V-Index: An Index based on Consistent Researcher Productivity

Ali Daud
Department of Computer Science and Software Engineering,
International Islamic University, Islamabad, 44000,
Pakistan
ali.daud@iiu.edu.pk

S M Saleem Yasir
Department of Computer Science and Software Engineering,
International Islamic University, Islamabad, 44000,
Pakistan

Faqir Muhammad
Department of Business Administration,
Air University, Islamabad, 44000,
Pakistan

***Abstract*—In the current era, tremendous amount of scientific research work is published by thousands of researchers annually. Different methods have been proposed for researcher productivity indexing based on quantity and quality of publications. Unfortunately, none of them considered the variation among the number of citations received by a researcher for his papers. In this paper, a novel method named Variation-Index (v-index) is proposed to handle this issue. It will consider variation in number of citations received by the researcher's publications. V-index considers the consistency in citations of researcher's publication in addition to their number of publications (quantity) and number of citations (quality) for indexing. Experimental results show the effectiveness of our proposed v-index over the existing state of the art indexing methods. This citation variation enhancement is quite general and can be merged in any of the existing indexing measures with ease.**

Keywords—Indexing; Variation; V-index; Consistent Researcher Productivity

I. INTRODUCTION

Researcher's success is usually based on the papers published by him/her in different journals, conferences and workshops. A large amount of money is being invested on scientific research in advanced countries due to which competition is getting tougher every day. Currently massive amount of scientific research work is published and organizations need to evaluate researcher's work for finding suitable researchers for emerging industry requirements [7,11]. All of the scientific progress depends on the quality of research work produced by the researcher's. Scopus, Thompson ISI, Google Scholar and Microsoft Academic Search has built and maintained a large database of researchers publication in different journals, conferences and workshops with number of citations received by them. They provide h and g-indexes of researchers which are mostly used for judging researcher productivity.

Different indexing methods have been proposed and used to measure the quantity and quality of the work of researchers. In the past, impact factor (IF) [13] was considered to be the best indexing method for evaluating the journal articles. It uses the information of average number of citations received by an article published in journals. Journals with high IF were considered as more productive than of those with the lower IF. Impact Factor was limited to the journals indexing, consequently a general indexing scheme equally useful for journals, conferences and researchers named h-index [15] was proposed. H-Index considered the number of articles and citations received by those articles. One can use h-index to assess the work of individual researcher as well as group of researchers or team. Later g-index [9] was proposed, which has used the same method as h-index to calculate the impact and quantity of published work by researcher but it is more sensitive than h-index by providing more importance to researchers with highly cited papers. Number of different variants of h-index and g-index were proposed by different researchers suggesting new enhancements to the existing methods by removing their weaknesses [5,6,16,17]. H-index and g-index are also merged to get benefit from both at the same time [2]. All the existing indexing methods ignore the variation in citations of papers for researchers.

In this paper, a new method is proposed which will consider and analyze variations in citations of papers for researchers. Our proposed method named v-index finds the consistency of quality work produced by researcher based on the variation among citations received by her/his papers. The main idea is that a researcher with low or less variance will be considered more consistent in producing quality work among others with same h-index and g-index values. For example, there are researcher A and B and they both have published 10 papers and also received same number of citations say 100 for their papers. Researcher A received 10 citations for his each paper while researcher B received 90 citations for his one paper and for other 9 papers he received just 10 citations. It shows that researcher A is receiving attention of other researchers for his all publications due to consistently publishing good quality work, while on the other hand researcher B's only one paper received 90 citations. It shows that he is not consistently publishing good quality work and his only one work received high appreciation from other researchers.

The major contributions of this work are (1) highlighting the importance of consistent citations of publication for researcher productivity indexing (2) a proposal of a simple method for calculating variation among citations received by the papers of a researcher. To the best of our knowledge this is the first work which considers citations variation of papers for researcher productivity indexing.

