

Analysis of Academic Evaluation Indicators Based on Citation Quality

Mingyue Zhang*, Jin Shi*, Jin Wang**, and Chang Liu***

Abstract

The academic research performance is often quantitatively measured by means of using citation frequency. The citation frequency-based indicators, such as h-index and impact factor, are commonly used reflecting the citation quality to some extent. However, these frequency-based indicators are usually carried out based on the assumption that all citations are equal. This may lead to biased evaluations in that, the attributes of the citing objects and cited objects are significant. A high-accuracy evaluation method is needed. In this paper, we review various citation quality-based evaluation indicators, and categorize them considering the algorithms being applied. We discuss the pros and cons of these indicators, and compare them from four dimensions. The outcomes will be useful for our further research on distinguishing citation quality.

Keywords

Academic Evaluation Indicators, Citation Analysis, Citation Impact, Citation Quality

1. Introduction

In order to measure the value, influence or quality of academic activities and associated matters, quantitative criteria and methods are often used for developing evaluation indicators. Academic evaluation indicators vary with evaluation purposes, levels, depth, and content, therefore many academic evaluation indicators appeared.

Citation analysis is a major and commonly used method in constructing these indicators. From the viewpoint that applying quantitative measurements based on citation data, the impact, quality or value of works or researchers can be objectively reflected by the citations that they have received. As impact factor, h-index, and other citation-based metrics have become the most pervasive indicators in use, their limitations [1,2] are recognized and become more evident. It is observed that these indicators are all based on citation frequency/counts on the assumption that all citations are equal. It is obviously not true for two reasons. First, sole citation quantity cannot represent citation quality. Second, many other dimensions such as the citation databases, influence of a journal and author, the location and role of citation in the paper, relevance, culture and cognition have direct effects on the citation quality and cannot be overlooked.

※ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Manuscript received March 15, 2018; accepted May 20, 2018.

Corresponding Author: Jin Shi (shijin@nju.edu.cn)

* School of Information Management, Nanjing University, Jiangsu, China (mg1614027@smail.nju.edu.cn, shijin@nju.edu.cn)

** School of Computer & Communication Engineering, Changsha University of Science & Technology, Hunan, China (jinwang@yzu.edu.cn)

*** Dept. of Logistics and Information Management, Jiangnan University, Hubei, China (liuchang@jhu.edu.cn)

Numerous studies [3,4] have proposed methods to distinguish citations by considering citation quality other than solely citation frequency, however, the citation quality have not been systematically addressed in quantitative analysis literatures. Here, we examine a variety of quality-based indicators, extract the involved determinants, discuss their pros and cons, and compare them in practical dimension. This study lays the foundation for our eventual goal that is to develop objective and precise indicators which can be used for academic evaluation.

The remainder of this paper is organized as follows. In Section 2, we review three categories of citation quality evaluation indicators. In Section 3, we compare them in the evaluation dimension, research objects, computational complexity, and the degree of application. In Section 4, we conclude the findings and point out the future work.

2. Citation Quality Evaluation Indicators

In this section, we review three representative citation quality-based evaluation indicators. They are designed according to different principles, distinguishing the citation quality from the different angles.

2.1 Simple Frequency-Weighted Indicators

IF (impact factor), the first evaluation indicator promoted in 1972 by Garfield, assumes the citation with the different bibliographic record features equally. However, citations with the different citation bibliographic record have the different citation quality. From this perspective, simple frequency-weighted indicators can be roughly classified into the following four levels.

Frequency-weighted indicators considering citation time interval

It is observed that the influence of cited papers is correlated to its publishing time [3], i.e., the more recently published cited objects are considered to yield greater influence. Based on this finding, Ding et al. [5] found that most papers' citation time interval of library science and information science are not more than 2 years. Walker et al. [6] found that it would enter the recently published papers more likely in the process of random into a paper by analyzing the citation network. It means the shorter the citation time interval, the greater the influence of the cited papers. On this basis, Jarvelin and Persson [7] used attenuation parameters to reduce the weight of papers that published earlier. Yan and Ding [8] gave higher weights to papers with shorter citation interval by fitting the papers' time-citation curve.

