IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 13, 2019, accepted August 6, 2019, date of publication August 30, 2019, date of current version September 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2938657

Self-Citation Analysis on Google Scholar Dataset for H-Index Corrections

FIAZ MAJEED[1], MUHAMMAD SHAFIQ[2], AMJAD ALI[3], MUHAMMAD AWAIS HASSAN[4], SYED ALI ABBAS[5], MOHAMMAD EID ALZAHRANI[6], MUHAMMAD QAISER SALEEM[6], HANNAN BIN LIAQAT[1], AKBER GARDEZI[7], AND AZEEM IRSHAD[8]

[1]Department of Information Technology, University of Gujrat, Gujrat 50700, Pakistan
[2]Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, South Korea
[3]Department of Computer Science, COMSATS University Islamabad, Lahore Campus, Lahore 54590, Pakistan
[4]Department of Computer Science and Engineering, University of Engineering and Technology, Lahore 54920, Pakistan
[5]Department of Computer Science and Information Technology, University of Azad Jammu and Kashmir, Muzaffarabad 13100, Pakistan
[6]College of Computer Science and Information Technology, Al Baha University, Al Baha 1988, Saudi Arabia
[7]Department of Computer Science, COMSATS University Islamabad, Islamabad 45550, Pakistan
[8]Department of Computer Science and Software Engineering, International Islamic University, Islamabad 44000, Pakistan

Corresponding author: Muhammad Shafiq (shafiq.pu@gmail.com)

This work was supported by the National Research Foundation of Korea (NRF) through the Brain Korea 21 Plus Program under Grant 22A20130012814.

ABSTRACT For the recent decades, self-citations have been extensively studied by the academia, therein Web of Science (WoS) citations count and h-index are being considered as the benchmark parameters. The WoS is used to determine citations based on the Institute for Scientific Information's (ISI) master list. Towards this end, Google Scholar maintains a broad source of the research articles. However, Google Scholar does not exclude self-citations from the list of citations of one particular journal, author or co-author. The Google Scholar citation statistics are, therefore, not regarded as highly accurate. In this paper, we propose an updated h-index for Google Scholar by first quantifying and thereafter excluding the self-citations from the h-index. We target the two aspects of Google Scholar that belong to the evaluation of the quality of Google Scholar sources and the self-citation' records available in the citation lists. In our analysis, we have taken two datasets. The first dataset is composed of scientists awarded by the Scientometrics, which is recorded from the Google Scholar. According to this dataset, 28 scientists have been awarded as best researchers on the basis of their maximum citations and the h-indexes. The second dataset includes 16 non-award winner scientists. Both datasets include records falling in the period of 1984 to 2017. Based on analysis of award winner scientists' data, the aggregated journal self-citations are observed as 3.95%, whereas author and coauthor self-citations are found as 2.86% and 3.33%, respectively. In contrast, non-award winner scientists have average journal biased citation as 1.22%, author biased citations as 0.41% and co-author biased citations as 0.90%. We consider three types of the self-citations, i.e., journal, author, and coauthor for each scientist to cumulatively calculate the revised h-index. We obtained a new ranking of the scientists, which is based on a more accurate updated h-index. The updated h-index for the Google Scholar can be used for a more accurate academic ranking of the authors and the research articles.

INDEX TERMS Bibliometrics, Google Scholar, h-index, self-citations, scientists ranking.

I. INTRODUCTION

In academics, the h-index is used to measure the productivity of the individual researchers, h-index of an article h in the descending order of citations cannot have citations less than the value of h. For example, article 43 must has citations greater than or equal to 43. If the article 43 has citations less than 43 then the article cannot be included in h-core. On the other hand, the h-core is the number of articles equal to or greater than h citations, e.g., h-index 43 means 43 articles are available in h-core [1]. We can find several h-index variants in the existing literature [2]–[10]. For example, 37 h-index variants are reviewed based on the meta-analysis in [2], in which it is concluded that most of the variants are redundant since they are correlated to the h-index. An assessment of the scientists based on the publications having citations more than their references is articulated in [3]. In this work, the author

The associate editor coordinating the review of this article and approving it for publication was Dimitrios Katsaros.

has evaluated 20 variants on the dataset of mature scientists and found that ranking the publications based on their number of citations is not a very useful tool. In [4], it is observed that the score of various indices under study increases the number of publications. In [5], 108 variants have been reviewed to find out correlation among those articles. Therein, most of the indices focused on the individual performance of the researchers by ignoring the other aspects. In [6], a related analysis of several indicators is presented for the assessment of a group of scientists, in which more than one indicators are used for a group of scientists. More recently, ha index is proposed to find out credit in the co-authorship [8]. This index gives more credit to the high h-index co-authors without considering the contribution of the authors. Further, hla index is suggested to add up other aspects in the calculation of h-index such as career length and discipline in [9]. It has been extended with the PageRank algorithm to remove limitations of h-index [10].

Unfortunately, self-citations are done by the researchers to inflate their h-index [11]. Usually, the less productive researchers who cannot attract others for citations practice self-citation [12]. The self-citation is motivated by self-interest since the author who cites himself/herself is cited by the others [13]. The self-citations up to 10-20% are perceived as normal and above this average is considered offensive [14]. We can find three types of self-citations that are known as the journal, author and co-author self-citations. The journal self-citation refers to when one article cites anonymous papers of the journal where this article is itself published [15]. The author level self-citation happens when one researcher cites his/her own published paper [16]. However, whenever one researcher, e.g. A, cites a published paper of his/her coauthor, i.e., B, is what we called as co-author self-citation [17]. The analysis of in-text citations on PubMed and Elsevier datasets identify that the number of self-citations in the citing document increases the total number of citations [18], [19]. In one study, the top 96 universities of USA considered and concluded that the highest Institutional Self-Citations Rates (ISCR) were identified as 31% for one of the universities [20]. Another study on Norway publications articulates that self-citations exist as 36% on average [14], which can lead to a measurable inaccuracy in the academic level rankings of the journals, authors and their articles.

The Google Scholar database facilitates retrieval of the worldwide scholarly literature. It provides access to the large range of databases including public and personal databases their access is not commonly available. Google Scholar freely grants access privilege to all those databases alternatively in addition to organizing the related documents by Google search engine. On contrary to this, the Web of Science (WoS) and Scopus are paid sources and are considered as reliable sources to analyze scientific activity. Compared to Google Scholar, WoS and Scopus do not cover the exclusive sources. However, Google Scholar contains unauthentic material, e.g., unpublished information, conference presentations, and internal reports, etc. that are otherwise not significant [21]. This is how Google Scholar is not considered reliable to compute the h-index because it does not exclude self-citations from the list of citations. The correct ranking by Google Scholar can be achieved by examining the top researchers and eliminating their information related to the self-citations.

In the existing literature, extensive studies have been done in biology, ecology, biomedicine and material sciences with respect to self-citations [16], [22]–[24]. Further, journal self-citations have been analyzed to assess the quality of journals since the Journal Impact Factor (JIF) is also increased significantly by self-citations [25]. So, the journals can also be verified as abnormal with having maximum self-citations. Such journals are automatically recognized based on empirical analysis of the dataset [26]. There are 6.9% journals that suspect out of 1138 journals included in Journal Citation Reports (JCR) 2014 [27]. Further, the exclusion of self-citations from the papers available in h-index updates the list of h-index sufficiently [28], [29]. The self-citations affect h-index and its variants differently [30], [31]. Some of the indexes capture manipulations while others inflate [32]. Additionally, institutional self-citations have been considered for their accurate ranking.

