra and Judit Bar-Ilan shall be similar. Finally, our MH-index assigns the index values of the above four scholars as 26.96, 22.85, 18.74, and 17.87, which well reflects their citation curve property and is consistent with the general expectation on scientific impact.

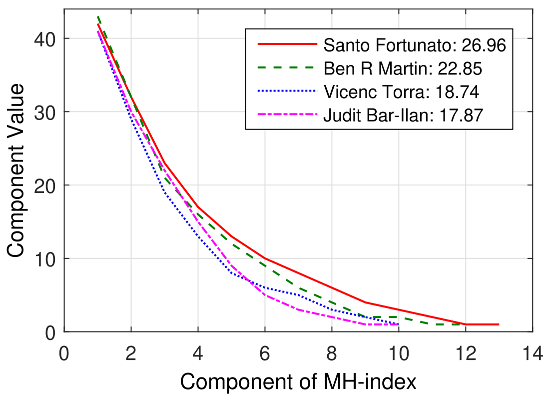


FIGURE 4. **Component level analysis of MH-index for four scholars with similar h-index. The h-index value of each author is indicated by the first element of the MH-index component vector.**

In Fig. 5, the publication citation curves of two scholars, Linton C. Freeman and Loet Leydesdorff, are illustrated, together with their index scores by seven different schemes. It is observed that the top-8 article citation counts of Linton C. Freeman is notably above that of Loet Leydesdorff. But for the remaining articles, the latter has better citation record. The h-index value of the latter is 87, which is significantly larger than the former whose h-index is only 49, which is

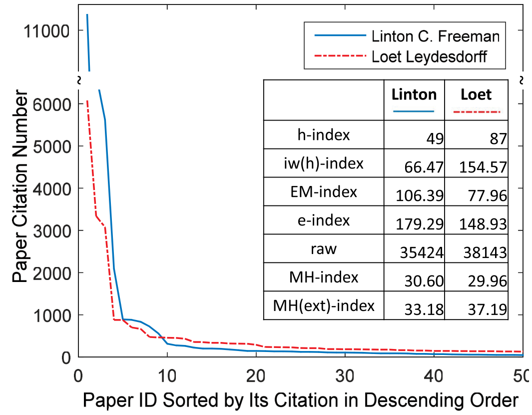


FIGURE 5. **Publication citation curves of two scholars: Linton C. Freeman and Loet Leydesdorff, and the results by different index schemes. In the figure, "raw" means the naive index scheme of simple total citation count.**

due to the ignorance of those excess citation count of those highly-cited articles. On the other hand, EM-index runs to another extreme by giving a significantly larger index value 106.39 to Linton C. Freeman, comparing to the EM-index value of 77.96 for Loet Leydesdorff, thanks to the sensitivity of EM-index to the cliff-like drop in the top citation counts. Our MH-index is free of the above disadvantages and assigns very close index values (30.60 vs. 29.96) to the two referred scholars. Fig. 6 shows another example showing similar dilemma of the comparison index methods and the fine-grained result of our MH-index.

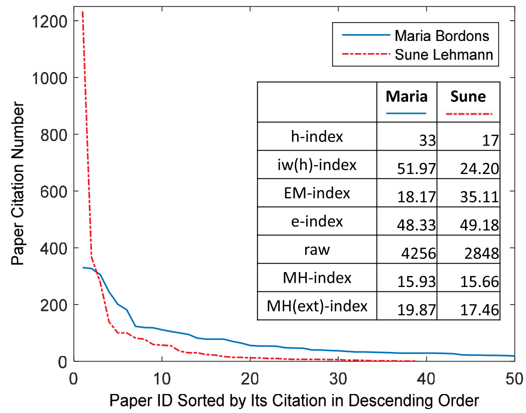


FIGURE 6. **Publication citation curves of two scholars: Maria Bordons and Sune Lehmann, and the results by different index schemes. In the figure, "raw" means the naive index scheme of simple total citation count.**

In Fig. 7, we further compare the seven index schemes based on the publication citation records of Heidi Winklhofer and Daniel HD. Heidi Winklhofer has 501 more total citation count than Daniel HD (6270 vs. 5769), which is dominantly attributed to the fact the highest citation of Heidi Winklhofer is much larger than that of Daniel HD (4024 vs. 671). However, besides that, Daniel HD has similar citation in the second publication and consistently more citations than Heidi Winklhofer in the remaining ones,

TABLE 4. Spearman rank correlation of different indices.