The remaining paper is arranged as follows. In section II, existing literature about researcher productivity indexing methods is given. Section III is about consistent research productivity indexing methods as well as it explains h-index and g-index as baselines. In section IV, dataset and results and discussions are provided and section V finally concludes this work. The terms author, scientist and researcher are used interchangeably in the paper as well as in this area of research.

II. RELATED WORK

Garfield [13] proposed a method to assess the quality of work published by the journal which is known as Impact Factor (IF). Journal which has higher Impact Factor was considered to be valuable among others. Impact Factor of a journal is calculated as the average number of citations for each of the published paper gained during the previous two years. Impact Factor was used to rank the journals and it cannot be applied to individual researcher's work directly [1]. As IF was representative of whole journal and a researcher who published a paper in that journal and even his paper did not get citations will get the same IF as other researchers published in that journal whose papers got many citations.

Individuals should be indexed based on the quality and quantity of their own publications and citations received by their publications and not by the journal in which they publish. To compare and evaluate the individual research Hirsch [15] proposed h-index. In it papers are arranged in descending order according to the citations received by them. The h-index is the paper number N, equals to or less than the number of citations of respective paper and all the proceeding documents have N or fewer citations. The h-index was robust in the sense that it did not punish a researcher for the number of papers which are not cited to the ones with high citation rate [4,12]. One everlasting limitation of these indexing methods is also discussed that they cannot be used to measure the impact of a researchers awarded with Nobel Prize on their extra ordinary work [15].

During the recent years h-index is used to be the most practicing index to measure and assess the quality and quantity of work of individual researchers directly unlike IF which can measure researcher productivity indirectly through journal citations. It can be applied to journal publications as well as article appears in different conferences, but it has been found that h-index appears to be less sensitive to tackle different factors like giving more importance to highly cited papers. Consequently, Egghe [9,10] proposed new index called g-index, which gives extra weight age to highly cited papers. If publications of scientist are ranked in descending order then g-index is the largest document number such that top g publications collectively received at least g^2 citations. The g-index calculation resembles to the h-index and it makes the procedure of ranking the scientist more sensitive [8], but as both of the indices used natural number to calculate so they both have deficiency of discriminatory authority.

Both h-index and g-index ignored the career length of researcher which is discussed by Burrell [5] and an enhancement named m-quotient to existing h-index by including career length was proposed. In M-quotient the h-index value is divided by the number of years of research activity.

Later, Burrell [6] proposed a-index by saying most prolific core of scientist output can be expressed as the average number of citations of a published paper in h core. Instead of using arithmetic average to measure the central tendency of citation distributions, new method based of median named m-index was introduced [20], by discussing the extreme values effect on arithmetic average.

Another variation of g-index and h-index was presented by Kosmulski [17] known as h(2) index. Calculation of h(2) index just like original g-index, has added more sensitivity to h-index and gives importance to more cited papers like g-index. The h(2) index of a scientist is the natural number equals to h(2) such that most cited h(2) publications received at least (h(2))2 citations collectively [17]. A weakness of a-index was discussed [16], that its process involves the division by h-index which affects the result of a good researcher with higher h-index. Jin et al. [16] handled this unfair behavior of a-index and proposed new solution in the shape of r-index. In r-index instead of dividing by h-index value of a researcher, author used method of taking square root of the sum of the citations of published papers in Hirsch's core to calculate the index. Jin et al. [16] along with r-index also proposed the Ar-index which adapted the power of r-index. It considered not only the intensity of the citations of the published article but also make use of the life time of the publication, which make it more sensitive as with the passage of time index of a scientist not only increases but can be decreased.