Few popular h-index variants have been introduced to improve the h-index that include w-index [33], g-index [34], a-index [35] and AR-index [36]. In addition, some measures like the ch-index have been introduced to identify the self-citations. The ch-index exclude self-citations from the list of citations for a particular paper. It does not consider co-authorship and journals for the subtraction of self-citations [37]. Another measure is the b-index that compute self-citations without analyzing complete citations of the articles by using a mathematical formula. It returns h-index of the article by eliminating self-citations. Moreover, the b-index considers top ranked papers related to the field and estimate self-citations. It also computes self-citations for individual authors but not for co-authors and the journals [38].

The dh-index is another measure that removes self-citations from the h-index without prior knowledge about the distribution of citations. It considers individual author and journal but not co-authors. The dh-index presents a new measure in comparison to the h-index for self-citations [11]. More recently, the nested self-citations have appeared in the literature. The part of the paper such as a diagram, table or text box is cited in the papers. Such constructs should not be separately considered as the citation [39]. It is also identified that fewer erroneous publications are added in the author's profiles of Google Scholar. Each author must update his account by deleting such publications [40]. Further, Thomson Reuters already suspended indexing a given journal whose number of self-citations grow above a certain threshold. The problem with this approach is to define what is an "acceptable" percentage of the self-citations [11].

Few variants of h-index can be found in [1]. However, no one eliminates author, coauthor and journal self-citations as a whole. Hence, there exists ample space to completely eliminate all these types of self-citations. We need to generate

TABLE 1. Comparison of proposed work with the existing schemes.

Authors	Journal	Author	Co-author	Filter	Dataset	Modify Author Index for Journals	Combine Author, Co-Author & Journals	Real-time Feasibility
Yu *et al.* [27]	✓	✗	✗	✗	JCR	✗	✗	✓
Donner [29]	✗	✓	✓	✗	Scopus	✗	✗	✓
Milz & Seifert [24]	✗	✓	✓	✗	DBLP	✗	✗	✓
Zhao *et al.* [19]	✗	✓	✓	✗	PubMed	✗	✗	✓
Costas *et al.* [16]	✗	✓	✓	✗	WoS	✗	✗	✓
Ferrara & Romero [11]	✗	✓	✓	✓	Google Scholar	✗	✗	✓
Hirsch [1]	✗	✓	✓	✗	ISI-NCR	✗	✗	✓
Frandsen [25]	✓	✗	✗	✗	SSCI	✗	✗	✓
Yu, Yu, and Wang [26]	✓	✗	✗	✗	JCR	✗	✗	✓
Proposed Work	✓	✓	✓	✓	Google Scholar	✓	✓	✓

a true productivity measure of the scientists because authors not only involve in self-citations but they are also supporting journals to increase their impact factors. In this paper, we propose a variant of h-index called updated h-index to eliminate self-citations. We analyze two datasets therein one contains scientists awarded by the Scientometrics and other is composed of non-award winner scientists. These are collected from the Google Scholar within the period of 1984 to 2017 so as to compute the self-citations influence in the h-index. Finally, a new ranking of the scientists is achieved from the updated h-index. Our updated h-index adds journal level self-citations calculation in the author level measure.

To the best of authors' knowledge, no comparative work exists that consider journals in the calculation of self-citations. This could be helpful to notice that the journals are important to be included in the calculation of self-citations because these are not only used to manipulate their individual IF but also support to add up author's and co-author's self-citations. In return, the impact of the author's paper is attached to the target journal impact. In addition, professional journals having journal citations more than a threshold should be added in the list of suspected journals. Therefore, the impact of journals must also be included in the author level indexes to calculate the correct author's h-index and so the accurate deduction of self-citations.

We briefly compare our work with the existing approaches in terms of qualitative parameters to highlight its novelty in Table 1. The qualitative parameters include **Journal**, **Author**, and **Co-Author** in addition to the following. **Filter:** To test whether a work eliminates all of the self-citations or it applies some criteria to keep the legitimate self-citations. **Dataset:** The benchmark of the dataset to authenticate the unbiased and accurate results. **Modify author index for journals:** To check whether the modification of author level h-index is applicable for the journals. **Combine Author, Co-Author, and Journals:** To evaluate whether a given work together takes into account all three types of self-citations, i.e. Author, Co-Author, and Journal. **Real-time feasibility:** To evaluate the deployed tools and applications of a given work are relevant to the real environment.

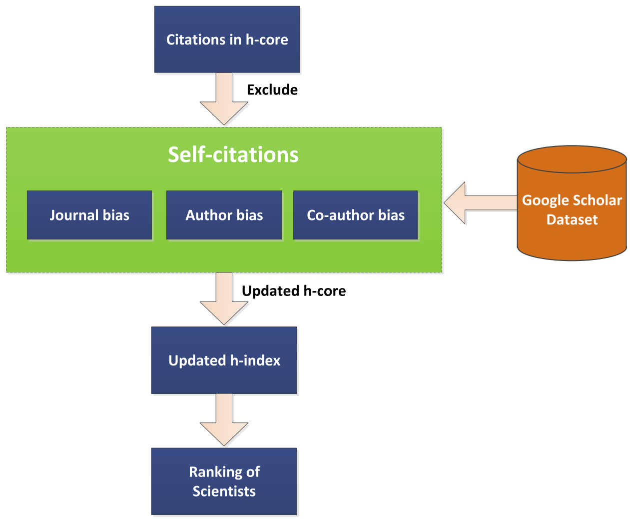

FIGURE 1. The proposed architecture.

To put in a nutshell, we have summarized the contributions of this paper in the following:

- We propose a variant of h-index called updated h-index that first time considers the inclusion of journals in the author level index.
- We define a rational individual threshold for the journal, author and co-author to subtract the legitimate self-citations only.
- We have built the updated datasets of the award winner and non-award winner scientists from the Google Scholar which were not available before.
- A new ranking of the scientists is achieved, which is significant to evaluate the true productivity of the journals and the authors compared to the h-index.

The rest of the paper is organized as follows. Section II describes materials and methods. Section III presents results discussion. Finally, the last section concludes the paper.

II. MATERIALS AND METHODS

We scrutinize the papers available in h-core of scientists. In this connection, we analyze the datasets that are generated from the Google Scholar. We have shown the architecture of the proposed system in Figure 1. In the first step of the proposed system, the citation information of the publications available in the h-core of the h-index for each scientist (available in the Google Scholar datasets) is taken as the input. In the second step, journal, author and co-author self-citations are identified and thereafter they are excluded from the dataset. As a result, updated h-core is generated, which is used to calculate the updated h-index subsequently. From the updated h-index, we have eventually created the latest ranking of the scientists in the last step. Notice that the terms self-citation and bias are interchangeably in this paper.

The main purpose of this work is to exclude self-citations from the h-index. We distinguish the three type of self-citations as Journal bias, Author bias, and Co-author bias. These self-citations are required to be excluded from the citations of each article of an author. The proposed three types of self-citations (or bias) are defined bellow.

A. JOURNAL BIAS

The journal self-citation can be captured by exploring the references list where articles from the same journal are cited. For example, the citations of the paper published in Scientometrics belong to the same journal. Hence, journal bias can be written as,

$$jbias = \sum_{p=0}^{n} jcit_p, \tag{1}$$

where *jbias* for paper p is the total number of self-citations from papers published in the same journal as p, $jcit_p$ denotes self-citations for paper p published in a journal.