	h-index	iw(h)-index	EM-index	e-index	raw	**MH-index**	**MH(ext)-index**
h-index	1	0.97	0.71	0.88	0.95	0.90	0.96
iw(h)-index	0.97	1	0.61	0.78	0.90	0.82	0.92
EM-index	0.71	0.61	1	0.91	0.84	0.90	0.82
e-index	0.88	0.78	0.91	1.00	0.97	0.99	0.95
raw	0.95	0.90	0.84	0.97	1.00	0.97	0.99
MH-index	0.90	0.82	0.90	0.99	0.97	1	0.97
MH(ext)-index	0.96	0.92	0.82	0.95	0.99	0.97	1

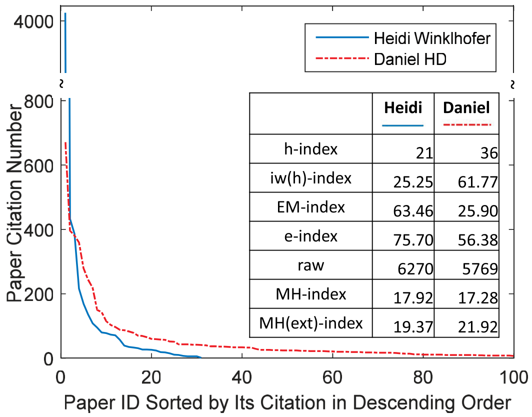


FIGURE 7. Publication citation curves of two scholars: Heidi Winklhofer and Daniel HD, and the results by different index schemes. In the figure, "raw" means the naive index scheme of simple total citation count.

demonstrated by the dashed red curve over the solid blue one with a notable gap. Besides, Heidi Winklhofer has much less publications than Daniel HD (31 vs. 148). Based on the above fact, the index by raw total citation count favors Heidi Winklhofer, which is followed by e-index and EM-index. In contrast, a distinct opposite result is suggested by h-index and iw(h)-index. Different from the above index schemes, our MH-index suggest similar assessment results for Heidi Winklhofer and Daniel HD (17.92 vs. 17.28). With the additional consideration of the citations below the conventional h-index threshold level, our MH(ext)-index favors Heidi Winklhofer less than Daniel HD (19.37 vs. 21.92), which is consistent with the judgement by common sense. Based on the above discussion, the key advantage of our index scheme over e-index and raw total citation count is that, we never risk to overemphasize the highly cited articles. With the subtly designed weighting scheme and the implicit consideration of publication productivity, our index method can achieve more comprehensive and reasonable results.

To measure the statistical dependency between different index methods, we compute the Spearman rank correlation on the ranking results among the studied indices. As shown in Table 4, our MH-index is highly correlated with e-index, raw (total citation count), h-index and EM-index, while our MH(ext)-index is more strongly correlated with raw (total citation count) and h-index. It indicates that our MH-index not only considers the overall significant impact as revealed by total citation count and h-index, but also captures the fine-grained importance of those highly-cited articles as revealed by EM-index. Although our MH-index and its extension are highly correlated with e-index and raw (total citation count), the key difference lies that our index schemes well handles the individual highly cited article without giving undue weight, as demonstrated in Fig. 7. Since the MH(ext)-index enriches the MH-index with those tailed publication articles, the contribution of those h-core publication articles are saturated to some extent, which is reflected by the fact that the correlation between MH(ext)-index and EM-index is low. In a nutshell, our proposed MH-index and MH (ext)-index schemes are more balanced and fine-grained for comprehensively evaluating and comparing the scientific impact of individual scientists.

V. CONCLUSION

In this paper, we propose a novel multi-scale h-index to measure the scientific impact of individual scholars. We mathematically formulate the index scheme from the perspective of correlation function to collaboratively investigate the publication and citation record. With a multi-scale paradigm, a multi-dimension index factor is generated, which is further summarized by adaptive weighting into a global index number. The proposed MH-index (or Ma-Huang index) makes a fine-grained measure and gives adaptive credit based on the significance level of the publication article citation count. We evaluate our index scheme with comprehensive study on the real scholars and quantitatively demonstrate the advantage of the proposed MH-index over the comparison indices.

REFERENCES

[1] J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 46, pp. 16569–16572, Nov. 2005.

[2] A. F. J. van Raan, "Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups," *Scientometrics*, vol. 67, no. 3, pp. 491–502, Jun. 2006.

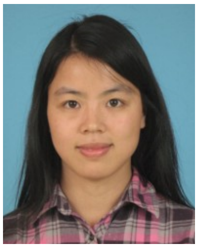
[3] T. Braun, W. Glänzel, and A. Schubert, "A Hirsch-type index for journals," *Scientometrics*, vol. 69, no. 1, pp. 169–173, Oct. 2006.

[4] J. Mingers, F. Macri, and D. Petrovici, "Using the h-index to measure the quality of journals in the field of business and management," *Inf. Process. Manage.*, vol. 48, no. 2, pp. 234–241, Mar. 2012.

[5] J. C. Guan and X. Gao, "Exploring the H-index at patent level," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 60, no. 1, pp. 35–40, Jan. 2009.

