IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 27, 2017, accepted August 15, 2017, date of publication September 22, 2017, date of current version October 12, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2744798

DS-Index: Ranking Authors Distinctively in an Academic Network

MUHAMMAD FAROOQ[1], HIKMAT ULLAH KHAN[2], SAQIB IQBAL[3], EHSAN ULLAH MUNIR[2], (Senior Member, IEEE), AJMAL SHAHZAD[4]

[1]Department of Computer Science, Government College Rehmatabad, Rawalpindi 46000, Pakistan
[2]Department of Computer Science, COMSATS Institute of Information Technology, Wah Cantt 47040, Pakistan
[3]Department of Software Engineering and Computer Science, College of Engineering and Information Technology, Al-Ain University of Science and Technology, Al Ain 15258, United Arab Emirates
[4]Centre for Advanced Studies in Engineering, Islamabad 44000, Pakistan

Corresponding author: Hikmat Ullah Khan (hikmat.ullah@ciitwah.edu.pk)

ABSTRACT The impact and productivity of researchers are assessed using bibliometric parameters, such as the number of publications and citation analysis. A number of indices exist that use these parameters, but almost all of them overlook citation pattern of the researchers, which results in assigning the same index value to the two different authors with different citation patterns. In this paper, a new index called DS-index is proposed, which differentiates among the authors having even a very small change in the citation pattern of their publications. It uniquely identifies the different index values and thus the proper ranking order for authors. The index is applied to the self-developed large DBLP data set having publication data of over 50 years. The results compared with the existing indices using the standard performance evaluation measures confirm that the proposed index performs better by ranking the authors in a distinctive order.

INDEX TERMS Ranking, bibliometircs, citation analysis, academic network.

I. INTRODUCTION

To measure the scientific output and research achievement of an author's research is an important research problem in the research domain of Scientometrics. The task of the evaluation of research carried out by the researchers helps in decisions about offering a research job, granting funding for projects, and nominating for research awards. One of the basic parameters to show the research productivity of the authors is their publications count. An author's impact in the research community is usually measured by the number of citations received by his/her publications from other publications in the academic network. The relevant literature presents a number of metrics as indicators for the assessment of the worth of a researcher. The early approaches consider features such as the number of publications and the number of the articles which cite these publications. Later the focus shifted to the introduction of indices which measure the impact of a research work of an author and provides the ranking of the authors.

The different indices in the relevant literature measure the significance of research work of an author and depict it with the help of a single numeric value. Among these indexing indicators, h-index [1] is widely used because of its simplicity and easiness of computing [2], [3]. H-index has a limitation that it does not consider the citations of top h-core publications even if the highly cited publication is different for two authors but their h-index will be same. To counter this limitation, g-index [4] is proposed. The g-index considers the citations of all the h-core publications. Another limitation of the h-index as well as the g-index is that both the indices consider only h-core values and lack to consider all the publications and the research age of the author. To counter such limitations, R-index and AR-index [5] are proposed. Both these indexes are capable of showing the productivity as well as its impact on the academic community. A common limitation of all the existing indices is that they assign the same index value to two different authors with diverse citation pattern. This limitation does not help us to decide among the authors and we cannot rank them distinctively. As the existing indices lack to provide a distinctive order to authors who have a different number of publications and their publications have received a diverse pattern of citations. There is a need to devise such an index which can differentiate between the impact and productivity of two different authors with very minute differences in their citation patterns.

In this paper, we propose DS-index, which is capable of ranking authors with small differences in their research productivity and its recognition within an academic network

TABLE 1. A list of symbols used in paper.

Symbol	Quantity
A	Set of Authors
P	Set of Papers/Publications
C	Set of Citations
a	$a \in A$
p	$p \in P$
c	$c \in C$
R_i	Rank of an ith author
h_i	h-index of an ith author
g_i	g-index of an ith author
R_i	h-index of an ith author
$C_j^{a_i}$	the number of citations of an author a_i
AR_i	g-index of an ith author
$N_p^{a_i}$	The total number of publications of an author a_i
$N_c^{a_i}$	The total number of citations received by an author's publications
$N_y^{a_i}$	Research age of an author

in the form of citations. It assigns different index values and thus ranks the authors distinctively. The proposed method is compared with indices in the relevant literature and results reveals that the proposed index provides a unique ranking to all the authors. The large data set of DBLP is prepared and the results are compared with standard performance evaluation measures of O-Sim [6], Kendall correlation and Spearman correlation [7], [8]. In addition, a new performance measure, Uniqueness, measures the degree of the number of distinct values, which are found using an index. The results confirm that the proposed index performs better than the existing indices. The list of symbols used in the paper is given in Table 1 which helps in understanding the concepts discussed in the paper.

The rest of the paper is organized as follows: Section II reviews the works in the relevant literature. Section III shares the need of new index and problem statement; Section IV presents the proposed index. Section V presents the experimental setup. In this section, details of dataset used and the performance evaluation measures applied are discussed. Section VI discusses the results before concluding the paper.

