

# Improving RA-index by Using the Weighting Mechanism Number of Citations to Filter "Spike" Signal of the Citation Data of Indonesian Authors

Adian Fatchur Rochim<sup>1</sup> and Riri Fitri Sari<sup>2</sup>

<sup>1</sup> *adian@undip.ac.id*

Departement of Computer Engineering, Diponegoro University, Semarang, 50275 (Indonesia)

<sup>2</sup> *riri@ui.ac.id*

Department of Electrical Engineering, Faculty of Engineering, Universitas Indonesia, Depok 16424 (Indonesia)

## Introduction

Number of citations and the number of papers were combines as H-index (Hirsch, 2005). H-index is an index to figure the profile of the authors. H-index is a well-known index that is used by the database indexers such as Clarivate Analytics, Scopus and Google Scholar. From the literature we found some weaknesses of H-index, including 1) the productive and perfectionist researcher were not accommodated by the H-index (Mesiar, et.al, 2016), 2) self-citation was calculated, 3) the citation weight of the main researcher is considered equal to other researchers, and the frequency of citation in a paper has not been considered (Bai et al., 2018) (Mesiar, et.al, 2016) (Gagolewski, et.al, 2009) (Zhu, et.al, 2015). Many H-index improvement proposals have been made. This includes the proposal of Egghe in 2006, which accommodates the impact value of perfectionist researchers (Egghe, 2006). Improvement and new indicator to measure the impact of researchers was needed for a better evaluation. Rochim, et.al. in 2018, proposed the RA-index as an alternative indicator of fairer-based bibliometrics to measure the impact of researchers (Rochim, et.al., 2018). Glanzel in 2016, stated that it is important to consider some methods and models to accommodate the needs (Glanzel, et.al, 2016).

This poster proposed an initial work to weighting mechanism and to filter the "spike" of citation. Subsequently, the filter is applied, and the result of citation data is calculated by the RA-index. RA-index is a more fairness-concerned variant of H-index (Rochim et al., 2018) (Rochim, et.al, 2017).

This work to measure and to differentiate of two authors with the same H-index value using the weighting citations and RA-index method. We investigate the phenomenon of the "spike" of the number of citation, and the initial solution to prevent/filter impact of the cartels/citation circle. The "spike" of citation phenomenon is the raise of the number of sudden citations within a short period of time, which is obtained from co-authors of multiple papers. Cartels/citation circle can be

defined as follows: 1) The activity of an author that act as also a reviewer for multiple papers at the same time and a joint-work among friends in a peer review ring to increase the record of papers and citation numbers (Gamboa, 2014), and 2) The activity of an author cite his/her friend's papers, and at the same time these friends also cite the author's papers (Witold Kienc, 2015). Tscharntke in 2007 classified the weighting for each author in a publication text into four weighting methods groups. The four groups are: 1) Sequence-determining-credit (SDC), 2) Equal Contribution (EQ), 3) First-author-emphasis (FLAE) and 4) Percent-contributed-percentage (PCI) (Tscharntke, 2007). In 2018, we have identified that a small number of Indonesian researchers conducted some activities of "citation circles" to increase their H-index values. "citation circle" is an activity in which someone cites the work of his friends, and will get a citation for the same way (Witold Kienc, 2015). This is a part of the "black hat" technique. The technique is not accepted or illegal for academics.

## Methodology

In order to prevent the activity of "creating citation circle", we recommend the weighting mechanism for the citation data. The citation data is weighted before it will be calculated by the RA-index method. This weighting mechanism is proposed to give an appreciate the first author and the corresponding authors. The corresponding author is normally the supervisor of the author. The proposed method accommodates the regulations of the Indonesian Government in granting credits for scientific publications. The method of the weighting mechanism is based on the combination of PCI and EC methods. For example, one paper has ten citations, and written by four authors i.e. main author (1), corresponding author (1) and other authors (2). The citation calculation obtained by each author is different and based on the following proportions as follows. The main author and correspondent get the maximum publication index value of 100% of the publication index value.

*author's publication index value =  $ma \times 100\%$  (1)*  
*ma value = the number of total citations of a paper.*

$$\text{others author's citation value} = (ma \times 50\%) / n \quad (2)$$

$n$  = total number of the others author.

Others author get a value of 50% of the maximum value divided by the number of other authors. We called the combination of the PCI and EC methods as the *maximization* or *ma* method for a weighting number of citations of author. After weighting, is done, the data were calculated by the RA-index method. The combination method of the *ma* method and the RA-index method is then called the RAMA-index method. In the previous work, a model for calculating the impact of researchers has been done using the RA-index method (Rochim et al., 2018). The following is the RA-index equation. The following is the example of the calculation process: Two authors are A and B. Author A has an H-index value of 10 and Author B of 12. Figure 1. shows the number of citations of the author each years.