New methods are suggested to complement existing h-index by removing the weakness of ignoring the details [18]. New idea is to create *h*-sequence and *h*-matrix of the scientist to find out rank at different scientific career time span, whereas one could also find out the original Hirsch index to that scientist in h-sequence and h-matrix. Egghe and Rousseau [11] proposed weighted h-index written as hw-index [11]. It depends on the number of citations obtained by the published papers in Hirsh core. It was presented in continuous settings and discrete. It was observed that in its continuous setting this index worked well and shows some good results, while in discrete setting some deviations from the ideal results are countered. Alonso et al. [2] tried to reduce the weaknesses of h-index and g-index by merging both of the indices. He merged the properties of both of the indices and created new index known as hg-index. The relationship of journals and researchers is discussed and an Indexing criteria by considering journal and the scientist at the time is proposed [3]. The intuition was that both entities are interrelated to each other as highly ranked journals have publications of highly ranked scientists. In the related work studied so far no one handles the problem of variations of citations of the published work of a researcher which motivated us to propose v-index. Less variation tells us about consistency, which is key attribute to judge the performance of all real world entities.

III. CONSISTENT RESEARCHER PRODUCTIVITY INDEXING

In this section, dataset, h-index and g-index is described before describing our proposed v-index. A synthetic data is taken for two scientists to show the importance of our proposed idea. In the dataset given in Table 1 and Table 2 both scientist have same number of published articles, same number of total received citations. H-index and g-index shown with highlighted rows for researcher A and B which is the same for both. Publication Rank is denoted by (PubR), citations is denoted by (Ci), Square of Publication Rank is denoted by ($PubR^2$) and cumulative citations is denoted by (CCi). Here, we selected H-Index and G-Index for comparative study as they are state-of-the-art, although our proposed method is quite flexible and can be merged in all existing indexing methods.

A. H-Index

Hirsch [15] proposed an index used to measure the productivity and impact of the published research or work of a researcher. This index uses the papers and citations relevant to each of these papers which they have gained in other researchers papers. It is defined as *"A scientist has index h if h of [his/her] N_p papers have at least h citations each, and the other ($N_p - h$) papers have at most h citations each"*. H-index covers both, the publications and the number of citations they have received and can be calculated by using the following equation:

$$h = \sqrt{\frac{N_c T}{a}} \quad (1)$$

where, NcT is the total number of citations received and "a" is the proportionality constant ranges between 3 and 5. Calculation of h-index can also be performed by the lotka power law given in [12,14].

TABLE 1. SCIENTIST A

PubR	Ci	$PubR^2$	∑Ci
1	50	1	50
2	44	4	94
3	40	9	134
4	31	16	165
5	25	25	190
6	18	36	208
7	12	49	220
8	10	64	230
9	**9**	81	239
10	5	100	244
11	4	122	248
12	1	144	249
13*	0	169	249
14*	0	196	249
15*	**0**	**225**	**249**
16*	0	256	249

TABLE 2. SCIENTIST B

PubR	Ci	$PubR^2$	CCi
1	108	1	108
2	50	4	158
3	20	9	178
4	14	16	192
5	13	25	205
6	13	36	218
7	11	49	229
8	10	64	239
9	**9**	81	248
10	1	100	249
11	0	122	249
12	0	144	249
13*	0	169	249
14*	0	196	249
15*	**0**	**225**	**249**
16*	0	256	249

B. G-Index

Egghe [9] proposed g-index also depends upon the number of citations received by the published work of individual researcher. The index is measured by counting the distribution of citations achieved by the published article of the given scientist [10]. Given is the set of published papers sorted in descending order according to the number of citations gained, and then g-index is the largest number such that the top g papers collectively have at least g^2 citations [10]. It can be calculated through Lotka function given below [9].

$$g = \left[\frac{\alpha - 1}{\alpha - 2}\right]^{\frac{\alpha-1}{\alpha}} T^{\frac{1}{\alpha}} \quad (2)$$

where, 'α' is the Lotka exponent and 'T' denotes the total number of sources (in the citation application this means the total number of ever cited papers).