II. RELATED WORK

One of the basic measures to find the productivity of the authors is their number of publications. It shows the quantity of the research work. Similarly, the number of citations received by their publications depicts the impact of the authors' work in a scholarly network [9], thus, enhancing the prestige of the journal or conference which published that research paper [10]. These basic concepts of publication count and citations count motivated the researchers that the research output should be measured by an arithmetic approach [11]. The bibliometric measures and indicators have their vast applications such as measuring the impact of an author in the community, to study the relationship between the universities and industries, and to forecast the future development trends [12]. A research study shared that more and more dimensions of complexity and quality of data are emerging raising challenges in modelling bibliographic data for bibliometric analysis of data analyzing techniques [13]. In addition to authors, research documents have also been ranked using content based analysis [14]. The link based algorithms have also been applied for ranking authors [15], [16]. Scientometrics research also focus on academic ranking of the world universities as well [17]. Recent research studies also focus on co-author ranking in heterogeneous network and the scientific ranking in heterogeneous academic hyper-networks [18], [19].

One of the first widely accepted efforts towards measuring the quality of research work of a scholar was the introduction of h-index [1]. The index is capable to depict the impact and magnitude of the research work of an author by a single value. The h-index is popular among the scientific community, as it is simple and easy to calculate [20]. It considers both the number of publications and the number of citations to measure the significance of an author in terms of h-index. H-index of an ith author a_i is calculated by organizing his/her set of publications P_i in decreasing order of their overall number of citationsC_i, then a rank R is allocated to each publication p in increasing order. The h-index of the author is the rank R_i in such that $R_i \leq C_i$ and $R_i + 1 \geq C_i$. The h-index has a limitation that if a publication is selected within the h-core publications for h-index, then further increase in its citations does affect the h-index value [21]. This problem is resolved by g-index (Egghe-2006), which takes into account the capability to credit the newly received citations by a publication. The g-index, a variation of h-index, is calculated by organizing the publications in descending order of their citation-counts, and then the cumulative sum of citations-count is calculated along with square of rank R_i of a publication. The g-index is the rank at which $(R_i)\,2$ is less than the cumulative sum of citations in such a way that the $(R_i + 1)\,2$ is greater than cumulative sum. Like h-index, g-index also does not provide satisfactory results to rank the authors distinctively [22]. To compare a young and senior author, both the h-index and the g-index fail to provide a good result. This problem is solved by m-index [23]. The m-index is calculated by dividing the h-index with the number of years of research work of an author. It is also called m-quotient. This index has a limitation that a small change in h-index causes a significant change in the value of m-quotient [24].

The h-index does not differentiate among the authors having same h-index even though they have a different citation pattern for their publications [25]. For instance, consider two authors, a^i and a^j. The author a^i has two publications having 201 and 3 citations in publication p_1^i and p_1^i respectively. On the other hand, the author a^j has two publications

having 4 and 3 citations in publication p_1^j and p_2^j respectively. There is a big difference in citation count, but their h-index is same, i.e., 2. This limitation was solved by porposing R-index [5]. The R-index value, for an author is calculated using Equation 1.

$$R_i = \sqrt{\sum_{j=1}^{h} C_j^{a_i}} \tag{1}$$

Where $N_c^{a_i}$ represents the number of citations of an author a_i. When it is required that existing research work should be given less significance and the new work should be given greater significance, then the choice is AR-index, which is calculated using Equation 2.

$$AR_i = \sqrt{\sum_{j=1}^{h} \frac{C_j^{a_i}}{N_y^{a_i}}} \tag{2}$$

Research age of an author a_i is represented by $N_y^{a_i}$ refers to the number of years from his/her first publication till the current year. The symbol h refers to the h-core publications of the author. This index is used when it is required that the most recent papers should be given more importance than old papers.

We find a number of research works which consider the role of co-authors in measuring the impact of an author. Hellsten et al. [26] proposed a technique to consider self-citation as these are useful for mining authors' patterns as co-authors in the publications. Such analysis is helpful to know about the co-authors of a paper and measure impact of their research. Ausloos [26] proposed a Scientometrics law by which a new bibliometric can be derived from the h-index of co-author. It states that a simple relationship holds between the number of joint publications J by its number of co-authors and their rank of importance r, i.e., $J \propto \frac{1}{r}$. Bougrine [27] proposed the extension of the proposed method of Ausloos and tested it on larger subfields and validated the results of the law. Miśkiewicz [28] applied the same technique proposed by Ausloos on books, articles and Wikipedia. This study revealed that the method is useful, but a little different behavior is observed. Galam [29] applied Tailor Based Allocations for multiple authors and used gh-index to find the impact of an author's research on his/her co-authors. Abbasi [30] proposed a theoretical model to explore the co-author based academic network. The proposed theory applied many network-based measures such as centrality, closeness and compared it with bibliometric like g-index. Ding [31] provided by his network based methods that only author's co- citation analysis is not enough because of its drawbacks such as its subjective nature of interpretation of results. He proposed that the factor analysis can be used with co-author analysis to avoid such problems.

