

## Counting Research Groups for Citation Assessment

Nakatoh, Tetsuya

Research Institute for Information Technology, Kyushu University

Hirokawa, Sachio

Research Institute for Information Technology, Kyushu University

<https://doi.org/10.5109/2230675>

---

出版情報 : Proceedings of Toward Effective Support for Academic Information Search Workshop. 1, pp.31-35, 2018-11-22. 九州大学大学院システム情報科学研究所

バージョン :

権利関係 :

# Counting Research Groups for Citation Assessment

Tetsuya Nakatoh<sup>[0000-0001-9299-7757]</sup> and Sachio Hirokawa<sup>[0000-0002-8050-1109]</sup>

Research Institute for Information Technology, Kyushu University,  
744 Motooka Nishi-ku Fukuoka 819-0395, JAPAN.  
{nakatoh,hirokawa}@cc.kyushu-u.ac.jp

**Abstract.** Assessment of a paper with respect to other research and researchers is one of the key issues in bibliometrics. In this paper, we propose a new measure “citation count by group (CCG)” to analyze the influence of a paper by considering researchers with co-authored papers as members of the same group. We collected 7,126 articles related to bibliometrics and analyzed how the conventional measure of citation count (CC) and the citation count by group (CCG) differ.

**Keywords:** Citation Count, Bibliometrics, Research Group

## 1 Introduction

In research activities, it is very important to find suitable papers. Many scientific publication databases have a function that lists bibliographic information of related articles by searching using query keywords. However, it is not easy to objectively evaluate the importance of these papers. As a substitute for objective evaluation, evaluation by the citation count of an article is generally used. For example, in Scopus, you can check the citation count of each paper obtained after searching for the paper with a query. In other words, the citation count of papers is heavily used as a more direct evaluation index of a thesis. In fact, Martin [5] reported that, as a result of a questionnaire survey on the evaluation criteria, the citation count gained much support. However, Kostoff et al. [4] showed that there are several problems in using the citation count as the criterion for research evaluation. As a solution to that problem, Nakatoh et al. [7] proposed evaluation indices to give more appropriate paper search results by using citation count limited to search keywords.

Journal Impact Factor (JIF) [2,3] is one of the most popular evaluation measures for scientific journals. Thomson Reuters updates and provides the score of journals in Journal Citation Reports (JCR) every year. The JIF of a journal describes the average of citation counts of articles published in the journal. It has become the de facto standard for evaluating researchers, research institutes, and articles, as well as journals. However, some problems have been pointed out in JIF [9]. Several proposals to solve these problems [1, 6, 8, 9] have also been made.

Instead of expressing citations with simple numerical values, we propose to evaluate by focusing on the composition of the researcher group citing papers. Studies cited by many research groups may qualitatively differ from research cited mainly by a

small group. In this research, for the proposed measure indicating the difference, we report case analyses on different types of citations.

## 2 Evaluation

### 2.1 Gathering Article Data

In this section, we explain the collection method of articles for analysis, and conduct basic analysis of the data. The article data was gathered from Scopus<sup>1</sup>. In this experiment, “bibliometrics” was chosen as a query keyword and 10,186 articles published from 1976 to 2015 were gathered using Scopus's search API. These data are written in JSON format. The items are as follows: “Content Type,” “Search identifier,” “Complete author list,” “Resource identifiers,” “Abstract Text,” “First author,” “Page range,” “SCOPUS Cited-by URI,” “Result URL,” “Document identifier,” “Publication date,” “Source title,” “Article title,” “Cited by count,” “ISSN,” “Issue number,” and “Volume.”

Next, we collected the articles that cite the gathered articles. The information on citing articles is shown in the JSON item's “link” as a URL. Since the API cannot obtain the information described by the pages corresponding to those URLs, we obtained the HTML files with the “wget” command. The information about a maximum of 20 citing articles is published in an HTML file. However, since there were also articles with 20 or more citing articles, we needed to divide and obtain the files repeatedly.