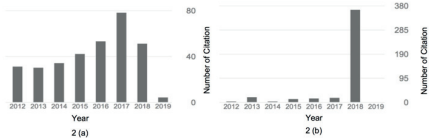


Figure 1. Comparison of the two authors.

Based on the visual observations, it appears that the number of papers and the number of citations of researcher A is greater than that of researcher B. However, the H-index value of researcher A is smaller than that of researcher B. The area of the number of citations of the researcher A is greater than researcher B in the period 2012 to 2019. Researcher B only had a large citation in 2018. For this reason, we need to filter the "spike" citation data of the author.

Table 1. Comparison the number of papers, citation and number of citation after weighted.

Researcher A					Researcher B				
No. of Papers	No. of Citation	Number Authors	Author Position	Number of Citation	No. of Papers	No. of Citation	Number of Authors	Author Position	Number of Citation after Weighted
1	15	3	*	25	1	40	10	*	2
2	74	3	1	74	2	32	14	*	1
3	33	5	*	3	3	25	10	*	1
4	23	3	*	4	4	25	24	*	1
5	13	6	1	13	5	22	5	*	2
6	12	6	*	1	6	18	6	*	2
7	12	2	*	3	7	16	14	*	0
8	10	5	*	1	8	15	10	*	1
9	10	5	1	10	9	14	5	*	1
10	10	5	*	1	10	14	2	1	14

\*: author is a co-author.

## Discussion

Table 1. shows the comparison of the authors based on number of citations, number of papers and authors position. In the initial investigation we used 10 highest citation of the authors' papers.

Table 2. shows the results of the H-index calculation and RAMA-index calculation. Researcher A gets the RAMA-index 11 and Researcher B gets the RAMA index value 1. The calculation results show that the RAMA-index method is able to filter out citation data that is not actually the impact of the researcher as a main author, so that it is not included in the calculation of the researchers' impact.

Table 2. Comparison of the H-index and RAMA-index value of the authors.

Researcher A		Researcher B	
H-index	RAMA-index	H-index	RAMA-index
10	11	10	1

## Conclusion

This work is a research in progress. The model was proposed to give weighting value of Indonesian researcher citation. The citation data was filtered by the combination of PCI and EC methods. The initial investigation shows that the weighting mechanism of the *maximization* can be used to filter the "spike" citation of the authors. We will continue to optimize the weighted values of the co-authors. In the future, it is planned to test the model by the real data of the Indonesian researchers will be conducted.

## References

- BiHui, J. (2007). The R- and AR-indices: Complementing the h-index. *Chinese Science Bulletin*, 52(6), 855–863. <https://doi.org/10.1007/s11434-007-0145-9>
- Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, 69(1), 131–152. <https://doi.org/10.1007/s11192-006-0144-7>
- Gagolewski, M., & Grzegorzewski, P. (2009). A geometric approach to the construction of scientific impact indices. *Scientometrics*, 81(3), 617–634. <https://doi.org/10.1007/s11192-008-2253-y>
- Gamboa, C. (2014). SAGE statement on Journal of Vibration and Control. Retrieved February 6, 2019, from <https://uk.sagepub.com/en-gb/asi/press/sage-statement-on-journal-of-vibration-and-control>
- Glänzel, W., Thijs, B., & Debackere, K. (2016). Productivity, performance, efficiency, impact—What do we measure anyway? *Journal of Informetrics*, 10(2), 658–660. <https://doi.org/10.1016/j.joi.2016.04.008>
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proc Natl Acad Sci U S A*, 102(46), 16569–16572. <https://doi.org/10.1073/pnas.0507655102>
- Mesiar, R., & Gagolewski, M. (2016). H-index and Other Sugeno Integrals: Some Defects and Their Compensation. *IEEE Transactions on Fuzzy Systems*, 6706(c), 1–1. <https://doi.org/10.1109/TFUZZ.2016.2516579>
- Rochim, A. F., Muis, A., & Sari, R. F. (2017). Discrimination Measurement Method on H-index and G-index Using Jain's Fairness Index. *ISSI 2017 - 16th International Conference Scientometrics and Informetrics*, 446–447.
- Rochim, A. F., Muis, A., & Sari, R. F. (2018). Improving Fairness of H-index: RA-index. *Journal of Library and Information Technology*, 38(6), 378–386.
- Witold Kienc. (2015). Should you care about your h-index, and if so, how to improve it? | Open Science. Retrieved October 11, 2018, from <https://openscience.com/should-you-care-about-your-h-index-and-if-so-how-to-improve-it/>
- Zhu, X., Turney, P., Lemire, D., & Vellino, A. (2015). Measuring academic influence: Not all citations are equal. *Journal of the Association for Information Science and Technology*, 66(2), 408–427. <https://doi.org/10.1002/asi.23179>