The limitations of the indices are illustrated in [32]. The problem of not distinctively ranking the authors is common in all the above-mentioned indices which are addressed in this research work. A survey of such indices provides an in-depth analysis of the advantages and disadvantages of the indices in the relevant literature [33]. Another review article presents the use and application of h-index and its variants in various field [2].

TABLE 2. A real data example showing Problem of existing indices.

	h-index	g-index	R-index
Sophie Cluet	5	9	8.8
James F. O'Brien	5	9	8.8
L.Paul Chew	5	9	8.8
Gerald Tesauro	5	9	8.8
Maogang Wang	5	9	8.8

TABLE 3. The statistics of dblp dataset.

Symbol	Quantity
Publications Years	1936-2013
Number of Authors	1351586
Number of Publications	3818185
Number of Conferences	6598
Number of Journals	1403
Average Publications per year	77922
Average Publications per author	2.83

III. NEED OF THE NEW INDEX AND PROBLEM STATEMENT

As mentioned earlier, the existing indices lack to rank authors distinctively. For example, consider two scientists A and B who have written 6 papers each. The author A's publications have received 9,5,3,3,1,0 citations, whereas the author B's publications have received 7,5,5,3,1,1 citations sorted in descending order. Both the authors enjoy the same value for h-index, g-index and R-index. We cannot rank A and B distinctively. In other words, the h-index, g-index and R-index may assign the same rank to more than one authors competing for a significant position. There is a need of an index, which rank the authors distinctively.

To explain the need of a new index, let us show the indices rank values of four authors from our dataset, given in Table 2 and it is evident that the existing indices are same for five authors so we cannot properly rank these authors using these indices and a new index is required.

Formally, In a set A of authors and for given two author a_i and a_j, having a set of publications $N_p^{a_i}$, $N_p^{a_j}$ and citations, $N_c^{a_i}$ and $N_c^{a_j}$ i.e., but their academic index values is same and thus their ranks R_i and R_j are equal, i.e., $R_i = R_j$. The aim is to propose an academic index such that for two dissimilar

TABLE 4. Top ten authors index values using various indexes (ranked w.r.t ds index).

Author	Publication Count	Citation Count	H-index	G-index	R-index	DS-index
Jeffrey D. Ullman	277	3412	28	56	51.497	367.702
Michael Stonebraker	254	2605	27	46	41.557	290.646
David J. Dewitt	188	2271	24	44	40.373	276.540
Philip A. Bernstein	177	1913	22	42	39.166	246.080
Jim Gray	167	1878	18	43	40.743	220.340
David Maier	234	1613	18	38	35.566	205.557
Won Kim	223	1392	21	35	32.357	199.260
Serge Abiteboul	268	1570	19	35	31.384	195.561
Catriel Beeri	107	1273	20	34	31.048	186.010
Yehoshua Sagiv	169	1288	20	34	30.983	184.072

TABLE 5. Top ten authors index values using various indexes (ranked w.r.t ds index).

Author	Publication Count	Citation Count	H-index	G-index	R-index	DS-index
Jeffrey D. Ullman	277	3412	1	1	1	1
Michael Stonebraker	254	2605	2	2	2	2
David J. Dewitt	188	2271	3	3	4	3
Philip A. Bernstein	177	1913	4	5	5	4
Jim Gray	167	1878	8	4	3	5
David Maier	234	1613	8	6	8	6
Won Kim	223	1392	5	7	10	7
Serge Abiteboul	268	1570	7	7	11	8
Catriel Beeri	107	1273	6	8	12	9
Yehoshua Sagiv	169	1288	6	8	14	10

research record of author a_i and a_j, their rank should be unique, i.e., $R_i \neq R_j$.

IV. THE PROPOSED DS-INDEX

The proposed DS-index is an extension of g-index to provide a distinctive ranking of authors. It is calculated by computing the square root of the citation counts of all the g-core Publication of an author. The g-core publications are such publications which are considered for calculation of the g-index of that particular author. Then it cumulatively sums the square root values of citations received for each publication of the author to compute the final DS-index value for the author. The DS-index is given in Equation 3 as follows:

$$DS_i = \sum_{j=1}^{g} \sqrt{C_j^{a_i}} \tag{3}$$

Where g-represents the g-core publications of the authors and $C_j^{a_i}$ represents the number of citations of jth publication of an author a_i.

The set of symbols used in the paper is presented in Table 1 for ease of understanding.

V. EXPERIMENTAL SETUP

This section presents the data set used and the performance evaluation measures applied.

