Evaluating Scholarly Impact Through Open Peer Review: A Case Study of F1000 Postprint

Authors: Yuan Guohua, Kou Jingjing, Zhang Jianyong, Han Zhengqi

Date: 2023-08-27T00:00:00+00:00

Abstract

[Purpose/Significance] This paper proposes a scholar influence evaluation method based on open peer review, which inductively converges open peer review content to form multiple tag clusters that characterize scholars' influence at different levels, thereby exploring scholar influence from multiple quality perspectives and verifying the effectiveness of the proposed method. [Method/Process] This study investigates and analyzes existing scholar influence evaluation methods domestically and internationally, designs and constructs a scholar influence evaluation method based on open peer review, and conducts empirical research and comparative analysis using the open peer review platform F1000 as a case study. [Results/Conclusion] The scholar evaluation method based on open peer review represents a new evaluation perspective that achieves multi-angle assessment of scholars and enhances the richness and distinctiveness of scholar influence. Analysis indicates that this method does not exhibit significant correlation with bibliometric evaluation methods and can serve as a supplement to traditional evaluation approaches. The integration of traditional quantitative analysis with open peer review methods will be the future development trend in academic evaluation.

Full Text

Preamble

ChinaXiv Collaborative Journal, Vol. 62, No. 13, July 2018

Research on Scholar Influence Evaluation Based on Open Peer Review: A Case Study of F1000

YUAN Guohua[1],[2], KOU Jingjing[1],[2], ZHANG Jianyong[1], HAN Zhengqi[1],[2]

[1] National Science Library, Chinese Academy of Sciences, Beijing 100190
[2] University of Chinese Academy of Sciences, Beijing 100049

Abstract

[Objective/Significance] This paper proposes a novel method for evaluating scholar influence based on open peer review. By systematically synthesizing open peer review content, the method constructs multiple tag clusters that characterize different dimensions of scholarly impact, thereby enabling multi-angle exploration of influence from a qualitative perspective and validating the effectiveness of the proposed approach. [Method/Process] The study first investigates and analyzes existing domestic and international methods for evaluating scholar influence. Building upon this foundation, we design and construct an evaluation framework based on open peer review, using the F1000 platform as a case study for empirical research and comparative analysis. [Result/Conclusion] The open peer review-based evaluation method represents a new evaluation perspective that achieves multi-dimensional assessment of scholars and enhances the richness and distinctiveness of influence characterization. Analysis indicates that this method does not exhibit significant correlation with bibliometric evaluation approaches, suggesting it can serve as a valuable supplement to traditional evaluation methods. The integration of traditional quantitative analysis with open peer review methods will be the future trend in academic evaluation.

Keywords: Scholar Influence; Evaluation; Open Peer Review; F1000

Introduction

Scholar influence evaluation plays a crucial role in research management and assessment, representing a long-standing focus in the field of academic evaluation. Effective evaluation of scholar influence promotes fair competition in scientific research and enhances scholars' enthusiasm for conducting research. The breadth and depth of a scholar's influence primarily depend on their research outputs and the extent to which these outputs are valued, recognized, and applied by others. Consequently, evaluating scholars ultimately boils down to assessing their academic achievements. Current academic evaluation relies mainly on bibliometric methods, with citation analysis being the most widely used approach. While this method has gained broad recognition, it has also faced continuous criticism. Researchers have consistently hoped to supplement quantitative evaluation with qualitative methods. Particularly as scientific publishing, dissemination, and communication become increasingly open, the flourishing development of social media provides more suitable platforms and media for open evaluation. Open peer review (OPR) has emerged in response to these trends, making large-scale qualitative evaluation possible.

In May 2017, the EU OpenAIRE project released a report titled "OpenAIRE Survey on Open Peer Review: Attitudes and Experience Amongst Editors, Authors and Reviewers" [1]. The survey revealed that three-quarters of respon-

dents had participated in open peer review in various ways. The emergence of open peer review has stimulated research enthusiasm in the academic evaluation field. Currently, some scholars have attempted to use open peer review to evaluate articles and scholars, and have conducted correlation studies comparing these methods with citation-based bibliometric approaches. For instance, Song Liping et al. compared the effectiveness and correlation of peer review and bibliometric methods in scientific evaluation based on F1000 and Web of Science [19], concluding that bibliometric indicators are positively correlated with peer review results. Shen Xiaoling et al. [20] explored webometric indicators associated with peer review to establish a comprehensive evaluation model as an alternative to peer review. However, existing research on academic evaluation using open peer review has primarily focused on quantitative indicators such as the F1000 factor and Altmetrics, while largely overlooking the qualitative review content generated by open peer review, which is equally important for academic evaluation. Therefore, unlike previous studies, this paper focuses on the qualitative content produced by open peer review, using F1000 tags as the basis for scholar influence evaluation to construct a novel assessment method.