The authors' own citations still have an effect on the coherence of their research by letting alone the citations from the same journal. Towards this end, some fields are more known that have many citations from the same journal. If all journal citations of an author are removed regarded as self-citations, then the authors' influence cannot be measured accurately. We need a reasonable threshold that should be applied to allow legitimate journal self-citations. For example, Thomson Reuters considers a journal as suspected if it has more than 20% self-citations [11]. In its continuity, we have applied a 20% filter while calculation of the journal self-citations. Therefore, the journal self-citations for each paper can be represented as a function,

$$f_j(jbias) = \begin{cases} jbias - (gcit \times 0.2) & jbias \geq gcit \times 0.2 \\ 0 & jbias < gcit \times 0.2, \end{cases} \tag{2}$$

where *gcit* represent gross-citations including citations and self-citations. We here set the *jbias* as 0 if the self-citations are less than 20% of the gross citations and *jbias* is set to the biased value exceeding 20% limit otherwise.

B. AUTHOR BIAS

The author self-citations can be identified by matching authors of the citing paper with the authors available in its references list [16], which is given as,

$$abias = \sum_{p=0}^{n} acit_p, \tag{3}$$

where *abias* for paper p is the total number of self-citations from papers written by the same author as p, $acit_p$ denotes self-citations for paper p written by an author.

It is not reasonable to consider an author's all own citations as the self-citations since few of them may provide a foundation of the current work. In this regards, we have applied a 5% filter on the self-citations to remain legitimate author's own citations [41]. Hence, the author self-citations for each paper can be represented as a function,

$$f_a(abias) = \begin{cases} abias - (gcit \times 0.05) & abias \geq gcit \times 0.05 \\ 0 & abias < gcit \times 0.05, \end{cases} \tag{4}$$

where we set the *abias* as 0 if the self-citations are less than 5% of the gross-citations for respective paper and otherwise *abias* is set to the biased value exceeding the 5% limit.

C. CO-AUTHOR BIAS

The co-author bias can be calculated by the identification of common co-authors of one particular researcher. The co-authors of h-index can be considered in the co-author bias as follows,

$$cbias = \sum_{p=0}^{n} ccit_p, \tag{5}$$

where *cbias* for paper p is the total number of self-citations from papers written by the same co-author as p, $ccit_p$ denotes self-citations for paper p co-written by a co-author.

We have also applied a 5% filter on the self-citations to keep a legitimate co-author's own citations as *abias*, *cbias* can maximum be up to 5% in a particular paper [41]. Therefore, co-author self-citations for a paper can be calculated as a function,

$$f_c(cbias) = \begin{cases} cbias - (gcit \times 0.05) & cbias \geq gcit \times 0.05 \\ 0 & cbias < gcit \times 0.05, \end{cases} \tag{6}$$

where we set the *cbias* as 0 if the self-citations are less than 5% of the citations for respective paper and *cbias* is set to the biased value exceeding the 5% limit otherwise.

From Equations (2,4,6), net-citations (denoted by *ncit*) can simply be calculated as,

$$ncit = gcit - \left[f_j(jbias) + f_a(abias) + f_c(cbias)\right]. \tag{7}$$

This would be helpful to notice that a single citation cannot be overlapped such that it can either be a journal-bias, author bias or co-author bias but not all at the same time. For example, an article has only one citation, it will either be counted as 1 in terms of a journal, an author or a co-author self-citation by marking the variable $ncit = 0$.

TABLE 2. Dataset specifications.

Window	Number of scientists	Total articles	Total citations	Total h-core citations
1984-2017	28	8406	376670	315113

TABLE 3. Derek de Solla Price Memorial Medal's data of the award winner scientists from 1984 to 2017.

No.	Year	Name	Articles	Citations	h-core citations	h-index	w-index	g-index	a-index	r-index
1	1984	Eugene Garfield	1513	27838	20674	62	20	166	333.45	143.78
2	1985	Michael J. Moravcsik	398	4258	2759	27	8	65	102.19	52.53
3	1986	Tibor Braun	445	8929	5676	46	13	94	123.39	75.34
4	1987	Vasily V. Nalimov	65	866	707	12	5	29	58.92	26.59
5	1987	Henry Small	109	8989	8578	36	15	94	238.27	92.62
6	1988	Francis Narin	132	13788	12920	49	18	117	263.67	113.67
7	1989	Bertram C. Brookes	158	3566	3158	26	9	59	121.46	56.20
8	1989	Jan Vlachy	170	1096	447	15	3	33	29.80	21.14
9	1993	Andras Schubert	247	10313	8702	47	14	101	185.15	93.28
10	1995	Anthony F.J. Van Raan	179	10568	8899	54	17	102	164.80	94.33
11	1995	Robert K. Merton	481	119048	111771	100	46	345	1117.71	334.32
12	1997	Ben Martin	206	11611	10364	43	16	107	241.02	101.80
13	1997	John Irvine	75	3349	3031	26	7	57	116.58	55.05
14	1997	Belver C. Griffith	101	6562	6083	28	10	81	217.25	77.99
15	1999	Wolfgang Glanzel	431	13583	9616	61	17	116	157.64	98.06
16	1999	Henk F. Moed	244	10839	9413	54	17	104	174.31	97.02
17	2001	Leo Egghe	340	8672	6407	39	10	92	164.28	80.04
18	2001	Ronald Rousseau	492	11129	7222	49	13	105	147.39	84.98
19	2003	Loet Leydesdorff	730	35907	27088	86	23	189	314.98	164.58
20	2005	Peter Ingwersen	282	9866	8487	36	13	99	235.75	92.12
21	2005	Howard D. White	116	7010	6436	31	12	83	207.61	80.22
22	2007	Katherine W. McCain	114	5590	5176	30	9	74	172.53	71.92
23	2009	Péter Vinkler	103	2602	2096	29	8	51	72.28	45.78
24	2009	Michel Zitt	79	2066	1875	25	8	45	75.0	43.30
25	2011	Olle Persson	145	4465	3952	27	10	66	146.37	62.86
26	2013	Blaise Cronin	378	8632	5896	46	12	92	128.17	76.79
27	2015	Mike Thelwall	475	18897	12561	73	19	137	172.07	112.08
28	2017	Judit Bar-Ilan	198	6631	5119	42	14	81	121.88	71.55

D. DATASET

In our analysis, we have chosen two datasets that are individually related to the award winner and non-award winner scientists. The datasets were built from Google Scholar in December 2017. The specifications of the award winner dataset (of Derek de Solla Price Memorial Medal) are shown in Table 2, which belongs to all those 28 scientists being awarded as the best researchers on the basis of their highest citations and h-index. In the award winner dataset, there are 8406 articles of all the scientists are and their total citations are 376670. According to h-index, Mr. Robert K. Merton is considered as the best researcher. He has 111771 citations in the h-core articles, while his h-index is 100.

We have shown the detailed dataset of award winner scientists in Table 3, in which information regarding the year of award-winning, the number of total published articles, total citations produced, and citations available in h-core are inscribed for quick reference. Further, various available indexes are also given in this table including h-index, w-index, g-index, a-index, and r-index. The breakdown of the dataset relating to different parameters, i.e., journals, authors, and co-authors of the h-core articles are also engendered in this dataset. On the contrary, the dataset related to the non-award winner scientists is given in Table 4, which contains the list of scientists who are not the award winners but have profiles closely related to that of the award winner scientists. The scientists in both datasets are falling from 1984 to 2017.