A. DATASET

For this research, we have prepared the DBLP[1] dataset. DBLP provides the data in XML files which is available free of cost for research purposes. The XML files have been crawled from the beginning of DBLP that is year 1936 till December 2013. The XML file of the data set is imported to the Oracle database by developing an application using Microsoft Visual Studio C#.NET 2012 version. The dataset statistics are provided in Table 3. A sample of part of data set records is presented as follows:

<article mdate="2002-01-03" key="persons/Codd69">
<author>E. F. Codd</author>
<title>Derivability, Redundancy and Consistency of Relations Stored in Large Data Banks.</title>
<journal>IBM Research Report, San Jose, California</journal>
<volume>RJ599</volume>

[1] http://dblp.uni-trier.de/xml/

<month>August</month>
<year>1969</year>
<cdrom>ibmTR/rj599.pdf</cdrom>
<ee>db/labs/ibm/RJ599.html</ee>
</article>

B. PERFORMANCE EVALUATION MEASURES

The results of the DS-index are compared with the existing indices using five performance evaluation measures, elaborated as follows:

1) DEGREE OF RANK UNIQUENESS

For measuring the distinctive nature of the proposed index, rank uniqueness, represented Uq, is calculated by Equation 4.

$$Uq = \frac{\textit{Total unique ranks}}{\textit{Total records under consideration}} \tag{4}$$

Where Uq ranges between 0 and 1.

2) PEARSON CORRELATION

It is used to calculate the linear association of two variables [8]. In this paper, it shows the association between the ranking orders of the existing indices and the proposed index. The symbols $\boldsymbol{R_1}$ and $\boldsymbol{R_2}$ represents the lists of two ranking results under discussion. It is calculated using Equation 5.

$$r = \frac{k\left(\sum \boldsymbol{R_1}\boldsymbol{R_2}\right) - (\boldsymbol{R_1})\left(\sum \boldsymbol{R_2}\right)}{\sqrt{[k]}} \tag{5}$$

3) SPEARMAN RANK CORRELATION

It is used to calculate the linear association of two variables [8]. In this paper, it shows the association between the ranking orders of the proposed and the existing indices. It is calculated using Equation 6.

$$r = \frac{n\left(\sum \boldsymbol{R_1}\boldsymbol{R_2}\right) - \left(\sum \boldsymbol{R_1}\right)\left(\sum \boldsymbol{R_2}\right)}{\sqrt{[k\sum \boldsymbol{R_1^2} - (\sum \boldsymbol{R_1})^2][k\sum \boldsymbol{R_2^2} - (\sum \boldsymbol{R_2})^2]}} \tag{6}$$

4) KENDALL RANK CORRELATION

Kendall Rank Correlation Coefficient is used to measure the variations in two results. Suppose there are n observation of two variables A and B. Let (a_1, b_1), $(a_2, b_2) \ldots (a_n, b_n)$ are pair of values of each A and B variables at same index. The pair of value is said to be concordant when $a_i > a_j$ and $b_i > b_j$ or $a_i < a_j$ and $b_i < b_i$. Similarly, the pair of values is said to be discordant when $a_i > a_j$ and $b_i < b_j$ or $a_i < a_j$ and $b_i > b_j$. The Kendall coefficient, represented by $\boldsymbol{\tau}$, is calculated using Equation 7, as shown at the bottom of the next page.

5) OSim

The OSim$(\boldsymbol{R_i}, \boldsymbol{R_j})$ [6] is used to calculate the overlapping between first k values of two ranking orders $\boldsymbol{R_i}$ ranking and $\boldsymbol{R_j}$ ranking. It is calculated using Equation 8.

$$\boldsymbol{OSim} = \frac{\boldsymbol{R_i} \cap \boldsymbol{R_j}}{\boldsymbol{k}} \tag{8}$$

TABLE 6. A comparison of ds-index vs existing indices using performance measures.

	Pearson Correlation	Spearman Correlation	Kendall Correlation
DS-index vs h-index	0.72	0.66	0.52
DS-index vs g-index	0.97	0.98	0.93
DS-index vs R-index	0.96	0.96	0.91

TABLE 7. OSim results among all indices.

	O-Sim
DS-Index vs h-Index	0.8
DS-Index vs g-Index	0.7
DS-Index vs R-Index	0.7

VI. RESULTS AND DISCUSSIONS

First we discuss the top ten authors ranked by DS-index, then we discuss the results with the help of performance evaluation measures. At the end, we discuss in detail the case of two authors who have the same values using all the existing indices but the proposed-index ranks each author by assigning a distinctive index value.

A. AUTHORS RANKING ANALYSIS

Let us compare the results of baseline and the proposed indices. Table 4 shows the values of the respective indices whereas the Table 5 presents their rank order. The h-index, g-index and R-index are taken as the baseline indices. The publication count represents the total number of publications of an author and citation count represents the collective sum of citations received by all the papers of an author.

It is noteworthy that the range of values in Table 4 is relatively smaller for the existing indices as compared to the proposed index. The highest DS-index value is 367.70 and the lowest DS-index value is 184.07 so the range, the difference between maximum and minimum value is 183.63. On the other hand, the range of h-index values is merely 8. The DS-index values for only top 10 authors are quite dispersed and all the resultant index values are unique.