We extracted information on the articles cited from the obtained HTML files and linked the articles by citation relation. This allowed us to quickly extract lists of authors citing a paper. At this time, we limited the ties between the 10,186 articles originally collected. To compare them under the same conditions, the citation count was also limited to those from the 10,186 papers. In other words, the citation count is a Focused Citation Count [7].

### 2.2 Construction of Researcher Group

This section explains how to extract a researcher group. Originally, to accurately identify the group of researchers, detailed investigation of the activities of each researcher was necessary, which is not easy. However, since the purpose of this paper is determining the number of researcher groups as an evaluation method to replace the cited number of papers, we considered that it would contribute to the analysis sufficiently with only the co-author information obtained from the bibliographic information of the cited paper.

The specific extraction methods of the researcher group are as follows. First, I will focus on one article,  $p_i$ . Let  $C(p_i)$  be the paper set citing this paper  $p_i$ . We will combine (union) lists with the same author for each author's list of articles in  $C(p_i)$ . For

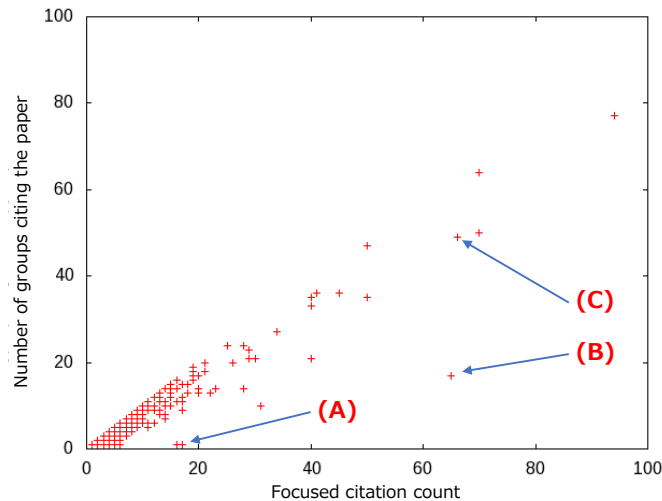
---

<sup>1</sup> <http://www.scopus.com/>

example, if the author list of  $c_1$  is  $(a_1, a_2, a_3)$  and the author list of  $c_2$  is  $(a_4, a_2, a_5)$ , their union list  $(a_1, a_2, a_3, a_4, a_5)$  will be constructed as a new researcher group. Additionally, it is maintained as a common researcher group for  $c_1$  and  $c_2$ . By repeating the same process, all authors of papers, including those belonging to the research group, will join that group. In addition, even if there are several papers with authors of this group, the citation count of the researcher group is counted as one. The authors of papers that do not share co-authors with other papers have built a research group themselves, and the citation count is one.

Fig. 1 plots the citation count and the citation count by group for each article. Since the citation count of the group is limited to the bibliographic information collected in Section 2.1, the horizontal axis for comparison is Focused Citation Count [7], limited to articles in the bibliometrics field. The vertical axis is the citation count by group.

In Fig. 1, many papers are plotted near the diagonal. They are articles that are cited by many research groups, not biased to a specific research group. There are also several papers below the diagonal line. This indicates that there is a research group citing the relevant paper from multiple papers. In the next section, we will analyze some samples in detail.



**Fig. 1.** Number of groups citing the paper

### 2.3 Analysis on Citations from Researcher Group

From Fig. 1, we can select some characteristic papers. Among them, the following three were chosen and analyzed.

- A) A paper located at the lower left of the graph, which has 17 citations while it is quoted from one group
- B) A paper located at the lower right of a graph that received more than 60

- citations, but the number of cited groups is less than 20
- C) A paper that is about the same as (B) and has almost the same number of quotations, and the number of cited groups is not small

**(A) Article-id: 2-s2.0-1442285649.**

Kostoff, R.N., Shlesinger, M.F., Tshiteya, R.: Nonlinear dynamics text mining using bibliometrics and database tomography, *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering*, Volume 14, Issue 1, pp.61–92. (2004)

This paper has 17 citations, but the citations are only by the author of the paper. In other words, all citations were self-citations. This paper relates to a technique for extracting information from a database of literature, and it is cited repeatedly in his own papers afterward, but there are no citations from other researchers. Kostoff points out in his own paper [4] that citations can be raised by self-citation as an issue inherent to paper evaluation by citation count. It is easy to remove self-citation from the evaluation, in fact you can exclude self-citation in the citation analysis provided by Scopus. The purpose of this research is not exclusion of self-citation, but evaluation using the number of research groups citing a certain paper. This example is not a good example for the purpose of this research, and it seems to be an appropriate sample in the sense that it was not included in the subject data of this research.