2 Research Progress

2.1 Existing Scholar Influence Evaluation Methods

Peer review dates back nearly 300 years to the mid-18th century when the Royal Society of Edinburgh and the Royal Society of London invited their members to select articles for publication [2]. Since then, academic evaluation has evolved from peer review-dominated to a coexistence of peer review and bibliometrics, and finally to bibliometrics-dominated approaches. In the domain of scholar influence evaluation, Lotka proposed an empirical law describing the relationship between authors and paper productivity in 1926, later known as "Lotka's Law" [3], which laid the foundation for linking scholar output to influence evaluation. The term "bibliometrics" was formally introduced in 1969, and since then, numerous bibliometric-based methods for evaluating scholar influence have emerged, ranging from Garfield's citation frequency to Hirsch's h-index [4], and various h-index variants and PageRank derivatives. Researchers have made persistent efforts in this area, including Egghe's g-index [5], Zhang's e-index [6], and domestic scholars' AuthorRank [7] and LeaderRank [8]. These methods can be categorized into three types: mean-type indicators, h-type indicators, and correlation-type indicators [9].

Mean-type indicators, such as average citations per paper and impact factor, use mean values as their core concept to eliminate the influence of article quantity or publication year, but they have not achieved ideal results. For example, average citations per paper ignores the impact of citation motivation on evaluation outcomes. H-type indicators, derived from the h-index, have been extensively studied, with Spain even hosting a dedicated website "H-index and Variants" [10] providing comprehensive information on the h-index principle, applications, and variants. According to incomplete statistics, there are currently

over 30 h-type indicators, each addressing different limitations of the original h-index. Correlation-type indicators evaluate scholars based on relationships between them, specifically through mutual citation and collaboration networks. Popular research in this category involves PageRank improvements, leveraging similarities between web links and author citation/collaboration relationships to develop various indicators, such as AuthorRank proposed by Sun Haisheng et al., LeaderRank by Wang Xiaomei et al., and Ma Ruimin et al.'s method using the Eigenfactor algorithm to evaluate scholars [11].

2.2 Emergence and Development of Open Peer Review

With the continuous development of information technology and social media platforms, the ways academic achievements are published, disseminated, and evaluated are evolving. The shift toward online and social media-based access and evaluation provides suitable platforms and media for open evaluation. Open peer review represents a manifestation of peer review in the era of open science. The term first emerged in the late 20th century as a supplement to traditional peer review and has since permeated both pre-publication and post-publication processes. This developmental trajectory has led to diverse definitions in academia. For instance, N. McCormack [12] defines open peer review as mutual knowledge of author and reviewer identities, while A. Mulligan [13] argues that identities should be disclosed not only between authors and reviewers but also to the public. D. Shotton [14] and P. Perakakis [15] view open peer review as a process of public participation in determining manuscript acceptance, where articles receive public comments immediately after submission, and these comments, together with formal peer review results, determine publication fate. Domestic scholar Liu Chunli [16] defines it as a process where peer experts in a network community rapidly identify important literature in academic publications based on their expertise. Jiang Chunlin [17] considers it an online review process where peer experts use social tools to evaluate papers on open access platforms while interacting with authors. E. Ford [18] systematically identified eight characteristics of open review, five describing the review process (whether reviews are signed, whether author and reviewer identities are mutually disclosed and allow discussion, whether editors control the process, whether results are publicly disclosed, and whether readers can comment) and three describing review timing (pre-publication, during publication, and post-publication review).

This study focuses on post-publication open peer review, which specifically refers to articles already accepted through traditional peer review and subsequently evaluated by other peers or readers.

3 Scholar Influence Evaluation Based on Open Peer Review

The core idea of our evaluation method is to systematically synthesize open peer review content to form multiple tag clusters that characterize different dimensions of scholarly influence, thereby enabling multi-angle exploration of

influence from a qualitative perspective. Since the evaluation approach is based on F1000, it is necessary to first provide a detailed explanation of F1000's open peer review system.