Table 5 represents the rankings produced by each index of the authors. These rankings are illustrated against their respected indexes values. The ranks are allocated based on their values computed using the indices. The author with the highest index value is assigned rank 1, the second is assigned rank 2 and so on. When two or more authors have the same index value, then they are assigned the same rank. The results show that the first 2 authors have same rank in each indexing scheme. This shows that DS-rank not only gives the unique orders, but also the correct ranking.

The correlation between the DS-index and other ranking schemes is shown in Table 5. According to the Pearson rank correlation, the highest correlation of DS-index rank is with

TABLE 8. Top 10 authors and their ranks by each index.

Author	H-index	Author	G-index	Author	R-index	Author	DS-index
J. D. Ullman	1	J. D. Ullman	1	J. D. Ullman	1	J. D. Ullman	1
M. Stonebraker	2	M.Stonebraker	2	M. Stonebraker	2	M. Stonebraker	2
D. J. Dewitt	3	D. J. Dewitt	3	Jim Gray	3	D. J. Dewitt	3
P. A. Bernstein	4	Jim Gray	4	D. J. Dewitt	4	P. A. Bernstein	4
Won Kim	5	P. A. Bernstein	5	P. A. Bernstein	5	Jim Gray	5
M. J. Carey	6	E. F. Codd	6	R. A. Lorie	6	David Maier	6
C. Beeri	6	R. A. Lorie	6	E. F. Codd	7	Won Kim	7
Y. Sagiv	6	David Maier	6	David Maier	8	S. Abiteboul	8
R. Agrawal	6	S. Abiteboul	7	N. Goodman	9	C. Beeri	9
S. Abitebou	7	Won Kim	7	Won Kim	10	Y. Sagiv	10

TABLE 9. Uniqeness rank score for top authors.

Top Authors	h-index	g-index	R-index	DS-index
50	0.26	0.36	1	1
100	0.16	0.22	0.94	1
150	0.11	0.17	0.88	1
200	0.9	0.14	0.83	1
250	0.7	0.11	0.76	1

TABLE 10. Problem preseted in Table 2 is solved by DS-index.

	h-index	g-index	R-index	DS-index
Sophie Cluet	5	9	8.8	81.1
James F. O'Brien	5	9	8.8	78.4
L.Paul Chew	5	9	8.8	75.1
Gerald Tesauro	5	9	8.8	73.2
Maogang Wang	5	9	8.8	71.2

g-index and lowest correlation is with h-index rank. The higher values of the correlation depict again the correctness of the DS-index ranking.

Let us first share the correctness of our method and later we will discuss the uniqueness of the proposed index. The OSim results presented in Table 6 represent that the results

TABLE 11. Authors with same h-index, g-index and R-index but Different DS-index.

Author	h-index	g-index	R-index	DS-index
Ron Barber	3	8	9.1	14.4
Seth J. White	3	8	9.1	17.6
Himanshu Gupta	4	9	9.1	19.7
Yue Zhuge	4	9	9.1	22.5
Marvin H. Solomon	4	9	9.1	20.2

of the DS-index are much similar to those of other indexing schemes for top 10 authors of each index. In Table 7 the OSim values among the results of Table 6 are shown. The OSim values show that the 80% of the authors selected by DS-index and h-index are the same for top 10 values of each index regardless of their position in the list. Similarly, 70% authors selected by DS-index are same with respect to g-index and R-index.

1) DEGREE OF RANK UNIQUENESS OF INDEXES

The uniqueness measure the degree of unique ranks assigned to each author. For example, in case the value of uniqueness is 1, then it means there is 100% unique index value for all the authors. Resultantly, each author is assigned a separate rank. This helps to determine who should be given the top rank using deep analysis.

The Table 8 presents the ranking of authors using four indices. The results reveal that with the DS-index, each author is assigned a unique rank. For example, h-index assigns rank values 6 to as many as four different authors even they have different research records.

The Table 9 shows the uniqueness of rank values for different indexes by considering for top 50 to 250 authors.

$$\tau = \frac{(number\ of\ concordant\ pairs) - (number\ of\ discordant\ pairs)}{\frac{1}{2}k\,(k-1)} \tag{7}$$

TABLE 12. A comparison of two authors with different citations count.

Ron Barber		Seth J. White	
Publication	Citations Count	Publication	Citations Count
Efficient and Effective Querying by Image Content.	42	Shoring Up Persistent Applications.	56
The QBIC Project: Querying Images by Content, Using Color, Texture, and Shape.	39	A Performance Study of Alternative Object Faulting and Pointer Swizzling Strategies.	17
Interactive Outlining: An Improved Approach Using Active Contours.	3	QuickStore: A High Performance Mapped Object Store.	11
Automatic and Semiautomatic Methods for Image Annotation and Retrieval in Query by Image Content (QBIC).	0	Implementing Crash Recovery in QuickStore: A Performance Study.	3

TABLE 13. Exploring the role of co-authorship for top authors(ex-score represents exclusivity score).