In the non-award winner dataset, the total articles of all scientists are 5613 and their total citations are 1446589. According to h-index, Mr. Milton Friedman has 76381 citations in the h-core articles and has h-index as 101. We have shown the chosen dataset of non-award winners scientists with more detail in Table 4. The non-award scientists have similar h-index to that of award winner scientists. These datasets are selected for comparison purpose so as to analyze the self-citations behavior of the scientists.

In our analysis, the journal bias is counted based on *j_bias* and subtracted from the citations of an author's paper in h-index. The author bias is counted for each paper available in h-index from citing articles having authorship of the same author. Further, co-author bias is counted based on the co-authors available in h-index papers that exist in the citing papers. To the best of our knowledge, the updated dataset was not available in any of the current research. Thus, the updated datasets have been built and used for the current state of the

TABLE 4. Data of the non-award winner scientists from 1984 to 2017.

No.	Name	Articles	Citations	h-core citations	h-index	w-index	g-index	a-index	r-index
1	Milton Friedman	2360	136173	76381	101	37	369	756.25	276.37
2	Philip Kim	579	74304	69467	89	42	273	780.53	263.57
3	Stephan Meyer	371	74612	65706	75	31	273	876.08	256.33
4	John Terning	212	74172	42422	61	19	272	695.44	205.97
5	W James Kent	135	77165	61724	67	36	278	921.25	248.44
6	Christopher M Bishop	260	75743	17579	53	22	275	331.68	132.59
7	Claude E Shannon	313	138557	29111	55	25	372	529.29	170.62
8	R Sekhar Chivukula	232	74442	30704	46	15	273	667.48	175.23
9	David Lowe	114	80579	72311	48	27	284	1506.48	268.91
10	Florencio Lopez de Silanes	260	81847	35372	48	18	286	736.92	188.07
11	Juan E Peralta	161	76308	15413	40	11	276	385.33	124.15
12	Thomas L Madden	52	83731	78162	30	24	289	2605.4	279.57
13	Fred D Davis	170	111894	83654	38	17	335	2201.42	289.23
14	Koichiro Tamura	120	122497	110991	30	19	350	3699.7	333.15
15	Peter Dalgaard	94	82765	2877	21	8	288	137	53.64
16	R Core Team	180	81800	3450	23	11	286	150	58.74

results. It can also be taken as the latest version of the works published ever since. In these datasets, the co-authors are enumerated maximum up to a 3rd position in the authorship. The updated h-index complements the h-index in the sense that it addresses current problems, which were not available in the past but now they have emerged in the race of increasing the impact of research.

III. RESULTS AND DISCUSSION

We have performed experiments on both of the chosen datasets to conduct a comparative analysis. We here discuss the key results of our analysis in the following.

A. AWARD WINNER SCIENTISTS

We mention that the updated h-index is calculated based on the proposed three self-citations (journal, author, and co-author) from Equations (1, 3, 5). We discuss one of the scientists Mr. Mike Thelwall as a toy example to understand how three bias affect when self-citations are excluded from his top-15 articles. Therein, the breakdown of the same papers with reference to the types of self-citations is further detailed. Then, the excluded articles from h-core of Mr. Mike Thelwall are presented. We calculated the updated h-index on the same pattern to generate the ranking of the scientists otherwise.

The updated h-index provides a true picture of the citations of the author's articles and their ranks. It significantly changes the h-index of the scientists. We would like to discuss a toy example to understand how our analysis works. For the scientist Mr. Mike Thelwall, top 15 papers have maximum citations as shown in Table 5. The total citation column shows current citations of a particular paper whereas updated citations column exhibits citations after excluding the self-citations. For example, biased citations of Mr. Mike for paper 5 are 78. So, Mr. Mike's actual citations reached 318 when biased citations are removed from his total citations (i.e., 396). In general, Mr. Mike has 12561 citations in h-core from 475 articles. The total biased citations are 2365, which is a notable number. The aggregated self-citations (Journal

TABLE 5. Detail of Mr. Mike Thelwall's articles.

No.	Total citation	Updated citations	Self-citations
1	839	839	0
2	567	567	0
3	435	435	0
4	427	427	0
5	396	318	78
6	391	391	0
7	387	334	53
8	309	257	52
9	304	304	0
10	297	240	57
11	273	273	0
12	271	197	74
13	270	127	143
14	238	238	0
15	224	102	122

bias, Author bias, and Co-author bias) are revealed for each paper. Commonly, self-citations are increased proportionally with the increase of total citations. However, in this case, Mr. Mike Thelwall is ranked from 3rd position in the h-index to 4th position according to updated h-index. The h-index of Mr. Mike is 73, which is reduced to 59 after the computation of updated h-index, when 14 articles have been excluded from the h-core.

We refer to the biased citations of Mr. Mike Thelwall in Table 6. For instance, the self-citations of Paper 5 are 78. It includes 0 journal biased citations, 39 author biased citations and 39 co-author biased citations. It should be noticed that the self-cited papers might be most relevant to the proposed work in one particular paper. Few of such papers might have been excluded by the updated h-index with the tag of biased articles. The author may refer to his previous (relevant) work in his paper, but the updated h-index has dropped out it by considering in the category of self-citations. So such citations should be scrutinized and excluded from the author self-citations. Co-author bias is the most critical in which chances of bias are increased since authors refer papers of their peers in order to increase the number of citations. Each

TABLE 6. Breakdown of Mr. Mike Thelwall's articles with respect to individual bias.

No.	Total citations	Total bias	Journal bias	Author bias	Co-author bias
1	839	0	0	0	0
2	567	0	0	0	0
3	435	0	0	0	0
4	427	0	0	0	0
5	396	78	0	39	39
6	391	0	0	0	0
7	387	53	0	28	25
8	309	52	0	34	18
9	304	0	0	0	0
10	297	57	0	36	21
11	273	0	0	0	0
12	271	74	0	40	34
13	270	143	0	78	65
14	238	0	0	0	0
15	224	122	0	71	51

TABLE 7. Mr. Mike Thelwall's excluded articles from h-core.

No.	Total citation	Journal bias	Author bias	Co-author bias	Total bias	Actual Citations
1	100	0	24	18	42	58
2	83	0	6	19	25	58
3	90	0	15	18	33	57
4	73	0	17	0	17	56
5	90	0	22	13	35	55
6	91	0	25	15	40	51
7	73	0	23	0	23	50
8	88	0	13	27	40	48
9	80	0	26	7	33	47
10	89	0	31	12	43	46
11	87	0	28	15	43	44
12	94	0	32	19	51	43
13	77	0	12	22	34	43
14	78	0	21	17	38	40

parameter needs detailed analysis individually to achieve an accurate figure of biased behavior in this regards.

As shown in Table 7, 14 articles have been similarly dropped from the h-core of Mr. Mike Thelwall. 73 papers in h-core exist before the computation of self-citations. The data has been ordered by actual citations. Article 1 has maximum 100 citations in the list of excluded articles. So, the actual citations have been reached to 46 due to 54 biased citations. The author bias is observed high than that of any other biased citation. However, any other type of bias may be increased otherwise since this observation is not constant. The maximum total biased citations of article 12 are 51 that shifted this article at one of the ending positions of the h-core. The excluded articles (mentioned in the parenthesis followed by the scientists name) from h-core of top 10 scientists are as follows: Mr. Robert K. Merton (0), Mr. Loet Leydesdorff (11), Mr. Mike Thelwall (14), Mr. Eugene Garfield (1), Mr. Wolfgang Glanzel (11), Mr. Anthony F.J. Van Raan (4), Mr. Henk F. Moed (6), Mr. Francis Narin (2), Mr. Ronald Rousseau (8), and Mr. Andras Schubert (11). Mr. Mike Thelwall is observed with a maximum of 14 articles that are excluded from the h-core. The presented results are related to top scientists. The self-citations may get worse when dataset of other scientists is analyzed.