Frequency-weighted indicators considering authors' influence, journals' influence, and papers' influence

From the perspective of citation description information, citations have some specific features that enable the assigning of different weights to citations. On the author level, Li [9] put forward the concept of citation quality-weighted impact factor based on the journal impact factor formula, using the h-index to weight citation frequency. On the journal level, Zheng and Liang [10] used weight W_j , the proportion that the cited journal's impact factor accounted for all journal's impact factor of this field, to distinguish the citation quality from different journals. On this basis, Lin [11] revised the weight W_j and proved that the revised indicator can distinguish the citation quality better. On the article level, Yan and Ding [8] adopted the article influence score provided by Thomson Reuters to weight the citations.

2.2 Network-based Citation Quality Evaluation Indicators

The network-based citation quality evaluation indicators build a citation network before designing the metrics. They emphasize the citation quality that obtains from other nodes though the citation network.

PageRank

The PageRank algorithm was initially used in the webpage ranking and was later improved for academic evaluation, distinguishing the citation quality from the perspective of journal reputation though citation network. The basic idea is: according to the links between the webpages, if webpage A links to webpage B, then B gets A's contribution to it, and this contribution score depends on the importance of A, that is, the more links from highly influential webpages, then the higher the impact of the page [12]. The method is inspired by bibliometric. For distinguishing citations quality, Pinski and Narin [13] proposed the iteration method to weight the citation quality. For a journal i , its iteration starting point is formula (1), then the weight of journal i is formula (2), where S_i represents the total number that journal i cites other journals, W_k represents the influence of journal k , C_{ki} indicates the total number that the journal k cites journal i . On this basis, various PageRank modifications emerge, such as CiteRank [14] which tries to overcome the problem of the aging effect in citation network.

$$W_1 = \frac{\text{The number of citations that journal } i \text{ receives}}{\text{the total number that journal } i \text{ cites other journals}} \quad (1)$$

$$W_i = \sum_{k=1}^n \frac{W_k C_{ki}}{S_i}. \quad (2)$$

Eigenfactor

In 2007, West et al. [15] proposed the concept of eigenfactor, achieving the theory of Pinski and Narin [13] successfully and distinguishing the citation quality from the perspective of journal reputation though citation network. After the eigenfactor was presented, Thomson Reuters published the Journal Citation Report, which added eigenfactor in 2009. The eigenfactor includes two indicators (eigenfactor score, article influence score), the former is the total effect of the journal, and the latter is the average impact of the articles.

The principle of eigenfactor can be described as follows. Supposing that a researcher randomly selects an article in a journal, then randomly selects a reference to the article and follows the reference relationship between the articles into the next journal, repeat the behavior. In this repetitive process, the greater the influence of the journal, the more the number that researchers enter the journal. The probability that a researcher enters a journal is the eigenfactor of the journal.

SJR (SCImago Journal Rank)

In 2007, the Spanish SCImago team presented a new academic evaluation indicator SJR (SCImago Journal Rank) based on Scopus database [16]. In 2008, the indicator was reported by Nature and its basic idea could be described as follows: the more the journal A is cited by prestigious journals, the higher the prestige of journal A. The results are broadly similar by using SJR or impact factor, but there are differences. These differences can be understood by popularity (citations from common journals/

articles) and prestige (citations from famous journals/articles). High popularity journals cited by low prestige journals have high impact factors and low SJR, while prestigious journals may be cited less frequently, and the impact factor is smaller, but citations from more prestigious journals give them higher SJR.

After the SJR was proposed, the SJR2 was proposed by SCImago team in 2012. Compared to SJR, the SJR2 added the cosine between the journals and the interdisciplinary weights of journals to further distinguish the citation quality.

2.3 Contend-based Citation Quality Evaluation Indicators

There are two aspects to evaluate the quality of citations based on citation content. Here are some typical methods from the two aspects to differentiate citation quality based on citation content.

The citation location and the actual number of occurrences in the cited content

Herlach [17] found that one-third of the articles were cited in an article more than once in his research and the cited content location was divided into four categories. Moreover, he argued that if a paper was cited in the introduction or literature review and subsequently mentioned again in the method or discussion section, then the paper might have a significant impact on the whole paper.