Author	H-index	Ex-Score	Author	G-index	Ex-Score	Author	R-index	Ex-Score	Author	DS-index	Ex-Score
J. D. Ullman	1	147.1	J. D. Ullman	1	147.0	J. D. Ullman	1	147.1	J. D. Ullman	1	147.1
M. Stonebraker	2	101.2	M.Stonebraker	2	101.2	M. Stonebraker	2	101.2	M. Stonebraker	2	101.2
D. J. Dewitt	3	100.9	D. J. Dewitt	3	100.97	Jim Gray	3	58.8	D. J. Dewitt	3	100.9
P. A. Bernstein	4	84.5	Jim Gray	4	58.8	D. J. Dewitt	4	100.9	P. A. Bernstein	4	84.5
Won Kim	5	69.7	P. A. Bernstein	5	84.5	P. A. Bernstein	5	84.5	Jim Gray	5	58.8
M. J. Carey	6	98.4	E. F. Codd	6	5.3	R. A. Lorie	6	20.8	David Maier	6	107.4
C. Beeri	6	55.9	R. A. Lorie	6	20.8	E. F. Codd	7	5.2	Won Kim	7	69.7
Y. Sagiv	6	84.4	David Maier	6	107.4	David Maier	8	107.4	S. Abiteboul	8	134.9
R. Agrawal	6	106.2	S. Abiteboul	7	134.9	N. Goodman	9	49.6	C. Beeri	9	55.9
S. Abitebou	7	134.9	Won Kim	7	69.7	Won Kim	10	69.7	Y. Sagiv	10	84.7

The results confirm that R-index assigns unique values to top 50 authors, but the DS-index assigns a unique rank to each author to all the top authors.

The Figure 1 proves that the DS-index assigns a unique rank to each author in top 250 authors and has the capability to assign more unique ranks than other indexing schemes

2) AUTHORS WITH SAME h-INDEX, g-INDEX and R-INDEX, BUT DIFFERENT DS-INDEX

Let us have a discussion of authors who have the same values for the existing indices. In Table 10, the authors having the same value for h-index, g-index and R-index are presented. It is noteworthy that the DS-index assigns unique values to all the authors. DS-index plays significant role in differentiating these authors while the existing authors fail to differentiate using existing indices of by h-index, g-index and R-index. To further explain, the table IX presents a similar data where the authors having similar indices values are presented. In addition, the detail of citation count pattern of two authors, Ron Barber and Seth J. White from Table 9 are provided in Table 12, which clearly indicates that they have different citation-count pattern, but same h-index, g-index and R-index and different DS-index.

In Table 12, the detail of the papers and their citations of the two authors is given to show that they have a different citation count, but has same h-index, g-index and R-indexes but different DS-index.

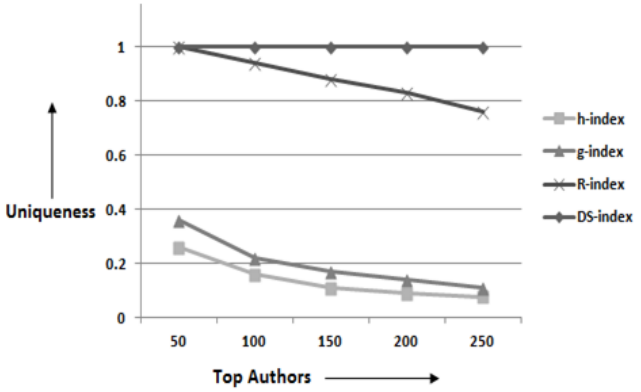

FIGURE 1. A comparison of Indices to degree of rank distinctiveness for authors.

VII. EXPLORING THE ROLE OF CO-AUTHORSHIP

In addition to the analysis of results provided by applying DS-index, we present the role of co-authors as well. In the existing literature, the weight of co-authorship is calculated by measuring exclusivity. Exclusivity is calculated to find the strength of a co-authorship relationship in the article. It represents that how exclusive a relationship of co-authors has been within a particular article. This relationship varies with the number of authors within an article. An increased number of authors within chosen article indicate less co-authorship strength. Thus, if there are a number of authors in a paper, then overall author has less influence because influence is divided among the authors. This concept is represented by using given equation 9.

$$Exclusivity_{a_p,a_q,a_r} = \frac{1}{f(p_i) - 1} \tag{9}$$

Where a_p, a_q, a_r represents the co-authors and the function $f(p_i)$ computes the number of authors in i^{th} article p. For instance, suppose an article $1(p_1)$ is written by 4 authors, article $2(p_2)$ by 3 authors, then the value of $f(p_1) = 4$, $f(p_2) = 3$. By applying the above equation, exclusive relationship is calculated for every participating co-author in all articles. For example, author a_1 is a co-author in both articles. Thus, the co-authorship strength of authors a_1 is 0.85 $(0.33 + 0.5)$. It shows that for measuring exclusivity score of each author, their scores are added for each publication.