C. Variation-Index (Proposed Index)

In this section, our proposed method named v-index is given, where 'V' stands for the variation among the received citations of researchers publications. The case of two scientists A and B given in Table 1 and Table 2 is considered to show that how variation in received citations of published papers play an important role to differentiate between work of researchers of same h-index and g-index value. We are dealing here with the situation in which both of the researchers have same number of published papers with same number of total citations. In this case quality of the work produced by the researchers is not differentiated by h and g indices due to their inability to handle citation variation. Our proposed v-index takes into consideration the quantity, quality and citation variation of papers altogether.

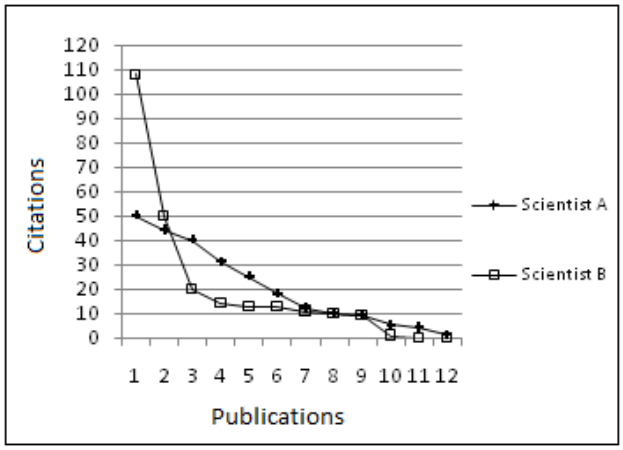

Fig. 1. Comparison between the received citations of author A and author B

Table 1 and Table 2 show that both researcher A and B have the same h and g indexes. Both of the indexes are less sensitive that they are unable to find the difference between the works of scientists by considering the variations of citations distribution. Figure 1 shows the citation variation for the work published by researcher A and B. It clearly shows that the citation variation for researcher A is less as compared to citation variation of researcher B. One can say that researcher A has a more stable graph of citations or productivity.

For adding the citation variation factor in h or g index like indexes standard deviation is used which is a commonly used method of finding variation in data, which can be calculated using Eq. 3. Standard deviation can also be calculated by taking the square root of variance[1]. Variance is calculated by taking the arithmetic mean of the square of difference of each value and the mean.

After the citation variation is calculated, v-index is obtained by simply dividing the scientist existing index value i.e. h or g indexes by the calculated standard deviation.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \mu)^2}{d-1}} \quad (3)$$

The new value of v-index shows ranking of scientist with all the good features of *h* and *g* indices along with the consistency of their quality work. In this work, the citation variation effect is added only to h and g indexes (Eq. 4) but this enhancement is very general and can be added to all other existing index as well.

$$V_{index} = \frac{h}{\sigma} or \frac{g}{\sigma} \quad (4)$$

where, *h* and *g* are the existing indices and σ is the calculated citation variation calculated through standard deviation.

Table 3 shows the calculated indexes values of scientist A and B for h, g and v indexes. It can be noticed that v-index poses more sensitivity than widely used existing *h* and *g* indices. Both of the existing indices do not consider the consistency of producing good work. There are chances that one researcher produces a paper which receives great number of citations whereas all other publications receive average or less citation count while second scientist consistently produces good work with high number of citations. Both scientists have 9 h-index and 15 g-index values and both of these indices are insensitive to find the consistency of publishing highly cited work. Table 3 shows the final results and comparison of V-index and other indices with calculated standard deviation. Our proposed method v-index values are (0.56, 0.90) for scientist A and (0.30, 0.55) for scientist B for h and g indexes values, respectively. The higher values of v-index for researcher A as compared to researcher B which for the h and g indexes, shows that scientist A is better than scientist B due to more consistency in producing quality work.

TABLE 3. H, G AND V INDEXES

Scientist	Indices (h,g)		Standard Deviation (σ)	v-index
A	H	9	16.19	0.56
	G	15	16.69	0.90
B	H	9	29.20	0.30
	G	15	27.41	0.55

IV. EXPERIMENTS

A. Dataset

Different online publication databases contain the information of researchers publications e.g. Google Scholar, Scopus and Web of Science (WoS). They also provide a facility to calculate h-index and g-index of researchers. Aforementioned online databases are maintaining their own publication data so there will be different result for the publications of same researcher for different databases and so the different *h* and *g* indexes.

