Postprint: Research on Evaluating Representative Academic Papers Based on P-4Cs

Authors: Yu Xi, Gao Yuhang

Date: 2024-06-13T00:00:00+00:00

Abstract

Specific and actionable methods for evaluating academic achievements are of great significance to academic evaluation, talent recruitment, discipline development, and social progress. This paper proposes constructing a P-4Cs performance-semantic four-dimensional extensible evaluation system based on an analysis of the current status and existing problems in representative work evaluation. The system conducts performance indicator evaluation, research topic positioning, citation context analysis, and multi-dimensional analysis based on scientometrics for representative works. The P-4Cs evaluation system provides judgment from external performance indicators to internal academic value, realizing comprehensive and multi-angle analysis and evaluation of papers, and providing references for further peer review.

Full Text

Research on Representative Academic Paper Evaluation Based on the P-4Cs Framework

Abstract: Specific and operable methods for evaluating academic achievements are of great significance for academic evaluation, talent introduction, disciplinary development, and social progress. Based on an analysis of the current status and existing problems of representative work evaluation, this paper proposes constructing a performance-semantic four-dimensional scalable evaluation system (P-4Cs). The system evaluates representative works through performance indicators, research theme positioning, citation context analysis, and multi-dimensional analysis based on scientometrics. The P-4Cs evaluation system provides judgment from external performance indicators to internal academic value, enabling comprehensive and multi-angle analysis and evaluation of papers, and provides references for further peer review.

Keywords: Performance; Semantics; Scientometrics; Visual analysis; P-4Cs

1 Background

Evaluation based on single indicators treats metrics as the sole criterion for judgment, but this approach has many drawbacks. Academic evaluation should be comprehensive, incorporating visual analysis methods to conduct integrated assessments of representative works from external evaluation metrics to citation contexts and paper content. This provides policymakers and talent performance management departments with objective references. Relying on single evaluation indicators or simple multi-dimensional indicators directly affects faculty hiring, promotion, and resource allocation.

From a third-party perspective, evaluation based on bibliometrics and scientometrics serves as a direct assessment method for the scientific community. The evaluation of representative works should use scientometrics to provide objective references for peer review, enabling evaluation subjects to make value judgments close to the true worth of the assessed object and facilitating experts' understanding of the evaluation object.

Since the national policy environment proposed changing the previous indicator-based evaluation and advocating a representative work evaluation system, a series of relevant policies and documents have been issued continuously. The "Guiding Opinions on Deepening the Reform of the Professional Title System for University Teachers" issued by the Ministry of Human Resources and Social Security and the Ministry of Education proposed overcoming tendencies such as "paper-only" and "project-only" evaluation, and not using educational background and other factors as restrictive conditions. It also stated that Science Citation Index (SCI) and Social Sciences Citation Index (SSCI) indicators should not be used as preconditions and direct basis for judgment [2]. This is the most instructive opinion to date, but specific operational methods still need to be explored.

Since academic papers constitute the main component of research achievements, this study focuses primarily on the evaluation of academic papers. Current research on academic paper evaluation mainly includes: academic value judgment based on citation counts [3-5], using alternative metrics (altmetrics) such as download counts and reprint counts to evaluate paper quality [6-8], academic paper evaluation using social network analysis [9-11], and academic paper quality evaluation based on peer review [15]. In addition to these evaluation methods, a popular approach in recent years is Ye Jiyuan's "full evaluation system" [16], which conducts comprehensive, all-element evaluations of evaluation objects. However, this method mainly targets humanities and social sciences and lacks operability. Shang Hairu et al. [17] proposed an evaluation method based on full-text bibliometric analysis, but such methods are still limited. These approaches essentially revolve around quickly extractable external features of papers or simple evaluative indicators, and to some extent ignore the disciplinary attributes, research characteristics, and impact on the field.

In this context, service departments should reflect on the evaluation of academic achievements. Even before the state issued a series of policies on scientific achievement evaluation, distortions had emerged in the application of bibliometrics and scientometrics theories in practice, such as the disconnect between metrological theory and practice, linear extension of indicators according to the subjective will of the evaluated, and singular thinking. In 2011, Zhang Hanxin, an academician in the field of fluid mechanics, published his views on evaluation criteria for scientific research achievements. The evaluation of representative works should not simply give a score or grade, but should provide a value judgment close to the discipline's development direction, changing scientific development patterns, or solving scientific uncertainties. Different disciplines and problems may require different evaluation methods. Under the current evaluation environment, we cannot precisely evaluate the quality of scientific achievements in a short time, so the evaluation focus should be placed as much as possible on the scientific community. Different disciplines should be treated differently and cannot be generalized.

2 Models of Scientific Knowledge Development

The first and most common model of scientific knowledge development is cumulative growth, where new knowledge grows logically from old knowledge. The second model involves new theoretical hypotheses that compete with traditional theories and are either accepted or falsified. The third model is scientific revolution [20], where new scientific theories are not derived from recent scientific development but are inspired by theories from a certain stage in scientific history. The first model emphasizes continuous cumulative growth, while the third model includes both periodic cumulative growth and periodic leap growth.