McCain and Turner [18] proposed utility index that evaluates the citation quality according to the different positions in the article, and gives the weight of the citation.

$$UI = W_{SC}[W_i \ln(X_i + 1) + W_m \ln(X_m + 1) + W_d \ln(X_d + 1) + W_r \ln(X_r + 1)] \quad (3)$$

W_{SC} is the aggregate index value depending on the relationship between key paper and source paper. If the key paper and source paper have the same author, W_{SC} is 0.1. If they have the same institution, W_{SC} is 0.5. Otherwise W_{SC} is 1. W_i, W_m, W_d, W_r indicate the importance of citations in different locations respectively: introduction, method, discussion, and review. X represents the number of occurrences of the corresponding position. In addition, four different weighted strategies were designed.

Domestic scholar Ding et al. [4] applied text-mining method to analyze the 32,496 citations of 866 articles. It was found that the highly cited articles appeared in the background and the literature review. Wan and Liu [3] proposed the concept of citation strength. Using 6 indicators determined the different citation levels, then using the 6 indicators and machine learning methods distinguished the citation quality. And the validity of the model was tested by two indexes. Finally, citation strength was used to evaluate the influence of the papers and authors in the calculation of impact factor and PageRank.

The theme, emotion, and function of cited content

This type methods evaluates citation quality mainly from the perspective of citing motivation and cited content feature. Finney [19] divided citations into 7 categories based on clues: recognized knowledge, experimental knowledge, methods, confirmation, denial, interpretation, future research. Nanba and Okumura [20] divided citations into 3 categories based on Finney [19]: Type B (base on), Type C (compare to), Type O (others). Peritz [21] classified the citations into 8 categories: outlining the state of the research field, providing background information, describing the method applied, contrasts,

opposing or supporting the new results, documentary referring to e.g., data collections, historical information, and perfunctory citations. Vinkler [22] divided citing motivations as academic-related citations and behavior-related citations, such as the relationship between authors.

Chinese scholars, Zhao [23], divided citations into positive citations, negative citations and neutral citations by emotion, and divided citations into deep citations, moderate citations and mild citations by citation depth. On this basis, Ye [24] concluded that the positive citations and the depth of citations were important for the evaluation of citation quality. However, some neutral citations, negative citations and mild citations cannot evaluate the quality of the cited articles. In addition, the weight should be made using appropriate citation data and adjusted according to the research object and purpose. Liu [25] used Wang Lan's classification method and the method of clue word matching to automatically determine the type of citation in order to weight citation quality.

3. Comparative Analysis

Based on principle and by concerns of practical applicability, this section compares the fore mentioned three categories of indicators from the perspectives of evaluation dimension, research objects, computational complexity, and the degree of application. The results are summarized in Table 1. Secondly, we selected the papers by a Turing Award Author as a sample and made an empirical study of three types of indicators. The results are showed in Fig. 1. Finally, we comment on our observations of the citation quality classification based on citation behavior.

Table 1. Comparison of academic evaluation indicators based on citation quality

Method	Simple frequency-weighted indicators	Network-based citation quality evaluation indicators	Content-based citation quality evaluation indicators
Representative research	[8], [9], [10]	[13], [15], [16]	[3], [5], [18], [25]
Principle	The citation quality is distinguished from the dimension of time, authors, journals and articles with the simple calculation method based on citation frequency.	The principle of PageRank is used to distinguish the citation quality according to the influence of the cited journals.	The citation quality is measured by citation location, the actual number of occurrences in the cited content, the citing content length, citing content density, citation motivation, as well as citation emotion.
Evaluation dimension	The dimension of time interval, authors, journals and articles	The dimension of journals	The dimension of citing content
Research objects	Citation description information	Citation network	Full-text
Computational complexity	Simple	Complex	Moderate
Application	No	JCR, Scopus	No