We have used the Google scholar online database to extract the data of researchers by using Publish or Perish (POP) utility. Google scholar has been growing as a huge source of online publications data and it is free and easy to access. We extracted the information of more than 12 thousand researchers with their respective publications and citation records using POP. The variables used in this study are paper id, researcher names and citations of paper.

B. Results and Discussions

a. Researchers with same h-index

In this section we are discussing the situation that if two or more researchers have same h-index values. There are many researchers which share same h-index value. We have selected 5 researchers with h-index 15 as an example. Our proposed v-index is applied to rank them as well. Results in Table 4 clearly show that v-index have detected the difference between the contributions of each scientist on the basis of consistency of received citations for their publications. So that the researchers which have consistency (less standard deviation or variation) among the citations have higher v-index while h-index was unable to make any difference between these researchers.

TABLE 4. RESEARCHER WITH H-INDEX 15

Scientist	h-index	Standard Deviation (σ)	v-index
MR Azimi-Sadjadi	15	17.89	84
F Hirsch	15	22.69	66
TF Syeda-Mahmood	15	30.61	49
PK Bhattacharya	15	32.28	47
M Bhattacharya	15	62.56	24

b. Researchers with same g-index

In this section we are discussing the situation that if two or more researchers have same g-index values. There are many researchers which share same g-index value. We have selected 5 researchers with g-index 25 as an example. Our proposed v-index is applied to rank them as well. Results in Table 5 clearly show that v-index have detected the difference between the contributions of each researcher on the basis of consistency of received citations for their publications. On the other hand g-index was unable to make any difference between these researchers.

TABLE 5. RESEARCHER WITH G-INDEX 25

Scientist	g-index	Standard Deviation (σ)	v-index
L Egghe	25	11.28	222
SY Chung	25	19.63	127
R Bhattacharya	25	20.32	123
T Syeda-Mahmood	25	20.48	122
M Mahmood	25	29.29	85

c. Researchers with same h and g-indexes

Finally we consider that if two or more researchers have same h and g-index values. The number of researchers having same h and g-index is much smaller as compared to number of researchers having only similar h-index or similar g-index values. We have selected 2 researchers with same h-index 7 and g-index 25 as an example. Our proposed v-index is applied to rank them as well. Results in Table 6 clearly show that v-index have detected the difference between the contributions of both researcher on the basis of consistency of received citations for their publications. The researchers A Dev which have consistency among the citations have higher v-index as compared to DP Chakraborty which has less consistent citations for his publications. In this case both h and g indexes were unable to differentiate between the performances of these researchers.

TABLE 6. RESEARCHER WITH H-INDEX 7 AND G-INDEX 15

Scientist	H, g Indexes		Standard Deviation (σ)	v-index
A Dev	**H**	7	13.17	53
	G	15	13.17	114
DP Chakraborty	**H**	7	25.61	27
	G	15	25.61	59

V. CONCLUSIONS

Existing methods for indexing researchers or groups are not considering very important factor of their consistent productivity. The addition of consistent productivity factor in terms of citation variation of researcher publications is novel as researcher with more consistent citation record is more productive. The idea of consistent productivity is quite general and can be applied to all existing researcher productivity indexing methods by simply dividing their values by citation variation value calculated through standard deviation same as we did here in case of h and g indexes values.

The time factor normalization like m-quotient for research career length [6] can be easily performed for citation variation by considering the year wise citations variation received by the publications. As different papers of researchers will have different year wise citation variation.

ACKNOWLEDGMENT

The work is supported by Higher Education Commission (HEC), Pakistan startup research grant under Interim Placement of Fresh PhDs program 2011.