Kuhn proposed that scientific advancement is an endless process built upon scientific revolutions, during which scientific revolutions emerge one after another [21]. The key to generating new ideas through scientific revolution lies in whether more convincing explanations can be made for observed phenomena. Kuhn's scientific revolution represents the alternation and rise and fall of old and new scientific paradigms. Crises in scientific understanding and paradigm conversions caused by these crises will leave traces in academic literature. If the scientific process is truly as Kuhn observed, we should be able to identify the rise and fall of paradigms from scientific literature and conduct multi-dimensional analysis of papers to determine whether they can trigger scientific revolutions and play key roles in paradigm conversion.

Ronald Burt from the University of Chicago proposed structural holes theory when studying social networks and social value. In social networks, not everyone has direct connections with others, creating structural incompleteness that constrains information flow. People around structural holes often have greater advantages due to access to diverse information that provides greater imaginative space [22-23]. This theory can be extended to citation networks, where structural holes connect different clusters. Exploring citation structures based on structural holes theory can help identify important literature within a field and papers that play key roles in paradigm conversion [24].

Academic knowledge systems are complex adaptive systems that can respond to disturbances caused by new literature in multiple ways [26]. Each new literature can change the system's current state to varying degrees—either maintaining the current structure or changing it. Transformative research will ultimately change existing research fields or promote the formation of new fields. Key node literature in the paradigm shift process or literature that can change the current system structure represents truly important achievements in scientific fields. For example, Japanese biologist Shinya Yamanaka's pioneering work on induced pluripotent stem cells (iPSCs) and British scientist Gurdon's contributions in this area won the 2012 Nobel Prize in Physiology or Medicine. High burstiness papers play a crucial foundational role in such achievements.

3 The P-4Cs Performance-Semantic Four-Dimensional Scalable Evaluation System

3.1 System Logic Overview

The P-4Cs performance-semantic four-dimensional scalable evaluation system evaluates academic achievements using a combination of external performance indicators and internal academic value analysis. External performance indicators include both the paper's own metrics (such as Journal Impact Factor (JIF), Category Normalized Citation Impact (CNCI), Journal Normalized Citation Impact (JNCI), H-index) and analysis of citing literature performance (such as whether citing authors or papers are highly cited). Internal academic value analysis includes node characteristics in citation networks, the paper's position in the research 脉络 of the field, and citation context analysis based on content.

The evaluation system integrates scientometrics, visualization analysis, and peer review to provide multi-dimensional analysis of academic papers. By drawing semantic concept trees, it analyzes the hierarchical structure of research from a semantic perspective. It can also extract topic words from keywords to analyze whether the content represents mainstream or marginal research in the discipline, or entirely new research. This system achieves the combination of qualitative and quantitative, macro and micro perspectives, transforming the presentation from simple data to descriptions and 图谱 analysis.

3.2 The 4C Value Levels and Evaluation Criteria

The 4C framework refers to four progressively increasing value levels: **Cost** (basic value), **Common** (contribution), **Creative** (innovation), and **Change** (revolutionary). **Cost** represents the most basic affirmation of research achievement value—as long as a paper passes peer review and is published in journals like *Science*, it is recognized as having basic academic value. **Common** refers to research that maintains the integrity of the original knowledge system and enriches, supplements, and perfects it. **Creative** refers to research that changes existing research fields or promotes the formation of new fields. **Change** refers to achievements with the potential to trigger scientific revolutions.

These four levels correspond to different paradigms in scientific development. Determining a new paper's academic importance requires judging its impact on the field's knowledge structure—whether it maintains the existing structure, connects different fields, or changes the knowledge structure. The specific judgment criteria are shown in Table 1.

[Figure 2: see original paper] shows the subject (theme) positioning of the target paper across various databases and platforms.

4 Case Study

4.1 Research Object and Data Sources

The research sample is an actual representative work evaluation request from Tianjin Normal University's professional title evaluation process—a paper indexed in *Science* Citation Index. The author is from the College of Physics and Materials Science at Tianjin Normal University, specializing in astrophysics. Using the target paper and its citing literature as data sources, we conducted various performance indicator analyses, subject positioning, citation context analysis, and multi-dimensional scientometric analysis.

4.2 Evaluation Results

The target paper was a highly cited paper in its publication year. Subject positioning of academic papers can be achieved through the Leiden University Centre for Science and Technology Studies' clustering based on direct citation relationships, which assigns each paper to a micro-level subject field [27]. The CWTS classification system uses citation relationships between Article and Review document types in the Web of Science Core Collection to construct micro, meso, and macro-level classification themes [28]. Dimensions, a new scientific data platform, deeply indexes full texts to assign disciplines or research fields to each document [29].

[Figure 3: see original paper] shows the concept tree spatial distribution of subject terms in the target paper's citing literature. The concept tree provides hierarchical knowledge organization, making significant concepts appear as nodes near the root and contextual attributes appear in positions away from the root. This hierarchical structure differs from the parallel structure provided by keyword co-occurrence analysis.