[6] E. Csajbók, A. Berhidi, L. Vasas, and A. Schubert, "Hirsch-index for countries based on essential science indicators data," *Scientometrics*, vol. 73, no. 1, pp. 91–117, Oct. 2007.

[7] J. Bar-Ilan and M. Levene, "The hw-rank: An h-index variant for ranking Web pages," *Scientometrics Int. J. Quant. Aspects Sci. Sci. Policy*, vol. 102, no. 3, pp. 2247–2253, Mar. 2015.
[8] J. K. Vanclay, "On the robustness of the H-index," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 58, no. 10, pp. 1547–1550, 2007.
[9] P. Vinkler, "Eminence of scientists in the light of theh-index and other scientometric indicators," *J. Inf. Sci.*, vol. 33, no. 4, pp. 481–491, Aug. 2007.
[10] R. Rousseau and F. Y. Ye, "A proposal for a dynamic H-type index," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 59, no. 11, pp. 1853–1855, Sep. 2008.
[11] M. Schreiber, "A modification of the H-index: The HM-index accounts for multi-authored manuscripts," *J. Informetrics*, vol. 2, no. 3, pp. 211–216, Jul. 2008.
[12] L. Bornmann, R. Mutz, and H.-D. Daniel, "Are there better indices for evaluation purposes than the H index? A comparison of nine different variants of the H index using data from biomedicine," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 59, no. 5, pp. 830–837, 2008.
[13] L. Bornmann, R. Mutz, S. E. Hug, and H.-D. Daniel, "A multilevel meta-analysis of studies reporting correlations between the H index and 37 different H index variants," *J. Informetrics*, vol. 5, no. 3, pp. 346–359, Jul. 2011.
[14] L. Egghe, "Mathematical study of H-index sequences," *Inf. Process. Manage.*, vol. 45, no. 2, pp. 288–297, Mar. 2009.
[15] M.-H. Huang and W.-Y.-C. Lin, "Probing the effect of author self-citations on H index: A case study of environmental engineering," *J. Inf. Sci.*, vol. 37, no. 5, pp. 453–461, Oct. 2011.
[16] L. Egghe, "An improvement of the H-index: The G-index," *ISSI Newslett.*, vol. 2, no. 1, pp. 8–9, 2006.
[17] L. Egghe, "An econometric property of the g-index," *Inf. Process. Manage.*, vol. 45, no. 4, pp. 484–489, Jul. 2009.
[18] B. Jin, L. Liang, R. Rousseau, and L. Egghe, "The R- and AR-indices: Complementing the h-index," *Chin. Sci. Bull.*, vol. 52, no. 6, pp. 855–863, Mar. 2007.
[19] T. R. Anderson, R. K. S. Hankin, and P. D. Killworth, "Beyond the durfee square: Enhancing the H-index to score total publication output," *Scientometrics*, vol. 76, no. 3, pp. 577–588, Sep. 2008.
[20] S. Alonso, F. J. Cabrerizo, E. Herrera-Viedma, and F. Herrera, "HG-index: A new index to characterize the scientific output of researchers based on the H- and G-indices," *Scientometrics*, vol. 82, no. 2, pp. 391–400, Feb. 2010.
[21] F. J. Cabrerizo, S. Alonso, E. Herrera-Viedma, and F. Herrera, "Q2-index: Quantitative and qualitative evaluation based on the number and impact of papers in the Hirsch core," *J. Informetrics*, vol. 4, no. 1, pp. 23–28, Jan. 2010.
[22] C. Gao, Z. Wang, X. Li, Z. Zhang, and W. Zeng, "PR-index: Using the H-index and PageRank for determining true impact," *PLoS ONE*, vol. 11, no. 9, Sep. 2016, Art. no. e0161755.
[23] M. Kosmulski, "A new Hirsch-type index saves time and works equally well as the original h-index," *ISSI Newslett.*, vol. 2, no. 3, pp. 4–6, 2006.
[24] R. S. J. Tol, "The H-index and its alternatives: An application to the 100 most prolific economists," *Scientometrics*, vol. 80, no. 2, pp. 317–324, Aug. 2009.
[25] P. Vinkler, "The π-index: A new indicator for assessing scientific impact," *J. Inf. Sci.*, vol. 35, no. 5, pp. 602–612, Oct. 2009.
[26] Q. Wu, "The W-index: A measure to assess scientific impact by focusing on widely cited papers," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 61, no. 3, pp. 609–614, 2010.
[27] A. Bihari and S. Tripathi, "EM-index: A new measure to evaluate the scientific impact of scientists," *Scientometrics*, vol. 112, no. 1, pp. 659–677, Jul. 2017.
[28] C.-T. Zhang, "The e-index, complementing the H-index for excess citations," *PLoS ONE*, vol. 4, no. 5, p. e5429, May 2009.
[29] R. Todeschini, "The j-index: A new bibliometric index and multivariate comparisons between other common indices," *Scientometrics*, vol. 87, no. 3, pp. 621–639, Jun. 2011.