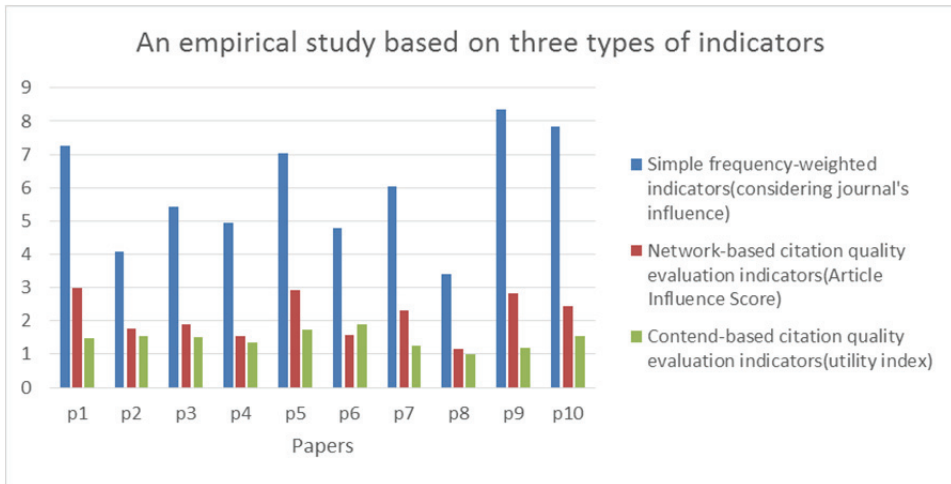


Fig. 1. An empirical study based on three types of indicators.

For the evaluation dimension, the results are described in Table 2. From the core of citation analysis, the citations' source and elements, summarize the evaluation dimension.

Table 2. Comparison of evaluation dimension

Method	Evaluation dimension	
	Where are the citations from	What elements of the citations have been considered of
Simple frequency-weighted indicators	The direct citing articles	Citation description information
Network-based citation quality evaluation indicators	The citing articles in the citation network	Citation description information
Content-based citation quality evaluation indicators	The direct citing articles	Full-text

For the research objects, the results are summarized in Table 3. Summarize the mainly research objects respectively.

Table 3. Comparison of research objects

Method	Research objects		
	Time	Journal impact, article impact, and author impact	Citation content
Simple frequency-weighted indicators	Yes	Yes	No
Network-based citation quality evaluation indicators	Yes	Yes	No
Content-based citation quality evaluation indicators	No	No	Yes

For the computational complexity, the results are described in Table 4. Summarize the computational complexity from the perspective of data preparation and data processing.

Table 4. Comparison of computational complexity

Method	Computational complexity	
	Data preparation	Data processing
Simple frequency-weighted indicators	Citation description information	Time interval, journals' impact, articles' impact and authors' impact are described by h-index or impact factor or other indexes
Network-based citation quality evaluation indicators	Citation description information, the citation relationship of a specified database	Time interval, journals' impact, articles' impact and authors' impact are described by h-index or impact factor or other indexes, iterating along the citation relationship.
Content-based citation quality evaluation indicators	Full-text of the original article	The citation location, the actual number of occurrences in the citing content, the citing content length, citing content density, citation motivation, as well as citation emotion

In order to intuitively compare the differences between the three types of indicators, we used the papers from the database Web of Science Core Set written by Naur Peter, who is the Turing Award Winner in 2005, to do an empirical study. We got 14 papers which written by Naur Peter and 264 citing papers which referred to this 14 papers in the core set of web of science by name retrieval and affiliation matching. Then, we selected the representative indicators of the three types of indicators: simple frequency-weighted indicators (considering journal's influence), network-based citation quality evaluation indicators (article influence score), and contend-based citation quality evaluation indicators (utility index). For simple frequency-weighted indicators (considering journal's influence), we used 2-year journal impact factor in 2016 which is calculated by

$$\text{2-year journal impact factor} = \frac{\text{citations in 2016 to items published in 2015,2014}}{\text{number of items published in 2015,2014}}$$

For network-based citation quality evaluation indicators (article influence score), we used the result of Web of Science. For contend-based citation quality evaluation indicators (utility index), we used formula (3) where W_i , W_m , W_d , and W_r are assigned 1, 3, 2, and 1, respectively, which was raised by McCain and Turner [17]. Finally, we took the top 10 papers written by Naur Peter and calculated the mean value of the three indicators corresponding to each article respectively. The results are shown in Fig. 1. From Fig. 1, we know that the value of simple frequency-weighted indicators (considering journal's influence) is larger, therefore, the distinction of evaluation is better. By contrast, the other two indicators are not. The first two types of indicators are highly correlated, if the value of the first type of indicators is large, the second type of indicators is also large. Therefore, the performance of the first type of indicators is better in terms of computational complexity and system closeness. In addition, if we want to use content-based citation quality evaluation indicators, we simply judge the location and frequency of occurrence, and whether the author is the same person or the same institution is far from meeting the evaluation needs. The differentiation of indicators designed in this way is not very good. At the same time, if the references are not standardized, it will result in actually citing many times, but only citing once. In addition, the placement of the position is more arbitrary, then the accuracy of this method will decline sharply. Therefore, integrating the content-based citation quality evaluation

method with the first and second indicators will improve the fault-tolerance and increase the accuracy.