VIII. CONCLUSION AND FUTURE WORK

This paper presents an academic index, DS-index, to rank the authors distinctively. It differentiates the authors even having same publication count, but the slight different citation pattern and outperforms the existing indices. The experiments performed on a huge dataset of DBLP confirm that the DS-index assigns a unique rank to each author and no two authors are assigned the same rank. In contrast, the existing indices assign the same ranking to more than one author. The DS-index is capable of finding the top authors having high impact as well as productivity. The citation pattern analysis confirms that DS-index is more effective and detects even the smallest differences in the impact and productivity of the authors. It can discriminate and thus select the esteemed authors of high research achievements. In the future, we would like to apply the proposed index to find the top journals, conferences and research groups and apply this index to other fields too.

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, 2005.

[2] S. Alonso, F. J. Cabrerizo, E. Herrera-Viedma, and F. Herrera, "*h*-index: A review focused in its variants, computation and standardization for different scientific fields," *J. Inform.*, vol. 3, no. 4, pp. 273–289, 2009.

[3] D.-L. Bui, T.-T. Nguyen, and Q.-T. Ha, "Measuring the influence of bloggers in their community based on the *h*-index family," in *Advanced Computational Methods for Knowledge Engineering*. Budapest, Hungary: Springer, 2014, pp. 313–324. [Online]. Available: http://dblp.uni-trier.de/rec/bibtex/conf/iccsama/BuiNH14

[4] L. Egghe, "Theory and practise of the *g*-index," *Scientometrics*, vol. 69, no. 1, pp. 131–152, 2006.

[5] J. Bihui, L. Liming, 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.

[6] T. H. Haveliwala, "Topic-sensitive PageRank: A context-sensitive ranking algorithm for Web search," *IEEE Trans. Knowl. Data Eng.*, vol. 15, no. 4, pp. 784–796, Jul./Aug. 2003.

[7] M.-T. Puth, M. Neuhäuser, and G. D. Ruxton, "Effective use of Spearman's and Kendall's correlation coefficients for association between two measured traits," *Animal Behaviour*, vol. 102, pp. 77–84, Apr. 2015.

[8] J. Hauke and T. Kossowski, "Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data," *Quaestiones Geogr.*, vol. 30, no. 2, pp. 87–93, 2011.

[9] J. Mingers, "Measuring the research contribution of management academics using the Hirsch-index," *J. Oper. Res. Soc.*, vol. 60, no. 9, pp. 1143–1153, 2009.

[10] E. Yan and Y. Ding, "Discovering author impact: A PageRank perspective," *Inf. Process. Manage.*, vol. 47, no. 1, pp. 125–134, 2011.

[11] D. Katsaros, L. Akritidis, and P. Bozanis, "The *f* index: Quantifying the impact of coterminal citations on scientists' ranking," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 60, no. 5, pp. 1051–1056, 2009.

[12] A. A. C. Teixeira and L. Mota, "A bibliometric portrait of the evolution, scientific roots and influence of the literature on university–industry links," *Scientometrics*, vol. 93, no. 3, pp. 719–743, 2012.

[13] A. Ferrara and S. Salini, "Ten challenges in modeling bibliographic data for bibliometric analysis," *Scientometrics*, vol. 93, no. 3, pp. 765–785, 2012.

[14] D. Zhou, S. A. Orshanskiy, H. Zha, and C. L. Giles, "Co-ranking authors and documents in a heterogeneous network," in *Proc. 7th IEEE Int. Conf. Data Mining (ICDM)*, Oct. 2007, pp. 739–744.

[15] Y. Ding, E. Yan, A. Frazho, and J. Caverlee, "PageRank for ranking authors in co-citation networks," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 60, no. 11, pp. 2229–2243, 2009.

[16] M. Farooq, H. U. Khan, T. A. Malik, and S. M. S. Shah, "A novel approach to rank authors in an academic network," *Int. J. Comput. Sci. Inf. Secur.*, vol. 14, no. 7, p. 617, 2016.

[17] N. C. Liu and Y. Cheng, "The academic ranking of world Universities," *Higher Edu. Eur.*, vol. 30, no. 2, pp. 127–136, Apr. 2011. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/03797720500260116

[18] L. Mookiah and W. Eberle, "Co-ranking authors in heterogeneous news networks," in *Proc. IEEE Int. Conf. Comput. Sci. Comput. Intell. (CSCI)*, Dec. 2016, pp. 1095–1100.

[19] R. Liang and X. Jiang, "Scientific ranking over heterogeneous academic hypernetwork," in *Proc. 13th AAAI Conf. Artif. Intell.*, 2016, pp. 20–26.

[20] D. Pradhan, P. S. Paul, U. Maheswari, S. Nandi, and T. Chakraborty, "C^3-index: Revisiting author's performance measure," in *Proc. 8th ACM Conf. Web Sci.*, 2016, pp. 318–319.

