

# AUTHOR PRODUCTIVITY INDEXING VIA TOPIC SENSITIVE WEIGHTED CITATIONS

Tehmina Amjad<sup>2</sup>, Shabnum Bibi<sup>2</sup>, Muhammad Akram Shaikh<sup>3</sup> and Ali Daud<sup>1,2\*</sup>

<sup>1</sup>Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

<sup>2</sup>International Islamic University, Islamabad, Pakistan

<sup>3</sup>Mehran University of Engineering and Technology, Jamshoro, Pakistan

[tehmnaamjad@iiu.edu.pk](mailto:tehmnaamjad@iiu.edu.pk), [akramshaikh@hotmail.com](mailto:akramshaikh@hotmail.com), [ali.daud@iiu.edu.pk](mailto:ali.daud@iiu.edu.pk)

\*Corresponding author: Dr Ali Daud, email: [ali.daud@iiu.edu.pk](mailto:ali.daud@iiu.edu.pk)

**ABSTRACT**— Different author productivity indexing methods have been proposed in order to rank scientists on the basis of their research work. The author productivity indexing methods present in literature do not consider the topic based contribution of authors for assigning them the weighted citations in a multi-authored paper. This study proposed TSWC-index which assigns Topic Sensitive Weighted Citations to authors of a paper according to their topic relatedness. Topic of co-authors in each paper against its first author has been checked and more weight is assigned to the co-authors if their topic is same as first author. The results are compared with h-index and k<sup>th</sup> rank index. Proposed method clearly shows significant difference among author's full citations score, weighted citations score and topic sensitive weighted citations score.

**Keywords**—Author productivity, Topic, citations count, topic based citations.

## I. INTRODUCTION

In scholarly networks, methods are required to find out the prominent researchers, to measure the performance of individuals in a collaborative task, and to rank journals and conferences. Different methods have been proposed in literature for evaluating the scientific performance of individuals, comparing researchers from same and different fields and their ranking. Almost all of techniques consider number of papers published by researcher and total number of citations received by those papers to evaluate a scientist research performance. The credit of received citations goes to all co-authors of a multi-authored paper. The average number of authors on scientific papers is increasing because complicated problems need more different subspecialties [1]. In case of multi-authored papers, some techniques are required to assign them credit according to their contributions. The ordering of the co-authors names is usually done on the basis of their contribution in a paper and trend of alphabetical ordering is reducing over time [2]. A change of counting methods modifies co-authorship and citation impact patterns [3]. While working in collaboration, the researchers influence the each other, this impact of influence is stronger if co-authors are senior [4,5].

The weighted criteria of contributions do not assign weights to the researchers according to their relatedness to that topic. Topic sensitive weighted citation means that weighted citations are assigned to authors of a paper according to their topic relatedness. Main contribution of this research is to assign Topic Sensitive Weighted Citations to authors in multi-authored papers. We have assigned weighted citations to authors of a paper by considering topic sensitivity as a key factor for evaluating researcher's work. The results of proposed index have been compared with our baseline methods showing significant improvement.

One of the well-known indexing methods named h-index [6] was proposed in 2005 that is a single valued index, used for evaluating the scientific performance of researchers. It measures the total number of papers and total number of citations received by those papers. H-index was insensitive towards highly cited papers [7,8] so g-index [7,9] and h(2)-index [8] were proposed later, which were an enhancement of

h-index, to remove its limitation of insensitivity towards highly cited papers. Different variations of h-index and g-index were proposed later to overcome some of their limitations and add improvements like A-index [10], R and AR-indices [11,12] m-index [13], e-index [14], k and w [15] etc. Flaw of these author productivity indexing methods is that they assign the total number of citations of a paper to each of its author in case of multi-authored paper, even when contribution of all authors in a paper is not same. To remove this limitation, some techniques were proposed that consider number of collaborators that worked together and assigned them credit according to their contributions (by considering different criteria) like hI-index [16], fractional h and g indices [17], fractional counting of citations [18], hp-Index [19], hap-index [20], hm-index [21], [22], harmonic h-index [23], k<sup>th</sup>-rank [24], w [25], gm-index [26], h-index [27], CCA h and g indices [28], hmc [29], k-norm and w-norm [15] etc. Some techniques were proposed to consider researcher's career length like m-quotient [30]. WL- index was proposed to evaluate authors by considering the frequency with which individual citations are mentioned in an article [31]. A topic based method was proposed for ranking of authors in a heterogeneous networks considering the impact of authors, papers and journals simultaneously [32]. Some indices based on the combination of existing indices like hg-index [33] and q2-index [34] were proposed to keep advantages of them collectively and remove their disadvantages. To our best knowledge, we are the first to quantify the weight of received citations of authors with respect to their topic in multi-authored papers.