**(B) Article-id: 2-s2.0-78650989464.**

Chen, Y.-C., Yeh, H.-Y., Wu, J.-C., Haschler, I., Chen, T.-J., Wetter, T.: Taiwan's National Health Insurance Research Database: Administrative health care database as study object in bibliometrics, *Scientometrics* Volume 86, Issue 2, pp. 365–380. (2011)

This paper has 72 citations, with 17 researcher groups citing this paper. There are 12 self-citations, so it has 60 citations from 16 other groups of researchers. The content of this paper is analytical research on Taiwan's National Health Insurance Research Database (NHIRD), and the authors of the articles citing the paper were biased towards researchers belonging to research institutes in Taiwan. There is a possibility that the group of researchers is limited due to the regionality of its content.

**(C) Article-id: 2-s2.0-33748074153**

Daim, T.U., Rueda, G., Martin, H., Gerdri, P.: “Forecasting emerging technologies: Use of bibliometrics and patent analysis,” *Technological Forecasting and Social Change* Volume 73, Issue 8, pp. 981–1012. (2006)

This paper has 66 citations from 49 research groups. Although this paper uses literature analysis as a method, it relates to the prediction methods of emerging technologies, the content of which is not limited to the field of bibliometrics. For this

reason, it seems that it was quoted by a group of researchers from a wide variety of fields.

### 3 Conclusion

In this paper, we investigated a method to quantitatively measure the influence of a paper by using groups of researchers citing the paper as the object of analysis. We proposed the number of researcher groups citing a paper as an evaluation measure of the paper and analyzed the case using the citation information from 7,126 documents in the bibliometrics field. From the case analyses, it was confirmed that the number of research groups citing a paper gives characteristics different from citation count. In the future, we plan to establish an objective measure of how the number of researchers citing a paper and their distribution affects the positioning of the paper.

### Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP18K11990.

### References

1. Bergstrom, C.: Eigenfactor: Measuring the value and prestige of scholarly journals, *College & Research Libraries News*, Vol.68(5) pp.314–316, (2007)
2. Garfield, E.: Citation indexes for science, *Science*, Vol.122(3159), pp.108–111, (1955)
3. Garfield, E.: The history and meaning of the journal impact factor, *Journal of the American Medical Association*, Vol. 295(1), pp.99–93, (2006)
4. Kostoff, R. N.: Performance measures for government-sponsored research: Overview and background, *Scientometrics*, Vol.36(3), pp281–292. (1996)
5. Martin, B. R.: The use of multiple indicators in the assessment of basic research, *Scientometrics*, 36.3, pp.343–362. (1996)
6. Marshakova-Shaikovich, Irina.: The standard impact factor as an evaluation tool of science fields and scientific journals. *Scientometrics*, Vol.35(2), pp.283–290. (1996)
7. Nakatoh, T., Nakanishi, H. and Hirokawa, S.: Focused Citation Count: A Combined Measure of Relevancy and Quality, *Proc. of IIAI International Congress on Advanced Applied Informatics 2015 (IIAI AAI 2015)*, pp.166–170. (2015)
8. Nakatoh, T., Nakanishi, H. and Hirokawa, S.: Journal Impact Factor Revised with Focused View, *Intelligent Decision Technologies – Proc. of the 7th KES International Conference on Intelligent Decision Technologies, KES-IDT 2015*, Vol. 39, pp.471–481. (2015)
9. Pudovkin, A. I. and Garfield, E.: Rank-Normalized Impact Factor: A Way to Compare Journal Performance Across Subject Categories, *Proc. of the ASIST Annual Meeting*, Vol. 41, pp.507–515. (2004)