[21] K. Kardam, A. Kejriwal, K. Sharma, and R. Kaushal, "Ranking scholarly work based on author reputation," in *Proc. Int. Conf. Adv. Comput., Commun. Inform. (ICACCI)*, Sep. 2016, pp. 2084–2088.

[22] 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, 2010.
[23] 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.
[24] D. Yu, W. Wang, S. Zhang, W. Zhang, and R. Liu, "A multiple-link, mutually reinforced journal-ranking model to measure the prestige of journals," *Scientometrics*, vol. 111, no. 1, pp. 521–542, 2017.
[25] J. Zhou, A. Zeng, Y. Fan, and Z. Di, "Ranking scientific publications with similarity-preferential mechanism," *Scientometrics*, vol. 106, no. 2, pp. 805–816, 2016.
[26] M. Ausloos, "A scientometrics law about co-authors and their ranking: The co-author core," *Scientometrics*, vol. 95, no. 3, pp. 895–909, 2013.
[27] H. Bougrine, "Subfield effects on the core of coauthors," *Scientometrics*, vol. 98, no. 2, pp. 1047–1064, 2014.
[28] J. Mińkiewicz, "Effects of publications in proceedings on the measure of the core size of coauthors," *Phys. A, Statist. Mech. Appl.*, vol. 392, no. 20, pp. 5119–5131, 2013.
[29] S. Galam, "Tailor based allocations for multiple authorship: A fractional *gh*-index," *Scientometrics*, vol. 89, no. 1, p. 365, 2011.
[30] A. Abbasi, J. Altmann, and L. Hossain, "Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures," *J. Inform.*, vol. 5, no. 4, pp. 594–607, 2011.
[31] Y. Ding, G. Chowdhury, and S. Foo, "Mapping the intellectual structure of information retrieval studies: An author co-citation analysis, 1987–1997," *J. Inf. Sci.*, vol. 25, no. 1, pp. 67–78, 1999.
[32] R. Adler, J. P. Ewing, and P. Taylor, "Joint committee on quantitative assessment of research: Citation statistics," *Austral. Math. Soc. Gazette*, vol. 35, no. 3, pp. 88–166, 2008.
[33] M. Maabreh and I. M. Alsmadi, "A survey of impact and citation indices: Limitations and issues," *Int. J. Adv. Sci. Technol.*, vol. 40, pp. 35–53, Mar. 2012.

MUHAMMAD FAROOQ received the master's degree in computer science from the COMSATS Institute of Information Technology, Attock Campus, Pakistan. He is currently serving as a Lecturer with the Department of Computer Science, Government College, Rehmatabad, Rawalpindi, Pakistan. He has over ten year experience in teaching, software development and research in various well known educational institutes and software organizations. He is an Oracle Certified Professional. He is currently serving as a Lecturer with the Department of Computer Science, Government College, Rehmatabad, Rawalpindi, Pakistan. His research interests include software engineering, software development, and scientometrics.

HIKMAT ULLAH KHAN received the master's degree in computer science and the Ph.D. degree in computer science from International Islamic University, Islamabad. He has been an Active Researcher for the last ten years. He is currently an Assistant Professor with the Department of Computer Science, COMSATS Institute of Information Technology, Wah Cantt, Pakistan. He has authored a number of research articles in top peer-reviewed journals and international conferences. His research interests include Social web mining, Semantic Web, data science, information retrieval, and scientometrics. He is a member of the Editorial board of a number of prestigious Impact Factor Journals.

SAQIB IQBAL received the M.Sc. degree in software engineering from the Queen Marry University of London and the Ph.D. degree in software engineering from the University of Huddersfield, U.K., the M.Sc. degree in computer science from Punjab University, Lahore, Pakistan and the Ph.D. degree in software engineering from the University of Huddersfield, U.K. He is currently an Assistant Professor with the Department of Software Engineering and Computer Science, College of Engineering and Information Technology, Al-Ain University of Science and Technology, Al Ain, United Arab Emirates. His research interests include software analysis and design, aspect-oriented software development, process modeling, model-based software development, requirements engineering, and design patterns in aspect-oriented programming.

EHSAN ULLAH MUNIR received the master's degree in computer science from the Barani Institute of Information Technology, Pakistan, in 2001, the Ph.D. degree in computer science and theory from the Harbin Institute of Technology, Harbin, China in 2008. He is currently serving as an Associate Professor and Head with the Department of Computer Science, COMSATS Institute of Information Technology, Wah Cantt, Pakistan. His research interests include heterogeneous parallel and distributed computing systems (cluster grid, cloud and peer-to-peer systems), computer and wireless networks, information systems and information retrieval.

AJMAL SHAHZAD is currently a MIS Manager with the Center for Advanced Studies in Engineering, Islamabad, Pakistan. His research interests include software engineering, software development and scientometrics.