Rest of the paper is organized as follows: section II provides the details of proposed method, section III includes experimental details and discussion about results and section IV concludes the paper.

## II. MATERIAL AND METHODS

Existing indexing methods discussed in literature, do not cover topic based weighted citations for authors in multi-authored papers. We propose two methods of assigning weight to authors. First is Normalized Weighted Citations

(NWC) that assigns weighted citations to authors of a paper. Second method is Topic Sensitive Weighted Citations (TSWC) that increases or decreases NWC score of authors on the basis of their topics. The topic of all co-authors of a paper is compared with the topic of first author. If topic of a co-author is same as that of the first author then his/her NWC score is maximized, otherwise, NWC score is reduced. We calculated NWC-index and TSWC-index, and compared the results with h-index and  $k^{\text{th}}$ -rank Index. Before calculating NWC and TSWC, we need to find the topics of interest of all authors in dataset. For that purpose, we used titles of published articles of authors and calculated their topic probabilities by using Latent Dirichlet Allocation (LDA). Hence, we divided the authors into 100 soft clusters, a topic that shows maximum probability for an author is considered to be his/her topic of interest. We follow the following steps to calculate the TSWC-index of an author.

1) Calculate Normalized Weighted Citations (NWC) score of each author in each paper as follows:

$$NWC_i = \frac{N-i+1}{\sum_{i=1}^n i} \times cit \quad (1)$$

Where  $i$  is the rank of author in a paper and  $N$  is the total number of authors of that paper.  $Cit$  is the total citations received by that paper.

2) Check topic of authors in each paper with its first author's topic.

a) If the topic of co-author is same as first author,  $NWC_i$  is calculated as in step 1

b) Else  $NWC_i = NWC_i / 2$

3) Calculate weights of authors who do not have the same topic as follows:

$$NWC_j = \frac{N-j+1}{\sum_{i=1}^n i-2} \times NWC \quad (2)$$

Where  $j$  is rank of same topic author and  $NWC$  is the value calculated in step 2\_b).

4) Calculate Topic Sensitive Weighted Citations (TSWC) of authors having same topic by adding the value calculated in step 3 to step 2\_a) for that author. Use the value calculated in step 2\_b) as TSWC of authors with different topic.

5) Calculate NWC-index and TSWC-index of each author as that of h-index.

### III. RESULTS AND DISCUSSION

Experimentation was performed on version 5 of DBLP-Citation-network dataset taken from arnetminer.org with 127,410 authors and 100,000 papers. We preprocessed the titles by removing stop words, punctuations and numbers to get correct results. Two baseline methods, h-index and  $k^{\text{th}}$ -rank index were implemented for the purpose of comparison and evaluation of results. H-index does not consider the

individual contribution of authors in papers and topic sensitivity for assigning weighted citations based on their individual contribution. In  $k^{\text{th}}$ -rank index, first author of a paper always receives all citations of that paper as in case of h-index, while others receive less weight of citations depending upon their position in co-authors list.

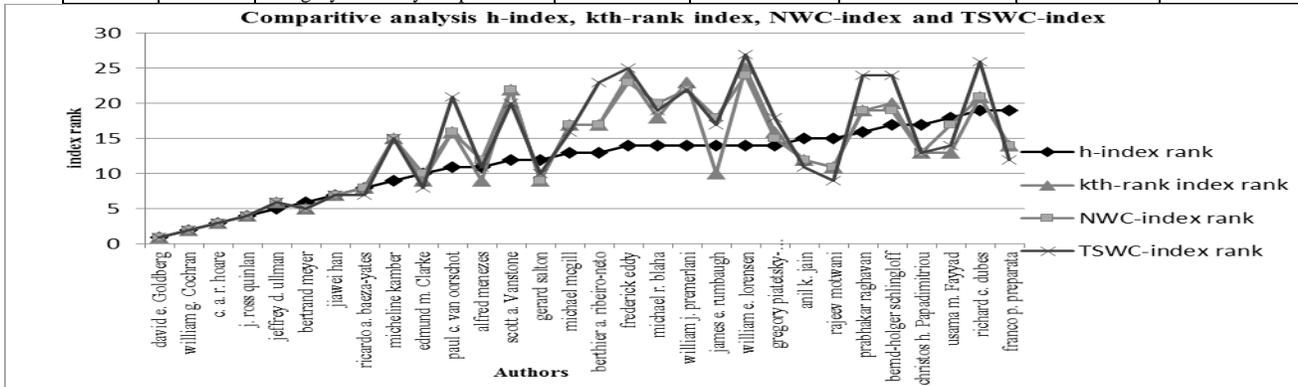
In this section, we compare the results of the proposed indices, NWC-index and TSWC-index, with the baseline methods. Table 1 shows the rank of authors after calculating TSWC score by checking the topics of authors in each paper i.e. for dissimilar topic author, his/her NWC score has been decreased and for same topic author, his/her NWC score has been increased which is shown in TSWC column. TSWC-index is then calculated for all authors. The citations of authors e.g. Ricardo A. Baeza-yates, Jiawei Han and Edmund M Clarke etc. were increased in TSWC because their topic of interest was same as that of the first authors of their co-authored papers. William G. Cochran and J. Ross Quinlan have same citations score as they gained in case of h-index, because they are the single authors of their papers. Jeffrey D. Ullman's citations score in TSWC was decreased because his topic of interest was not same as that of first author. Figure 1 shows the comparison of h-index,  $k^{\text{th}}$ -rank index, NWC-index and TSWC-index ranks of authors. The authors who have more number of single author papers have the NWC-index and in case of co-authored papers the authors receive higher rank if their topic in co-authored papers is same as first author, like Gerard Salton and Edmund M. Clarke, otherwise their rank is reduced like, Paul C. Van Oorschot and Michael McGill.