In Table 8, we present a complete detail, i.e. total published articles, citations in h-core, self-citations, h-index, and updated h-index, of the top 10 scientists. We juxtaposed the h-index and updated h-index to make a comparison. For example, h-index of Mr. Eugene Garfield is 62 whereas his updated h-index is 61. The maximum self-citations of the Mr. Loet Leydesdorff are 2806 from the h-core citations 27088. Based on the results, it can be noticed that the proportion of the self-citations increases as the function of publications and their associated citations. The readers may rank scientists on the basis of self-citations to get knowledge of their productivity. The reliability of Google Scholar can be judged further from this data because papers of the scientists are published in the quality venues.

The self-citations of Top-10 scientists are observed fraction to their citations. These can simply project other researchers having low-quality publications. The distribution of citations and self-citations of Top-10 scientists are shown in Figure 2. Therein, an order in the h-core citations exists because these are classified based on the updated h-index. For example, Mr. Eugene Garfield is ranked 3rd according to updated h-index. We observe that if a scientist has more citations, he/she highly likely to have higher h-index. However, the h-index of Mr. Anthony F. J. Van Raan contradicts to this

TABLE 8. Comparison of h-index and the updated h-index for top 10 award winner scientists.

No.	Name	Articles	h-core citations	Self-citations	h-index	Updated h-index
1	Robert K. Merton	481	111771	0	100	100
2	Loet Leydesdorff	730	27088	2806	86	75
3	Mike Thelwall	475	12561	2365	73	59
4	Eugene Garfield	1513	20674	198	62	61
5	Wolfgang Glanzel	431	9616	1901	61	50
6	Anthony F. J. Van Raan	179	8899	719	54	50
7	Henk F. Moed	244	9413	547	54	48
8	Francis Narin	132	12920	270	49	47
9	Ronald Rousseau	492	7222	992	49	41
10	Andras Schubert	247	8702	1728	47	36

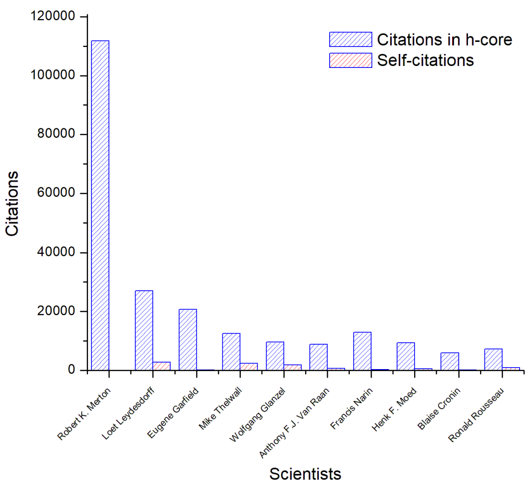

FIGURE 2. Distribution of citations and self-citations of the top 10 award winner scientists.

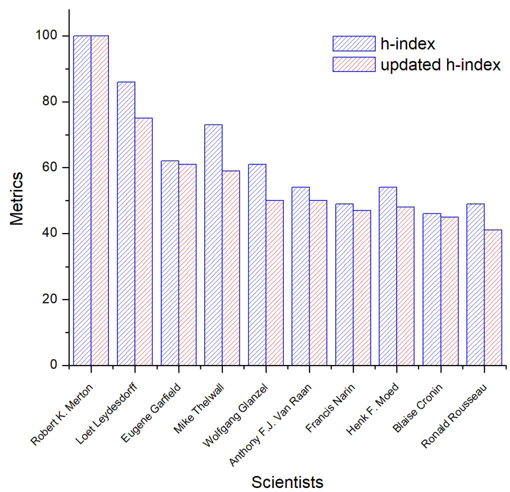

FIGURE 3. Comparison of h-index and updated h-index of award winner scientists.

behavior because it maintains citations across all h-core papers, i.e. less variability in the citations of the papers exist in h-core. We also witness that number of self-citations (compared to total citations in average) puts an impact on the h-index. For example, Mr. Wolfgang Glanzel has 9616 citations in h-core and 1901 self-citations. His h-index reduced from 61 to 50. This can be seen from the results that self-citations are inversely proportional to the ranking. That is when self-citations increased the rank is dropped, e.g., Mr. Andras Schubert dropped from the top 10th position.

We have shown the ranking of scientists based on the updated h-index in Table 9. According to updated h-index ranking, Mr. Robert K. Merton is consistently observed at the 1st position with h-index 100. Mr. Eugene Garfield has promoted by one rank and so placed at top 3rd position with the h-index 61. Mr. Mike Thelwall has demoted by one rank with h-index 59. The comparison of h-index and updated h-index is shown in Figure 3, in which we can see that the maximum 23% of h-index is reduced for Mr. Andras Schubert. In the sequence of percentage reduction, top-ranked scientists include Mike Thelwall (19%), Wolfgang Glanzel (18%), Ronald Rousseau (16%), Loet Leydesdorff (13%), Henk F. Moed (11%), Anthony F.J. Van Raan (7%), Francis Narin (4%), Eugene Garfield (2%) and Robert K. Merton (0%). We notice that if self-citations for a particular paper increase and subtraction of this value decrease citations of the paper from a threshold (h-index value), it is then excluded from the h-core. For example, Mr. Francis Narin and Mr. Ronald Rousseau hold index as 49 but their updated h-index is different, i.e. 47 and 41 respectively. We also observe that few of the fake papers are added in the authors' profiles that must be identified and thereafter excluded from the papers list. By doing so, it will also improve the quality of the h-index measured by Google Scholar.

We have presented a citation analysis in Figure 4. It shows the total citations of each scientist, actual citations, and biased citations. Therein, we can easily observe that biased citations reduce the total citations as follows: Mr. Wolfgang Glanzel (20%), Mr. Mike Thelwall (19%), Mr. Ronald Rousseau (14%), Mr. Loet Leydesdorff (10%), Mr. Anthony F.J. Van Raan (8%), Mr. Henk F. Moed (6%), Mr. Blaise Cronin (3%), Mr. Francis Narin (2%), Mr. Eugene Garfield (1%), and Mr. Robert K. Merton (0%). According to the results, actual and biased citations make up the total citations. There is a small difference between total and actual citations. It can be

TABLE 9. Ranking of the award winner scientists with the updated h-index.