[30] L. Zhai, X. Yan, and B. Zhu, "The h 1 -index: Improvement of H-index based on quality of citing papers," *Scientometrics*, vol. 98, no. 2, pp. 1021–1031, Feb. 2014.
[31] M. A. Garcia-Perez, "A multidimensional extension to Hirsch's H-index," *Scientometrics*, vol. 81, no. 3, pp. 779–785, 2009.
[32] G. Prathap, "Is there a place for a mock H-index?" *Scientometrics*, vol. 84, no. 1, pp. 153–165, Jul. 2010.
[33] P. Dorta-González and M.-I. Dorta-González, "Central indexes to the citation distribution: A complement to the h-index," *Scientometrics*, vol. 88, no. 3, pp. 729–745, Sep. 2011.
[34] M. A. García-Pérez, "An extension of the H index that covers the tail and the top of the citation curve and allows ranking researchers with similar H," *J. Informetrics*, vol. 6, no. 4, pp. 689–699, Oct. 2012.
[35] T. Lando and L. Bertoli-Barsotti, "A new bibliometric index based on the shape of the citation distribution," *PLoS ONE*, vol. 9, no. 12, Dec. 2014, Art. no. e115962.
[36] L. Egghe and R. Rousseau, "An h-index weighted by citation impact," *Inf. Process. Manage.*, vol. 44, no. 2, pp. 770–780, Mar. 2008.
[37] L. Waltman and N. J. van Eck, "A taxonomy of bibliometric performance indicators based on the property of consistency," RSM Erasmus Univ., Rotterdam, The Netherlands, Tech. Rep. ERS-2009-014-LIS, 2009.
[38] L. Waltman and N. J. Van Eck, "The inconsistency of the h-index," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 63, no. 2, pp. 406–415, 2012.
[39] G. Ellison, "How does the market use citation data? The Hirsch index in economics," *Amer. Econ. J., Appl. Econ.*, vol. 5, no. 3, pp. 63–90, Jul. 2013.
[40] S. Lehmann, A. D. Jackson, and B. E. Lautrup, "A quantitative analysis of indicators of scientific performance," *Scientometrics*, vol. 76, no. 2, pp. 369–390, Aug. 2008.
[41] L. Egghe and R. Rousseau, "A geometric relation between the h-index and the lorenz curve," *Scientometrics*, vol. 119, no. 2, pp. 1281–1284, May 2019.
[42] B. R. Martin, "The use of multiple indicators in the assessment of basic research," *Scientometrics*, vol. 36, no. 3, pp. 343–362, Jul. 1996.
[43] L. C. Brandão and J. C. C. B. Soares de Mello, "A multi-criteria approach to the H-index," *Eur. J. Oper. Res.*, vol. 276, no. 1, pp. 357–363, Jul. 2019.
[44] L. Egghe, "Characteristic scores and scales based on H-type indices," *J. Informetrics*, vol. 4, no. 1, pp. 14–22, Jan. 2010.
[45] R. Todeschini and A. Baccini, *Handbook of Bibliometric Indicators: Quantitative Tools for Studying and Evaluating Research*. Hoboken, NJ, USA: Wiley, 2016.
[46] T. Lindeberg, "Scale-space theory: A basic tool for analyzing structures at different scales," *J. Appl. Statist.*, vol. 21, nos. 1–2, pp. 225–270, Jan. 1994.
[47] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, 1999, pp. 1150–1157.
[48] V. Lepetit and P. Fua, "Keypoint recognition using randomized trees," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1465–1479, Sep. 2006.
[49] S. Li, Q. Hao, X. Kang, and J. A. Benediktsson, "Gaussian pyramid based multiscale feature fusion for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 9, pp. 3312–3324, Sep. 2018.



FENG MA was born in Hubei, China, in 1984. She received the B.S. degree in information management and information system from Central China Normal University, Wuhan, China, in June 2007, the M.S. degree in library and information science from the Institute of Science and Technology of Information of China, Beijing, China, in June 2009, and the Ph.D. degree in library and information science from Wuhan University, Wuhan, in June 2012.

From July 2012 to 2018, she was an Assistant Professor with the School of Economics and Management, Anhui Agricultural University, where she has been an Associate Professor, since 2018. She is the author of nine books and more than 30 articles. Her research interests include information management, academic impact evaluation, and statistical data analysis.

YUAN HUANG graduated from Beijing Forestry University, in 2006. She is currently a Librarian in library with the Hefei University of Technology. Her research interests include information retrieval, reference service, and academic impact analysis.