3.1 Overview of F1000 Open Peer Review

F1000, officially known as Faculty of 1000 [21], is a representative platform for post-publication open peer review. F1000 defines itself as a database providing secondary evaluation for life science articles. Its open review exists in three forms: rating system, textual review, and tag-based review. Reviewers are renowned experts from prestigious institutions in the US and Europe who evaluate papers based on their contribution to current biological and medical research and scientific value. The rating system allows reviewers to score articles as Good (one star), Very Good (two stars), or Exceptional (three stars). The composite scores from all reviewers constitute the article's F1000 factor, which is used to recommend a small number of outstanding papers published within the last month. Textual review involves experts writing detailed evaluation comments. Tag-based review requires experts to assign tags from a predefined tag library [22] rather than using arbitrary terms. Table 1 shows F1000's tag library and definitions.

In this paper, we select F1000 as the representative platform for open peer review to construct our evaluation method for the following reasons: (1) F1000 does not use journal prestige as the basis for article or author evaluation, treating all journals equally by focusing on both high-impact and ordinary journals. To date, F1000 has recommended articles from over 3,500 journals, providing a fairer evaluation method that recognizes high-quality work from lesser-known authors. (2) F1000's review results demonstrate high professionalism. The platform brings together over 8,000 top life scientists worldwide, ensuring both broad participation and professional expertise, yielding trustworthy evaluation results and rich data. (3) F1000's open review data is well-structured, presenting results through review ratings and content. The ratings form the F1000 factor, a direct indicator of article quality, while content appears as text and tags that are structurally related, with tags summarizing textual reviews. (4) F1000 provides rich review content with substantial usable data. While many academic platforms offer open review functions with minimal participation, F1000's approach, which incorporates certain mandatory elements, generates abundant reviews suitable for academic evaluation based on open review data.

3.2 Evaluation Framework Construction

When evaluating scholars, whether through traditional or novel methods, the assessment is based on research outcomes—specifically, scholarly publications. Our method follows this principle. Figure 2 [Figure 2: see original paper] illustrates the evaluation framework based on F1000 and scholarly papers. The process involves: (1) Information crawling: Extracting author names, titles, tags, F1000 factors, and other information from F1000. (2) Tag statistics: Counting

tags for each article to determine the frequency of each tag. (3) Tag clustering: To facilitate analysis and more accurately highlight author characteristics, tags are grouped into clusters representing different features, referred to as tag clusters. Table 2 shows the specific clustering results. (4) Tag migration: Identifying first authors and transferring tag clusters from articles to corresponding scholars, creating author-specific tag clusters called LabelScore that characterize different aspects of authors. (5) Empirical study: Conducting empirical research based on F1000 and performing correlation and comparative analyses with existing evaluation methods.

4 Empirical Research

4.1 Data Collection

Our experimental data were collected from F1000 and Web of Science (WoS). Using F1000 as the representative platform for open peer review, we selected the cardiovascular discipline and used Python-based web scraping tools to obtain information on 4,079 articles, including authors, titles, F1000 factors, and tags (data collection date: December 7, 2017), saved as .csv files. Partial results are shown in Figure 3 [Figure 3: see original paper]. WoS, as an authoritative citation database with broad coverage, provides direct citation-based assessment of academic value and serves as the representative for mainstream bibliometric evaluation methods.

4.2 Data Processing

The collected data were processed as follows: (1) Using Excel and Python to aggregate the count of each tag per article. (2) Applying the integrated tag clusters to categorize tags for each article. (3) Adopting the first-author contribution allocation method, where all article contributions are assigned to the first author. (4) Classifying and summarizing by first author name to identify the top 20 authors by publication count. (5) Calculating each author's characteristic values, referred to as LabelScore, using formula (1), where N represents the number of articles by the author on F1000, i indexes articles (i = 1, 2, ..., N), j indexes tag clusters (j = 1, 2, 3), and a_{ij} represents the score of article i on tag cluster j. This yields four LabelScore values characterizing each author. (6) Manually searching for each author's articles in WoS to record citation frequencies and calculate average citations per paper as the basis for bibliometric evaluation.

4.3 Empirical Results and Analysis

Table 3 presents evaluation results for the top 20 authors by publication count using our proposed method. Visualizing Table 3 yields Figure 5 [Figure 5: see original paper], where the horizontal axis represents LabelScore, the vertical axis shows author names, and different colored bars represent different LabelScore dimensions. Bar length indicates LabelScore magnitude. Table 3 and Figure

5 demonstrate that our method provides scholars with multiple scores reflecting various characteristics rather than a single evaluation score, yielding more comprehensive results.