No.	Name	Articles	Citations	h-core citations	Updated h-index
1	Robert K. Merton	481	119048	111771	100
2	Loet Leydesdorff	730	35907	27088	75
3	Eugene Garfield	1513	27838	20674	61
4	Mike Thelwall	475	18897	12561	59
5	Wolfgang Glanzel	431	13583	9616	50
6	Anthony F.J. Van Raan	179	10568	8899	50
7	Henk F. Moed	244	10839	9413	48
8	Francis Narin	132	13788	12920	47
9	Blaise Cronin	378	8632	5896	45
10	Ronald Rousseau	492	11129	7222	41

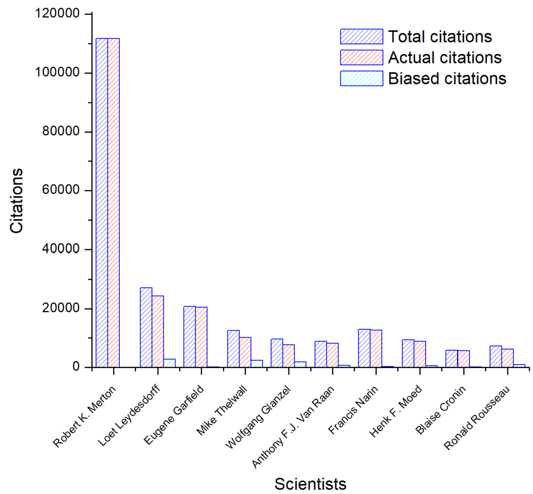

FIGURE 4. Comparison of all award winner scientists by total and biased citations.

seen that a decrease in actual citations likely decreases the h-index. For instance, Mr. Francis Narin has 11896 actual citations, 270 biased citations and his h-index is decreased from 49 to 47.

B. NON-AWARD WINNER SCIENTISTS

We have shown the comparison of h-index and updated h-index in Figure 5, in which we can see that the maximum 30% of h-index is reduced for Mr. R. Sekhar Chivukula. In the sequence of percentage reduction, top-10 scientists include Stephan Meyer (25%), John Terning (15%), Philip Kim (3%), W. James Kent (3%), David Lowe (2%), Milton Friedman (2%), Christopher M. Bishop (2%), Claude E. Shannon (0%) and Florencio Lopez de Silanes (0%). In comparison to the award winner scientists (where there is a maximum 23% h-index reduction), a non-award winner scientist has otherwise maximum 30% h-index reduction.

We have exhibited the citation analysis in Figure 6. It shows the total citations, actual citations, and biased citations of each scientist. We observe that biased citations reduce the total citations as follows: Mr. Claude E. Shannon (10%), Mr. Philip Kim (10%), Mr. R. Sekhar Chivukula (7%), Mr. Stephan Meyer (5%), Mr. John Terning (5%), Mr. W. James Kent (4%), Mr. David Lowe (2%), Mr. Christopher M. Bishop (2%), Mr. Milton Friedman (1%), and Mr. Florencio Lopez de Silanes (0%). Again, there no

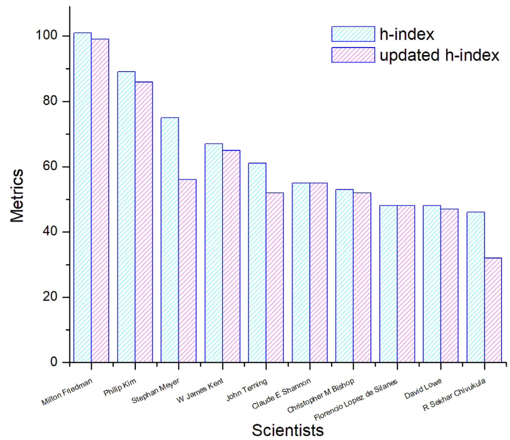

FIGURE 5. Comparison of h-index and updated h-index of non-award winner scientists.

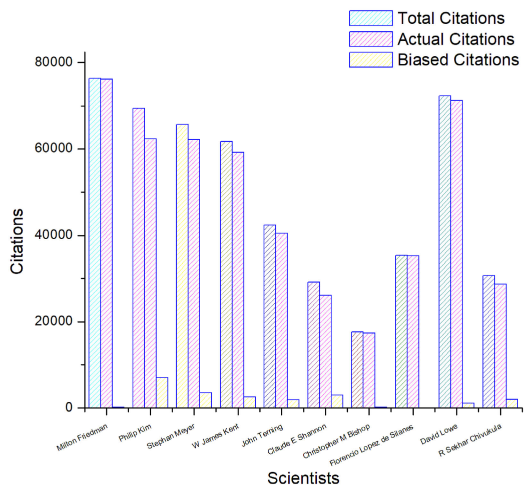

FIGURE 6. Comparison of all non-award winner scientists by total and biased citations.

considerable difference of biased citations for these scientists exists that that of the award winner scientists.

C. COMPARISON OF AWARD WINNER AND NON-AWARD WINNER SCIENTISTS

The difference in the ranking of researchers after the updated h-index are measured quantitatively in Table 10 and Table 11 for the award winner and non-award winner scientists, respectively. According to the "Difference" column in both

TABLE 10. Difference of ranking resulting from the updated h-index for award winner scientists.

No.	Name	h-index Rank#	updated h-index Rank#	Difference
1	Robert K. Merton	1	1	0
2	Loet Leydesdorff	2	2	0
3	Eugene Garfield	4	3	1
4	Mike Thelwall	3	4	1
5	Wolfgang Glanzel	5	5	0
6	Anthony F.J. Van Raan	6	6	0
7	Henk F. Moed	7	7	0
8	Francis Narin	8	8	0
9	Blaise Cronin	10	9	1
10	Ronald Rousseau	9	10	1

TABLE 11. Difference of ranking resulting from the updated h-index for non-award winner scientists.

No.	Name	h-index Rank#	updated h-index Rank#	Difference
1	Milton Friedman	1	1	0
2	Philip Kim	2	2	0
3	W James Kent	4	3	1
4	Stephan Meyer	3	4	1
5	Claude E Shannon	6	5	1
6	John Terning	5	6	1
7	Christopher M Bishop	7	7	0
8	Florencio Lopez de Silanes	8	8	0
9	David Lowe	9	9	0
10	R Sekhar Chivukula	10	10	0

TABLE 12. Comparison with actual applications.

Parameter	Source	Remove Self-citations	Journal Self-citations	Author Self-citations	Filter
CIDS	Google Scholar	✓	✗	✓	✗
Updated h-index	Google Scholar	✓	✓	✓	✓

TABLE 13. Self-citations computation by WoS, Google Scholar and Scopus.

Database	Self-citations	Source	Filter (Journal)	Filter (Author & Co-author)	Both filters (Combined)
WoS	✓	Limited	20%	✗	✗
Google Scholar	✗	vast	✗	✗	✗
Scopus	✓	Limited	✗	✗	✗
Updated h-index	✓	Google Scholar	20%	5%	✓

tables, the ranking is not significantly changed due to less number of self-citations.

D. COMPARISON WITH ACTUAL APPLICATIONS

One real application of excluding self-citations from the h-index is Citation Impact Discerning Self-citations (CIDS) [42]. The CIDS uses CiteSeers policy for the identification of self-citations. According to that policy, it checks the header of the indexed paper to mark as self-citation. On the other hand, our updated h-index is the variation of the h-index that not only counts self-citations from Google Scholar similar to CIDS but adds journal self-citations as well. As the journal self-citations are emphasized to be included in the self-citations calculation in the author level h-index, the legitimate and illegitimate self-citations are similarly required to be filtered by the updated h-index. The self-citations that are subtracted in the real applications are: 1) Illegitimate journal; 2) Illegitimate author; 3) Illegitimate co-author.

The most popular databases that maintain publications' data are WoS, Google Scholar, Scopus, etc. Few of them provide the facility to check self-citations. We summarize the self-citations that are practiced by these databases in Table 13. As shown in Table 12 and Table 13, we can notice that no existing application or database set a filter on the author and co-author level self-citations. Our proposed h-index has introduced a 5% filter which has been justified in the earlier section. It is safe to say that our updated h-index contains both journal and author, and co-author filter for h-index corrections.