Both  $k^{\text{th}}$ -rank index and TSWC index assign weighted citations to authors but the first author in multi-authored paper receives full citations of the paper in  $k^{\text{th}}$ -rank index and other authors receive citations according to their rank. However, the TSWC-index divides the total citations among authors of a paper using the criteria of topic sensitivity. The TSWC-index of an author may be either less than or equal to  $k^{\text{th}}$ -rank index depending on the situation that the author has same topic as first author or not and author is single author of a paper or not. Authors rank in calculating TSWC-index may be less, greater or equal to the  $k^{\text{th}}$ -rank index. The TSWC-index of the authors in table 2 was decreased because the actual citations of paper were divided among its co-authors. The TSWC-index rank of these authors was increased as compared to the  $k^{\text{th}}$ -rank. Rajeev Motwani's TSWC-index rank was increased by 2 and the other authors ranks were also increased.

In the table 3, rank of all authors was decreased in TSWC-index. Gregory Piatetsky-shapiro and Paul C. Van Oorschot have same  $k^{\text{th}}$ -rank of 16 and  $k^{\text{th}}$ -rank index of 33 but their TSWC-index was decreased to 24 and 20 and their ranks to 18 and 21 respectively because their citations were decreased when weighted topically. All authors in table 4 maintain same rank in both methods. First five authors have same  $k^{\text{th}}$ -rank index and TSWC-index while the subsequent authors have less TSWC-index as compare to the  $k^{\text{th}}$ -rank index.

**Table 1. Rank of authors by their TSWC-index**

S.NO	Rank	Authors	NWC	TSWC	TSWC-index	h-index rank	kth-rank index rank
1	1	David E. Goldberg	36085	36084	94	1	1
2	2	William G. Cochran	34194	34194	92	2	2
3	3	C.A. R. Hoare	28227	28227	84	3	3
4	4	J. Ross Quinan	16859	16859	64	4	4
5	5	Bertrand Meyer	14845	14834	60	6	5
6	6	Jeffrey d. Ullman	13175	12878	56	5	6
7	7	Ricardo A. Baeza-yates	6311	7292	42	8	8
8	7	Jiawei Han	7095	7110	42	7	7
9	8	Edmund M. Clarke	5469	6494	40	10	9
10	9	Rajeev Motwani	4875	6074	38	15	11
11	10	Gerard Salton	5686	5686	37	12	9
12	11	Alfred Menezes	4358	5311	36	11	9
13	11	Anil k. Jain	4386	5282	36	15	12
14	12	Franco P. Preparata	3987	4981	35	19	14
15	13	Christos H. Papadimitriou	4231	4813	34	17	13
16	14	Usama M. Fayyad	2872	3483	29	18	13
17	15	Micheline Kamber	3363	3363	28	9	15
18	16	Michael McGill	2813	2813	26	13	17
19	17	James E. Rumbaugh	2655	2655	25	14	10
20	18	Gregory Piatetsky-shapiro	3201	2403	24	14	16