REFERENCES

[1] R. Adler, J. Ewing, and P. Taylor, “Citation Statistics. Statistical Science,” vol. 24, no. (1), pp. 1-14, 2009.

[2] 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.

[3] D. Bouyssou, and T. Marchant, ”Consistent Bibliometric rankings of authors and of journals,” Journal of Informetrics, vol. 4, pp. 365-378, 2010.

[4] T. Braun, W. Glänzel, and A. Schubert, ”A Hirsch-type index for journals,” Jointly published by Akadémiai Kiadó, Budapest Scientometrics, and Springer, Dordrecht, vol. 69, no.1, pp. 169-173, 2006.

[5] Q.L. Burrell, “Hirsch’s h-index: a stochastic model,” Journal of Informatics, vol. 1, no. 1, pp. 16-25, 2007.

[6] Q.L. Burrell, “On the h-index, the size of the Hirsch core and Jin’s A-index,” Journal of Informetrics, vol. 1, no. 2, pp. 170-177, 2007.

[7] F.J. Cabrerizo, S. Alonso, E. Herrera-Viedma, and F. Herrera, “q2-Index: Quantitative and Qualitative Evaluation Based on the Number and Impact of Papers in the Hirsch Core,” Journal of Informatics, vol. 4, no. 1, pp. 23-28, 2009.

[8] R. Costas, and M. Bordons, “Is g-index better than h-index? An exploratory study at the individual level,” Jointly published by Akadémiai Kiadó, Budapest and Springer, Dordrecht, Scientometrics, vol, 77, no. 2, pp. 267-288, 2008.

[9] L. Egghe, “Theory and Practice of the g-index,” Jointly published by Akadémiai Kiadó, Budapest and Springer, Dordrecht, Scientometrics, vol. 69, no. 1, pp. 131-152, 2006.

[10] L. Egghe, “An Improvement to H-index: The G-index,” ISSI News-Letter, vol. 2, no. 1, pp. 8-9, 2006.

[11] L. Egghe, and R. Rousseau, “An h-index weighted by citation impact” Information Processing and Management,” vol. 44, no. 2, pp. 770-780, 2008.

[12] L. Egghe, and R. Rousseau, “An informetric model for the Hirsch-index,” Jointly published by Akadémiai Kiadó, Budapest and Springer, Dordrecht. Scientometrics, vol. 69, no. 1, pp. 121-129, 2006.

[13] E. Garfield, “Impact factors and why they won’t go away,” Nature, vol. 411, no. 6837, pp. 522-522, 2001.

[14] W. Glänzel, “On the h-index – a mathematical approach to a new measure of publication activity and citation impact,” Scientometrics, vol. 67, pp. 315-321, 2006.

[15] J.E. Hirsch, “An index to quantify an individual research output” Proceedings of the National Academy of Sciences of the United States of America, vol. 102, pp. 16569-16572, 2005.

[16] B.H. Jin, L.M. Liang, R. Rousseau, and L. Egghe, “The R- and AR-indices: Complementing the h-index” Chinese Science Bulletin, vol. 52, no. 6, pp. 855-863, 2007.

[17] M. Kosmulski, ”A new hirsch-type index saves time and works equally well as the original h-index,” ISSI newsletter, vol. 2, no. 3, pp. 4-6, 2006.

[18] L. Liang, “h-index sequence and h-index matrix: Constructions and applications,” Jointly published by Academia Kiadó, Budapest and Springer, Dordrecht, Scientometrics, vol. 69, no 1, pp. 153–159, 2006.

[19] R. Rousseau, ”Simple models and the corresponding h and g-indexes,” http://hdl.handle.net/1942/944, 2006.

[20] A. Sidiropoulos, D. Katsaros, and Y. Manolopoulos, “Generalized Hirsch h-index for disclosing latent facts in citation networks,” Jointly published by Akadémiai Kiadó, Budapest and Springer, Dordrecht Scientometrics, vol. 72, no. 2, pp. 253–280, 2007.