IV. CONCLUSION

In this paper, we have addressed the self-citation problem in h-index papers on Google Scholar datasets. Therein, two Google Scholar datasets are built for the award winner and non-award winner scientists. For the exclusion of self-citations, three types of biases have been proposed called Journal bias, author bias and co-author bias. These biases are considered by subtracting from the papers of the scientists that exist in the h-core. Based on the updated h-core, an updated h-index has been computed. Using this updated h-index, a new ranking of the scientists is generated. The overall journal self-citations, author self-citations and co-author self-citations are observed as 3.95%, 2.86%, and 3.33%, respectively. In contrast, non-award winner scientists have average journal biased citations as 1.22%, author biased citations as 0.41% and co-author biased citations as 0.90%. According to the results, the ranking of scientists is considerably modified by using updated h-index in comparison to h-index. The updated h-index can be applied on Google Scholar, WoS, Scopus, and other databases to exclude legitimate self-citations. The results could also be investigated for further analysis with datasets from WoS and Scopus.

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, Aug. 2005.

[2] 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.

[3] M. Kosmulski, "Calibration against a reference set: A quantitative approach to assessment of the methods of assessment of scientific output," *J. Informetrics*, vol. 6, no. 3, pp. 451–456, Jul. 2012.

[4] P. Vinkler, "Would it be possible to increase the Hirsch-index, π-index or CDS-index by increasing the number of publications or citations only by unity?" *J. Informetrics*, vol. 7, no. 1, pp. 72–83, Jan. 2013.

[5] L. Wildgaard, J. W. Schneider, and B. Larsen, "A review of the characteristics of 108 author-level bibliometric indicators," *Scientometrics*, vol. 101, no. 1, pp. 125–158, 2014.

[6] A. P. dos Santos Rubem, A. L. de Moura, and J. C. C. B. S. de Mello "Comparative analysis of some individual bibliometric indices when applied to groups of researchers," *Scientometrics*, vol. 102, no. 1, pp. 1019–1035, Jan. 2015.

[7] L. Bertoli-Barsotti and T. Lando, "A theoretical model of the relationship between the h-index and other simple citation indicators," *Scientometrics*, vol. 111, no. 3, pp. 1415–1448, Jun. 2017.

[8] J. E. Hirsch, "hα: An index to quantify an individual's scientific leadership," *Scientometrics*, vol. 118, no. 2, pp. 673–686, Feb. 2019.

[9] A.-W. Harzing, S. Alakangas, and D. Adams, "hIa: An individual annual h-index to accommodate disciplinary and career length differences," *Scientometrics*, vol. 99, no. 3, pp. 811–821, Jun. 2014.

[10] 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.

[11] E. Ferrara and A. E. Romero, "Scientific impact evaluation and the effect of self-citations: Mitigating the bias by discounting the h-index," *J. Assoc. Inf. Sci. Technol.*, vol. 64, no. 11, pp. 2332–2339, Nov. 2013.

[12] C. Bartneck and S. Kokkelmans, "Detecting h-index manipulation through self-citation analysis," *Scientometrics*, vol. 87, no. 1, pp. 85–98, Apr. 2011.

[13] J. H. Fowler and D. W. Aksnes, "Does self-citation pay?" *Scientometrics*, vol. 72, no. 3, pp. 427–437, Jul. 2007.

[14] D. W. Aksnes, "A macro study of self-citation," *Scientometrics*, vol. 56, no. 2, pp. 235–246, Feb. 2003.

[15] R. Rousseau, "Temporal differences in self-citation rates of scientific journals," *Scientometrics*, vol. 44, no. 3, pp. 521–531, Mar. 1999.

[16] R. Costas, T. N. van Leeuwen, and M. Bordons, "Self-citations at the meso and individual levels: Effects of different calculation methods," *Scientometrics*, vol. 82, no. 3, pp. 517–537, Mar. 2010.

[17] W. Glanzel and B. Thijs, "Does co-authorship inflate the share of self-citations?" *Scientometrics*, vol. 61, no. 3, pp. 395–404, Nov. 2004.

[18] K. W. Boyack, N. J. van Eck, G. Colavizza, and L. Waltman, "Characterizing in-text citations in scientific articles: A large-scale analysis," *J. Informetrics*, vol. 12, no. 1, pp. 59–73, Feb. 2018.

[19] D. Zhao, A. Strotmann, and A. Cappello, "In-text function of author self-citations: Implications for research evaluation practice," *J. Assoc. Inf. Sci. Technol.*, vol. 69, no. 7, pp. 949–952, Jul. 2018.

[20] D. Hendrix, "Institutional self-citation rates: A three year study of universities in the United States," *Scientometrics*, vol. 81, no. 2, pp. 321–331, Apr. 2009.

[21] I. F. Aguillo, "Is Google Scholar useful for bibliometrics? A webometric analysis," *Scientometrics*, vol. 91, no. 2, pp. 343–351, May 2012.

[22] E. Gianoli and M. A. Molina-Montenegro, "Insights into the relationship between the h-index and self-citations," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 60, no. 6, pp. 1283–1285, Jun. 2009.

[23] M. Leblond, "Author self-citations in the field of ecology," *Scientometrics*, vol. 91, no. 3, pp. 943–953, Jun. 2012.

[24] T. Milz and C. Seifert, "Analysing author self-citations in computer science publications," in *Proc. Int. Conf. Database Expert Syst. Appl.* Cham, Switzerland: Springer, 2018, pp. 289–300.

[25] T. F. Frandsen, "Journal self-citations—Analysing the JIF mechanism," *J. Informetrics*, vol. 1, no. 1, pp. 47–58, Jan. 2007.

[26] T. Yu, G. Yu, and M.-Y. Wang, "Classification method for detecting coercive self-citation in journals," *J. Informetrics*, vol. 8, no. 1, pp. 123–135, Jan. 2014.

[27] T. Yu, G. Yu, Y. Song, and M.-Y. Wang, "Toward the more effective identification of journals with anomalous self-citation," *Malaysian J. Library Inf. Sci.*, vol. 23, no. 2, pp. 25–46, Jul. 2018.

[28] M. Schreiber, "Self-citation corrections for the Hirsch index," *EPL Europhys. Lett.*, vol. 78, no. 3, Apr. 2007, Art. no. 30002.

[29] P. Donner, "Enhanced self-citation detection by fuzzy author name matching and complementary error estimates," *J. Assoc. Inf. Sci. Technol.*, vol. 67, no. 3, pp. 662–670, Mar. 2016.

[30] R. H. Galvez, "Assessing author self-citation as a mechanism of relevant knowledge diffusion," *Scientometrics*, vol. 111, no. 3, pp. 1801–1812, Jun. 2017.

[31] I. Hellsten, R. Lambiotte, A. Scharnhorst, and M. Ausloos, "Self-citations, co-authorships and keywords: A new approach to scientists' field mobility?" *Scientometrics*, vol. 72, no. 3, pp. 469–486, Sep. 2007.

[32] G.-A. Viiu, "A theoretical evaluation of Hirsch-type bibliometric indicators confronted with extreme self-citation," *J. Informetrics*, vol. 10, no. 2, pp. 552–566, May 2016.

[33] 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.

[34] L. Egghe, *An Improvement of the H-Index: The G-Index*. Milpitas, CA, USA: ISSI, 2006.

[35] B. Jin, "H-index: An evaluation indicator proposed by scientist," *Sci. Focus*, vol. 1, no. 1, pp. 8–9, 2006.

[36] 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, 2007.

[37] 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.

[38] R. J. C. Brown, "A simple method for excluding self-citation from the h-index: The b-index," *Online Inf. Rev.*, vol. 33, no. 6, pp. 1129–1136, 2009.