**Fig. 1. Comparison of h-index, kth-rank index, NWC-index and TSWC-index**

**Table 2. Position relocation with respect to kth-rank index: Position up**

S.NO	Authors	Kth-rank	Kth-rank index	TSWC-index	TSWC-index Rank	Earned position in TSWC-index
1	Ricardo A. Baeza-yates	8	48	42	7	+1
2	Edmund M. Clarke	9	46	40	8	+1
3	Rajeev Motwani	11	42	38	9	+2
4	Anil K. Jain	12	40	36	11	+1
5	Franco P. Preparata	14	38	35	12	+2
6	Michael McGill	17	32	26	16	+1
7	Scott A. Vanstone	22	26	21	20	+2
8	William J. Premerlani	23	25	19	22	+1

**Table 3. Position relocation with respect to kth-rank index: Position down**

S.NO	Authors	Kth-rank	Kth-rank index	TSWC-index	TSWC-index Rank	Position down in TSWC-index
1	Alfred Menezes	9	46	36	11	-2
2	Gerard Salton	9	46	37	10	-1
3	James E. Rumbaugh	10	44	25	17	-7
4	Usama M. Fayyad	13	39	29	14	-1
5	Gregory Piatetsky-shapiro	16	33	24	18	-2
6	Paul C. Van Orschoot	16	33	20	21	-5
7	Berthier A. Ribeiro-neto	17	32	18	23	-6
8	Michael R. Blaha	18	31	23	19	-1
9	Prabhakar Raghavan	19	30	17	24	-5
10	Bernd-holger Schlingloff	20	29	17	24	-4
11	Richard c. Dubes	21	27	15	26	-5
12	Frederick Eddy	24	22	16	25	-1
13	William E. Lorensen	25	19	11	27	-2

**Table 4. Position stable with respect to kth-index**

S.NO	Authors	Kth-rank	Kth-rank index	TSWC-index	TSWC-index Rank
------	---------	----------	----------------	------------	-----------------

1	David E. Goldberg	1	94	94	1
2	William G. Cochran	2	92	92	2
3	C. A. R. Hoare	3	84	84	3
4	J. Ross Quinlan	4	64	64	4
5	Bertrand Meyer	5	60	60	5
6	Jeffrey D. Ullman	6	58	56	6
7	Jiawei Han	7	51	42	7
8	Christos H. Papadimitriou	13	39	34	13
9	Micheline Kamber	15	35	28	15

## CONCLUSION

This study shows the significance of considering topics of co-authors when weighted citations are assigned to them in multi-authored papers. To evaluate scientists according to their topic based contribution, we proposed two indices: (1) NWC-index and (2) TSWC-index. Calculations of NWC-index are similar to calculations of h-index, once the Normalized Weighted Citations (NWC) score of authors in multi-authored papers according to their rank is allocated. Topic of each author was checked in each paper against its first author. The authors who have same topic as that of first author, their NWC score was increased. In case when co-author's topic was not same as that of first author, the NWC score assigned to them was decreased. The results were compared with the traditional h-index and  $k^{\text{th}}$  rank index. Variations in the ranked results were observed when weight of citations was assigned with consideration of topic of the authors. Results also show the effect on ranking of authors and variations in indices with respect to  $k^{\text{th}}$ -rank index and h-index. Our analysis shows that an author with single-authored papers has got the full citations score and his/her NWC-index and TSWC-index have same with h-index and  $k^{\text{th}}$ -rank index. In future, we intend to index the authors while considering that all co-authors of a paper have contributed equally. Another future work is to assign the coauthor's weights according to their correlation of topics with first author. If a coauthor's topic is closely correlated with first author topic then his/her weight should be minimized by a smaller amount and if his/her topic is hardly correlated with first author's topic then the weight should be minimized.