Citation quality can be divided into two areas: different citations to the same citing paper, and the same citation to the different citing papers. First, different citations have their own intrinsic attributes—published journal, writing authors, published time, total cited times, the length of the paper, research subject, literature review or research papers, and the degree of similarity between the citing papers and cited papers. These different attributes determine the citation quality of different citations to the same citing paper. Second, paper has different cited characteristics when it is cited by the different papers. Cited papers have different citation location, different actual number of occurrences in the citing content, different cited content length, different cited content density, different citation motivation, as well as different citation emotion. These cited characteristics determine the citation quality. Therefore, even for the same cited paper, the citation quality is different to the different citing papers. To solve the question, some researchers do the further research from the perspective of citation text.

4. Conclusion and Future Work

The citation quality is the concept presented relative to citation frequency. Citation frequency assumes that all the citations are equally important in statistics. However, different citations have their own intrinsic attributes, and paper has different cited characteristics when it is cited by the different papers. This paper presents a comprehensive understanding of current research situation about academic evaluation indicators based on citation quality through the comparison and analysis of the three types of indicators. At the same time, it lays the foundation for the further research on distinguishing the different cited articles' influence to the same citing article, distinguishing the same cited article's influence to the different citing articles, and identifying Matthew effect. It is beneficial for measuring citation quality, further optimizing the citation analysis method in the academic evaluation.

In fact, some extensions can be made in the future. For instance, a comprehensive citation quality-based evaluation model can be constructed. Analyze the correlation of these influencing factors. In addition, we should also pay attention to identify the Matthew effect in the citing behavior. An article is cited because it really has a high quality, or just a follow phenomenon. How to avoid the phenomenon that the citation quality is exaggerated by the existing evaluation methods is needed to be further studied.

References

- [1] L. Bornmann and R. Haunschild, "Does evaluative scientometrics lose its main focus on scientific quality by the new orientation towards societal impact," *Scientometrics*, vol. 110, no. 2, pp. 937-943, 2017.
- [2] D. Zhao and A. Strotmann, "Dimensions and uncertainties of author citation rankings: lessons learned from frequency-weighted in-text citation counting," *Journal of the Association for Information Science and Technology*, vol. 67, no. 3, pp. 671-682, 2016.
- [3] X. Wan and F. Liu, "Are all literature citations equally important? Automatic citation strength estimation and its applications," *Journal of the Association for Information Science and Technology*, vol. 65, no. 9, pp. 1929-1938, 2014.