From the horizontal perspective, authors' characteristics can be inferred from their scores and bar lengths across different tags. For clearer illustration, Figure 6 [Figure 6: see original paper] presents radar charts of tag cluster scores for three authors: J.J. McMurray, J. Liu, and M. Amann. The radar charts reveal distinct patterns: J.J. McMurray excels in "NewFinding," J. Liu shows relatively high performance in "ClinicalPractice" and "Review," while M. Amann's characteristics are primarily reflected in "Review." From the vertical perspective, the method identifies authors who excel in specific dimensions—for instance, X. Wang in practice, F. Liu in innovation, and J. Liu in controversial work. J. Liu, M.A. Laflamme, and M. Amann demonstrate strong performance in review/commentary, as indicated by longer bars. This vertical comparison helps quickly identify outstanding authors and their exceptional articles for specific characteristics.

4.4 Comparative Analysis

To validate the effectiveness and necessity of our method, we compared open peer review results with bibliometric evaluation represented by average citations per paper. Using Excel, we generated trend charts from Table 3 data (Figure 7 [Figure 7: see original paper]), then performed Pearson correlation tests using SPSS software among tag clusters, average F1000 factor, and average citation frequency. Results are shown in Table 4 .

The analysis reveals: (1) The correlation coefficient between average F1000 factor and average citation frequency is 0.625, significant at the 0.01 level (two-tailed), with similar trends in Figure 7, indicating high consistency between the quantitative open peer review indicator (F1000 factor) and the bibliometric indicator. This validates the effectiveness of open peer review in academic evaluation. (2) The correlation coefficients between the three tag clusters (ClinicalPractice, NewFinding, Review) and average citation frequency are -0.135, -0.275, and 0.178, respectively, with corresponding Sig values and trend lines (Figure 7) showing very weak or non-existent correlations. This suggests that qualitative evaluation based on open peer review addresses dimensions not covered by traditional methods, highlighting different authors than conventional approaches. (3) The correlation coefficients between the three tag clusters and the F1000 factor are 0.153, -0.209, and 0.005, respectively, with non-significant Sig values, indicating that while focusing on quantitative indicators from open review platforms and social media, we should also consider qualitative evaluation content.

In summary, open peer review-based scholar evaluation offers a new perspective that can supplement traditional methods. Scholar evaluation should not be mutually exclusive; 多元结合 is necessary for comprehensive assessment. Both

open peer review and bibliometrics have limitations when used alone. The combination of traditional quantitative analysis and open peer review represents the future development trend of academic evaluation.

Conclusion

This study systematically investigated existing domestic and international methods for evaluating scholar influence, proposing a novel evaluation framework based on open peer review and conducting empirical research using the F1000 platform. The empirical analysis revealed that some highly-cited authors do not perform particularly well under open peer review, while some authors with lower citation counts demonstrate outstanding performance from the open peer review perspective, suggesting a complementary relationship between the two approaches. Our findings indicate: (1) The proposed method assigns multiple characteristics to scholars, breaking the single-evaluation-angle dilemma and enabling multi-angle assessment that enhances the richness and distinctiveness of scholar profiles. (2) Statistically, the F1000-based open peer review method shows no significant correlation with citation analysis-based bibliometric methods, suggesting qualitative evaluation based on open peer review content addresses dimensions not covered by traditional methods and highlights different authors. (3) Open peer review-based evaluation provides a new perspective that can supplement traditional methods. The combination of traditional quantitative analysis and open peer review will be the future trend in academic evaluation.

However, this study has limitations. We did not comprehensively consider author contribution and attribution issues, focusing only on first authors' open peer review profiles, which may affect results to some extent. Future research should investigate the impact of authorship order and contribution factors on scholar influence evaluation to truly refine open peer review-based assessment.

References

[1] OpenAIRE. OpenAIRE survey on open peer review: attitudes and experience amongst editors, authors and reviewers [EB/OL]. [2018-01-22]. https://zenodo.org/record/570864#.wzvrz49oloy.

[2] KRONICK D A. Peer review in 18th-century scientific journalism [J]. Jama the journal of the American Medical Association, 1990, 263(10): 1321-1322.

[3] ZHANG Haiyan. Review of Lotka's law research based on bibliometrics [J]. Chinese Journal of Medical Library and Information Science, 2013, 22(8): 18-21.

[4] HIRSCH J. An index to quantify an individual's scientific research output [J]. Proceedings of the National Academy Of Sciences, 2005, 102(46): 16569-16572.

[5] EGGHE L. An improvement of the h-index: the g-index [J]. ISSI newsletter, 2006, 2(1): 8-9.

[6] ZHANG C T. The e-index, complementing the h-index for excess citation [J]. PLoS ONE, 2009, 4(5): 1-4.