## REFERENCES

- [1] D. Kennedy, "Multiple authors, multiple problems," 2003.
- [2] L. Waltman, "An empirical analysis of the use of alphabetical authorship in scientific publishing," *Journal of Informetrics*, vol. 6, no. 4, pp. 700–711, 2012.
- [3] A. Perianes-Rodríguez and J. Ruiz-Castillo, "Multiplicative versus fractional counting methods for co-authored publications. The case of the 500 universities in the Leiden Ranking," *Journal of Informetrics*, vol. 9, no. 4, pp. 974–989, 2015.
- [4] T. Amjad, A. Daud, D. Che, and A. Akram, "MuICE: Mutual Influence and Citation Exclusivity Author Rank," *Information Processing & Management*, 2015.
- [5] T. AMJAD, A. DAUD, and A. AKRAM, "Mutual Influence based Ranking of Authors."
- [6] J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 46, pp. 16569–16572, 2005.
- [7] L. Egghe, "Theory and practise of the g-index," *Scientometrics*, vol. 69, no. 1, pp. 131–152, 2006.
- [8] 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.
- [9] L. Egghe, "An improvement of the H-index: the G-index," *ISSI newsletter*, vol. 2, no. 1, pp. 8–9, 2006.
- [10] B. Jin, "H-index: an evaluation indicator proposed by scientist," *Science Focus*, vol. 1, no. 1, pp. 8–9, 2006.
- [11] B. Jin, "The AR-index: complementing the h-index," *ISSI newsletter*, vol. 3, no. 1, p. 6, 2007.
- [12] B. Jin, L. 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.
- [13] L. Bornmann, R. Mutz, and H.-D. Daniel, "Are there better indices for evaluation purposes than the h index? A comparison of nine different variants of the h index using data from biomedicine," *Journal of the American Society for Information Science and Technology*, vol. 59, no. 5, pp. 830–837, 2008.
- [14] C.-T. Zhang, "The e-index, complementing the h-index for excess citations," *PLoS One*, vol. 4, no. 5, p. e5429, 2009.
- [15] G. Anania and A. Caruso, "Two simple new bibliometric indexes to better evaluate research in disciplines where publications typically receive less citations," *Scientometrics*, vol. 96, no. 2, pp. 617–631, 2013.
- [16] P. D. Batista, M. G. Campiteli, and O. Kinouchi, "Is it possible to compare researchers with different scientific interests?," *Scientometrics*, vol. 68, no. 1, pp. 179–189, 2006.
- [17] L. Egghe, "Mathematical theory of the h-and g-index in case of fractional counting of authorship," *Journal of the American Society for Information Science and Technology*, vol. 59, no. 10, pp. 1608–1616, 2008.
- [18] D. Bouyssou and T. Marchant, "Ranking authors using fractional counting of citations: An axiomatic approach," *Journal of Informetrics*, vol. 10, no. 1, pp. 183–199, 2016.
- [19] J.-K. Wan, P.-H. Hua, and R. Rousseau, "The pure h-index: calculating an author's h-index by taking co-authors into account," *COLLNET Journal of Scientometrics and Information Management*, vol. 1, no. 2, pp. 1–5, 2007.
- [20] J. C. Chai, P. H. Hua, R. Rousseau, and J. K. Wan, "The adapted pure h-index," *Proceedings of WIS*, 2008.

- [21] M. Schreiber, "To share the fame in a fair way, hm modifies h for multi-authored manuscripts," *New Journal of Physics*, vol. 10, no. 4, p. 40201, 2008.
- [22] M. Schreiber, "A modification of the h-index: The h m-index accounts for multi-authored manuscripts," *Journal of Informetrics*, vol. 2, no. 3, pp. 211–216, 2008.
- [23] N. T. Hagen, "Harmonic allocation of authorship credit: Source-level correction of bibliometric bias assures accurate publication and citation analysis," *PLoS One*, vol. 3, no. 12, p. e4021, 2008.
- [24] C. H. Sekercioglu, "Quantifying coauthor contributions," *Science*, vol. 322, no. 5900, p. 371, 2008.
- [25] C.-T. Zhang, "A proposal for calculating weighted citations based on author rank," *EMBO reports*, vol. 10, no. 5, pp. 416–417, 2009.
- [26] M. Schreiber, "Fractionalized counting of publications for the g-Index," *Journal of the American Society for Information Science and Technology*, vol. 60, no. 10, pp. 2145–2150, 2009.
- [27] J. E. Hirsch, "An index to quantify an individual's scientific research output that takes into account the effect of multiple coauthorship," *Scientometrics*, vol. 85, no. 3, pp. 741–754, 2010.
- [28] X. Z. Liu and H. Fang, "Fairly sharing the credit of multi-authored papers and its application in the modification of h-index and g-index," *Scientometrics*, vol. 91, no. 1, pp. 37–49, 2011.
- [29] X. Z. Liu and H. Fang, "Modifying h-index by allocating credit of multi-authored papers whose author names rank based on contribution," *Journal of Informetrics*, vol. 6, no. 4, pp. 557–565, 2012.
- [30] Q. L. Burrell, "Hirsch's h-index: A stochastic model," *Journal of Informetrics*, vol. 1, no. 1, pp. 16–25, 2007.
- [31] X. Wan and F. Liu, "WL-index: Leveraging citation mention number to quantify an individual's scientific impact," *Journal of the Association for Information Science and Technology*, vol. 65, no. 12, pp. 2509–2517, 2014.
- [32] T. Amjad, Y. Ding, A. Daud, J. Xu, and V. Malic, "Topic-based heterogeneous rank," *Scientometrics*, vol. 104, no. 1, pp. 313–334, 2015.
- [33] S. Alonso, F. 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, 2009.
- [34] 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 Informetrics*, vol. 4, no. 1, pp. 23–28, 2010.