[39] J. A. T. da Silva, "Nested self-citation: The citation of a paper's least divisible unit," *Scientometrics*, vol. 111, no. 1, pp. 547–552, Apr. 2017.

[40] J. A. T. da Silva, "The Google Scholar h-index: Useful but burdensome metric," *Scientometrics*, vol. 117, no. 1, pp. 631–635, 2018.

[41] M. Brysbaert and S. Smyth, "Self-enhancement in scientific research: The self-citation bias," *Psychologica Belgica*, vol. 51, no. 2, pp. 37–129, Oct. 2011.

[42] F. M. Couto, C. Pesquita, T. Grego, and P. Verissimo, "Handling self-citations using Google Scholar," *Cybermetrics, Int. J. Scientometrics, Informetrics Bibliometrics*, vol. 13, no. 2, pp. 1–7, 2009.

FIAZ MAJEED received the Ph.D. degree in computer sciences from the University of Engineering and Technology, Lahore, Pakistan, in 2016. He is currently an Assistant Professor with the Faculty of Computing and Information Technology, University of Gujrat (UOG), Pakistan. His research interests include data warehousing, data streams, information retrieval, social networks, and ad hoc networks. He has published more than 15 articles in refereed journals and international conference proceedings in the above areas. He is a Reviewer for several renowned international journals.

MUHAMMAD SHAFIQ received the M.S. degree in computer science from the University Institute of Information Technology, Arid Agriculture University, Rawalpindi, Pakistan, the master's degree in information technology from the University of the Punjab, Gujranwala, Pakistan, and the Ph.D. degree in information and communication engineering from Yeungnam University, South Korea, in 2018. From 2010 to 2018, he was a Lecturer with the Department of Information Technology, University of Gujrat and with the Department of Computer Science, Federal Urdu University, Islamabad, Pakistan. He is currently a Postdoctoral Fellow with Yeungnam University, and is also an Assistant Professor with the Faculty of Computer Science, GC Women University, Sialkot, Pakistan. His research interests include the design of spectrum management, routing, and medium access control protocols for mobile ad hoc networks, the IoT, and cognitive radio networks.

AMJAD ALI received the B.S. and M.S. degrees in computer science from the COMSATS Institute of Information Technology, Lahore, Pakistan, in 2006 and 2008, respectively, and the Ph.D. degree in electronics and radio engineering from Kyung Hee University, South Korea, in 2015. Since 2015, he has been an Assistant Professor with the Department of Computer Science, COMSATS University Islamabad, Lahore Campus, Pakistan. He is currently a Postdoctoral Fellow with the MNC Laboratory, Department of Electrical Engineering, Korea University, Seoul, South Korea. He has published a number of research articles in refereed international journals and conferences. His main research interests include multimedia transmissions over cognitive radio networks, 5G cellular networks, multimedia cloud computing, the Internet of Things, and intelligent transportation systems and sensor networks.

MUHAMMAD AWAIS HASSAN received the M.S. degree from the University of Engineering and Technology (UET), Lahore, where he is currently pursuing the Ph.D. degree and is also an Assistant Professor with the Computer Science and Engineering Department. He is a Gold Medalist of Punjab University. His research interests include natural language processing, semantic web, software architecture, and open-source software development.

SYED ALI ABBAS received the B.Sc. degree in mathematics from the University of Azad Jammu and Kashmir, the master's degree in computer science from the University of the Punjab, Pakistan, and the Ph.D. degree in computer Science from Chongqing University, China. He is currently an Assistant Professor with the University of Azad Jammu and Kashmir. His primary areas of research are regression analysis diagnostics specifically related to software effort estimation.

MOHAMMAD EID ALZAHRANI received the B.Sc. degree from the College of Sciences, King Abdulaziz University, Jeddah, Saudi Arabia, the M.S. degree from the University of Queensland, Brisbane, Australia, in 2006, and the Ph.D. degree from Flinders University, Adelaide, Australia, in 2015, all in computer science. He was the Vice Dean of Admission and Registration, from 2015 to 2018. He has been the Head of the Computer Science Department, Collage of Computer Science and Information Technology, Al Baha University, Al Baha, Saudi Arabia, since 2016. His areas of interest are cybersecurity, distributed and semantic computing, and data mining.

MUHAMMAD QAISER SALEEM received the first master's degree in computer science from International Islamic University, Islamabad, Pakistan, in 1998, the M.S. degree in computer science from Malardalen University, Vasteras, Sweden, in 2006, and the Ph.D. degree in information technology from Universiti Teknologi PETRONAS (UTP), Malaysia, in 2013. He has been an Assistant Professor with the Collage of Computer Science and Information Technology, Al Baha University, Al Baha, Saudi Arabia, since August 2013. During his professional career, he remained involved in academia as well as IT industry and being involved in various development projects related to databases and data warehousing environment. He is also involved in various research projects. His current research interest includes the modeling of security requirement during business process modeling in SOA application.

HANNAN BIN LIAQAT received the B.S. degree in information technology and the M.S. degree in computer networks from the COMSATS Institute of Information Technology, Lahore, Pakistan, in 2006 and 2009, respectively, and the Ph.D. degree from the Dalian University of Technology, Dalian, China, in 2016. From 2009 to 2011, he was a Lecturer with the Information Technology Department, University of Gujrat, Gujrat, Pakistan. Since 2016, he has been with the Department of Information Technology, University of Gujrat, as an Assistant Professor and a Ph.D. and M.Phil. Supervisor. He has a number of publications to his credits in international journals and conferences. His research interests include *ad hoc* social networks, the IoT, cloud computing, mobile computing, and social computing. He is a Reviewer for several international journals.

AKBER GARDEZI received the master's and Ph.D. degrees from the University of Sussex, U.K. After the completion of the Ph.D. degree, he received a Postdoctoral Tutorial Fellowship from the Department of Engineering and Design, University of Sussex, where he was closely involved with projects on embedded systems for activity recognition for a period of two years. He is currently an Assistant Professor with the Department of Computer Science, COMSATS University Islamabad, Islamabad, Pakistan. He is involved in collaboration with the UNESCO Chair in ICTs for Development and United Nations Economic and Social Commission for Asia and the Pacific. The initiative is focusing on a three-tiered approach of Advocacy, Policy Making, and Development of novel technologies for the disabled. He is also involved in the projects relating toward ICTs for Development, hence utilizing his network of peers in terms of launching advocacy and awareness initiatives for the marginalized communities within the OIC member states. His research interest includes ICTs development reinforcing the belief that technology is indeed a vehicle for betterment of society.

AZEEM IRSHAD received the master's degree from Arid Agriculture University, Rawalpindi, Pakistan. He is currently pursuing the Ph.D. degree in security for multi-server architectures with International Islamic University, Islamabad, Pakistan. He has authored more than 40 international journal and conference publications, including 23 SCI-E journal publications. His research interests include strengthening of authenticated key agreements in SIP multimedia, the IoT, WBAN, TMIS, WSN, ad hoc networks, e-health clouds, and multi-server architectures. He has served as a Reviewer for more than 27 reputed journals, including the IEEE Systems Journal, the *IEEE Communications Magazine*, the IEEE Transactions on Industrial Informatics, the *IEEE Consumer Electronics Magazine*, *Computer Networks*, *Information Sciences*, CAEE, *Cluster Computing*, AIHC, *The Journal of Supercomputing*, and *Wireless Personal Communications*, notably.

• • •