[7] SUN Haisheng, HAN Hong. Empirical study on using AuthorRank citation identity for researcher evaluation [J]. New Technology of Library and Information Service, 2015, 31(9): 60-67.

[8] DENG Qiping, WANG Xiaomei. Using LeaderRank to identify influential authors [J]. Journal of the China Society for Scientific and Technical Information, 2011, 30(7): 30-33.

[9] KOU Jingjing. Research progress on Chinese scholar influence evaluation [J]. Information Research, 2017, 1(11): 129-134.

[10] H-index and variants [EB/OL]. [2017-08-22]. http://sci2s.ugr.es/hindex.

[11] MA Ruimin, HAN Xiaolin. Research on author influence evaluation based on improved Eigenfactor algorithm [J]. Journal of Chongqing University (Social Science Edition), 2015, 21(2): 106-110.

[12] MCCORMACK N. Peer review and legal publishing: what law librarians need to know about open, single-blind, and double-blind reviewing [J]. Law library journal, 2009, 101(1): 59-70.

[13] MULLIGAN A. Quality, certification and peer review [J]. Information Services & Use, 2008, 28(3-4): 197-214.

[14] SHOTTON D. The five stars of online journal articles—a framework for article evaluation [EB/OL]. [2017-09-13]. http://www.dlib.org/dlib/january12/shotton/01shotton.html.

[15] PERAKAKIS P, TAYLOR M, MAZZA M, et al. Understanding the role of open peer review and dynamic academic articles [J]. Scientometrics, 2011, 88(2): 669-673.

[16] LIU Chunli. A soft peer review-based method for evaluating scientific paper influence: the F1000 factor [J]. Chinese Journal of Scientific and Technical Periodicals, 2012, 23(3): 383-386.

[17] JIANG Chunlin, ZHANG Liwei, LIU Xue. Discussion on current status and issues of peer review research in China and abroad [J]. Chinese Journal of Scientific and Technical Periodicals, 2015(3): 163-166.

[18] FORD E. Defining and characterizing open peer review: a review of the literature [J]. Journal of scholarly publishing, 2013, 44(4): 311-326.

[19] SONG Liping, WANG Jianfang. Correlation study between peer review and bibliometrics based on F1000 and WoS [J]. Journal of Library Science in China, 2012(2): 62-69.

[20] SHEN Xiaoling, XU Yong, YAN Weizhong. Research on comprehensive evaluation of academic impact of scientific papers based on webometrics [J]. Library and Information Service, 2013, 57(21): 95-103.

[21] F1000 Prime FAQs [EB/OL]. [2017-08-22]. https://f1000.com/prime/about/faqs.

[22] F1000 Prime Classifications [EB/OL]. [2017-08-22]. https://f1000.com/prime/about/whatis/how.

[23] LOLL S J, VICTOR J L, YOUNG J M, et al. Genome-wide non-mendelian inheritance of extra-genomic information in Arabidopsis [J]. Nature, 2005, 434(7032): 505-509.

Author Contributions

YUAN Guohua: Designed and implemented the technical solution, performed data analysis, and wrote and revised the manuscript.
KOU Jingjing: Proposed the research idea, designed the study, and wrote and revised the manuscript.
ZHANG Jianyong: Provided guidance on manuscript writing.
HAN Zhengqi: Responsible for data cleaning and manuscript revision.

The Research of Scholar Influence Evaluation Based on Open Peer Review: Take the F1000 as an Example

YUAN Guohua[1],[2], KOU Jingjing[1],[2], ZHANG Jianyong[1], HAN Zhengqi[1],[2]

[1] National Science Library, Chinese Academy of Sciences, Beijing 100190
[2] University of Chinese Academy of Sciences, Beijing 100049

Abstract: [Purpose/significance] This paper designed and built an evaluation method of scholar influence based on open peer review. [Method/process] First the paper investigated the evaluation methods of scholar influence home and abroad. Then an evaluation method of scholar influence based on open peer review was designed and built. And Empirical research and comparative analysis were conducted. [Result/conclusion] The method of scholar evaluation based on open peer review is a new angle of evaluation, which achieves a multi-angle evaluation of scholars and increases the fullness and discernment of scholar influence. The analysis shows that there is no significant correlation between this method and the literature evaluation method, which can be used as a supplement to traditional evaluation methods. The combination of traditional quantitative analysis methods and open peer review methods will be the development trend of future academic evaluation.

Keywords: Scholar Influence; Evaluation; Open Peer Review; F1000

Note: Figure translations are in progress. See original paper for figures.