- [4] Y. Ding, L. Liu, C. Guo, and B. Cronin, "The distribution of references across texts: some implications for citation analysis," *Journal of Informetrics*, vol. 7, no. 3, pp. 583-592, 2013.
- [5] Y. Ding, G. Zhang, T. Chambers, M. Song, X. Wang, and C. Zhai, "Content-based citation analysis: the next generation of citation analysis," *Journal of the Association for Information Science and Technology*, vol. 65, no. 9, pp. 1820-1833, 2014.
- [6] D. Walker, H. Xie, K. K. Yan, and S. Maslov, "Ranking scientific publications using a model of network traffic," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2007, no. 6, article no. P06010, 2007.
- [7] K. Jarvelin and O. Persson, "The DCI index: discounted cumulated impact-based research evaluation," *Journal of the American Society for Information Science and Technology*, vol. 59, no. 9, pp. 1433-1440, 2008.
- [8] E. Yan and Y. Ding, "Weighted citation: an indicator of an article's prestige," *Journal of the American Society for Information Science and Technology*, vol. 61, no. 8, pp. 1635-1643, 2010.
- [9] C. Li, "Research on journal evaluation using the weighted impact factors considering cited quality: taking CSSCI source journals of LIS as an example," *Journal of Academic Libraries*, vol. 30, no. 1, pp. 29-34, 2012.
- [10] M. Zheng and F. Liang, "Impact factor modification based on the journal quality of citations," *Acta Editologica*, vol. 27, no. 1, pp. 19-21, 2015.
- [11] Z. Lin, "Application and discussion of weights of journal quality in the calculation formula of impact factor," *Chinese Journal of Scientific and Technical Periodicals*, vol. 26, no. 12, pp. 1295-1300, 2015.
- [12] D. Fiala, L. Subelj, S. Zitnik, and M. Bajec, "Do PageRank-based author rankings outperform simple citation counts?," *Journal of Informetrics*, vol. 9, no. 2, pp. 334-348, 2015.
- [13] G. Pinski and F. Narin, "Citation influence for journal aggregates of scientific publications: theory, with application to the literature of physics," *Information Processing & Management*, vol. 12, no. 5, pp. 297-312, 1976.
- [14] M. Dunaiski, W. Visser, and J. Geldenhuys, "Evaluating paper and author ranking algorithms using impact and contribution awards," *Journal of Informetrics*, vol. 10, no. 2, pp. 392-407, 2016.
- [15] J. D. West, T. C. Bergstrom, and C. T. Bergstrom, "The Eigenfactor Metrics: a network approach to assessing scholarly journals," *College & Research Libraries*, vol. 71, no. 3, pp. 236-244, 2010.
- [16] SCImago, (n.d.). SJR — SCImago Journal & Country Rank, 3 July 2018, from <http://www.scimagojr.com>.
- [17] G. Herlach, "Can retrieval of information from citation indexes be simplified? Multiple mention of a reference as a characteristic of the link between cited and citing article," *Journal of the American Society for Information Science*, vol. 29, no. 6, pp. 308-310, 1978.
- [18] K. W. McCain and K. Turner, "Citation context analysis and aging patterns of journal articles in molecular-genetics," *Scientometrics*, vol. 17, no. 1-2, pp. 127-163, 1989.
- [19] B. Finney, "The reference characteristics of scientific texts," City University of London, Centre for Information Science, 1979.
- [20] H. Nanba and M. Okumura, "Towards multi-paper summarization using reference information," *Journal of Natural Language Processing*, vol. 6, no. 82, pp. 79-86, 1999.
- [21] B. C. Peritz, "A classification of citation roles for the social sciences and related fields," *Scientometrics*, vol. 5, no. 5, pp. 303-312, 1983.
- [22] P. Vinkler, "A quasi-quantitative citation model," *Scientometrics*, vol. 12, no. 1-2, pp. 47-72, 1987.
- [23] Q. Zhao, "Research on citation character and citation depth in literature," *Journal of Information*, vol. 29, no. 10, pp. 46-50, 2010.
- [24] J. Ye, "Analyzing the essence of citations and their functions in academic assessment," *Journal of Library Science in China*, vol. 36, no. 1, pp. 35-39, 2010.
- [25] S. Liu, "Research on the citation evaluation based on citation context nature," *Information Studies: Theory & Application*, vol. 38, no. 3, pp. 77-81, 2015.



Mingyue Zhang <https://orcid.org/0000-0002-1739-0047>

She is an undergraduate student in School of Information Management, Nanjing University. Her research interests mainly include information management.



Jin Shi <https://orcid.org/0000-0002-1621-6944>

He is an associated professor in School of Information Management, Nanjing University. His research interests include information security, machine learning.



Jin Wang <https://orcid.org/0000-0002-6516-6787>

He received the B.S. and M.S. degree from Nanjing University of Posts and Telecommunications, China in 2002 and 2005, respectively. He received Ph.D. degree from Kyung Hee University Korea in 2010. Now, he is a professor in the School of Computer & Communication Engineering, Changsha University of Science & Technology. His research interests mainly include routing algorithm design, performance evaluation and optimization for wireless ad hoc and sensor networks. He is a Member of IEEE and ACM.



Chang Liu <https://orcid.org/0000-0002-3702-516X>

She is an associated professor in Department of Logistics and Information Management, Jiangnan University. Her research interests include information management.