[Figure 4: see original paper] illustrates the social network structure of the target paper. Betweenness centrality measures the extent to which a point lies on paths between other nodes in the graph. In co-authorship networks, authors in central positions play important roles in scientific dissemination. In the space sciences, which have relatively small interdisciplinary collaboration, betweenness centrality is generally small. The target paper's betweenness centrality is 0.05.

[Figure 5: see original paper] shows the target paper's distribution over time in the alluvial diagram. Alluvial diagrams visualize how research themes evolve along the time dimension, showing how theme streams split into multiple branches or converge. The longest time-spanning stream is highlighted in red, indicating the most prominent representative in the cluster. Although the target paper spans a long time in the stream (2006-2018), it did not become a representative paper in smaller clusters each year.

The comprehensive evaluation results for the target paper are shown in Table 2. Based on scientometric methods, these judgments provide relative and fuzzy grades that help us understand the academic status of the target paper and its likelihood of future attention and citation, providing a quick yet reliable reference for the academic community's final decisions.

5 Conclusion and Outlook

As long as there is demand for evaluating large numbers of scientific achievements, deviations and contradictions will continue to occur. No mechanism can fully explain the complexity of scientific development, and no evaluation method can precisely assess every academic achievement. The continuous improvement of evaluation methods aims to approach objective truth infinitely.

The original intention of the P-4Cs performance-semantic four-dimensional scalable evaluation system is to avoid erroneous value judgments that may arise from indicator-only evaluation methods. By delving into the content of target papers and comprehensively displaying academic achievements, it provides the scientific community and academic community with descriptive evaluations that save considerable time and effort.

While this paper focuses on research papers, the core idea of the P-4Cs evaluation system—that evaluation content should combine external performance indicators with internal academic value analysis—also applies to other achievements such as monographs and software. The system is replicable at the meso-level in terms of indicators and criteria, but at the micro-level, specific analysis is required for different disciplines and research fields. It has certain limitations for humanities and social sciences where citation phenomena are not prominent.

References

[2] Ministry of Human Resources and Social Security, Ministry of Education. Guiding Opinions on Deepening the Reform of Professional Title System for University Teachers [EB/OL]. [2023-09-02]. https://www.gov.cn/zhengce/zhengceku/2021-01/27/content_{5583094}.htm.

[3] Hirsch J E. An index to quantify an individual's scientific research output[J]. Proceedings of the National Academy of Sciences of the United States of America, 2005(46): 16569-16572.

[4] Egghe L. Theory and practice of the g-index[J]. Scientometrics, 2006(1): 131-152.

[5] Bihari A, Tripathi S, Deepak A. A review of h-index and alternative indices[J]. Journal of Information Science, 2023(3): 624-642.

[6] Melero R. Altmetrics—a complement to conventional metrics[J]. Biochemia Medica, 2015(2): 152-160.

[7] Piwowar H. Altmetrics: Value all research products[J]. Nature, 2013(7431): 159.

[8] Sud P, Thelwall M. Evaluating altmetrics[J]. Scientometrics, 2014(2): 1131-1143.

[9] Yogatama D, Heilman M, O'Connor B, et al. Predicting a scientific community's response to an article[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2011.

[10] Yu X, Gao Y. Single academic paper evaluation: A review[J]. Journal of Library Science in China, 2018(1): 50-61.

[11] Livne A, Adar E, Teevan J, et al. Predicting citation counts using graph and content mining[C]//Proceedings of the Workshop on Computational Scientometrics: Theory and Applications. New York: ACM, 2013: 5-8.

[12] McNamara D, Wong P, Christen P, et al. Predicting the impact of academic papers using citation network features[C]//Trends and Applications of Knowledge Discovery and Data Mining. Berlin, Heidelberg: Springer, 2013: 14-25.

[13] Upham P, Rosenkopf L, Ungar L. Positioning knowledge: Schools of thought and new knowledge creation[J]. Scientometrics, 2010(2): 555-581.

[14] Ye J. Construction of paper evaluation index system based on social network analysis method[J]. Library and Information Service, 2017(4): 97-102, 144.

[15] Zhang X. Research on identifying academic paradigms based on visualization techniques[J]. Library and Information Service, 2010(14): 28-31.

[16] Ye J. The construction and improvement of the academic evaluation system with Chinese characteristics for philosophy and social sciences[J]. Social Sciences Abstracts, 2021(7): 4-6.

[17] Shang H, Liu Z. Evaluating academic papers using academic influence: Two new indicators about academic inheritance effect and long-term citation[J]. Journal of Library Science in China, 2022(3): 133-146.

[18] Zhang H. Evaluating scientific research achievements from the perspective of fluid mechanics development[J]. Chinese Science Bulletin, 2011(26): 2853-2860.

[19] Kuhn T S. The Structure of Scientific Revolutions[M]. Chicago: University of Chicago Press, 1962: 92-110.

[20] Burt R S. Structural holes and good ideas[J]. American Journal of Sociology, 2004(2): 349-399.

[21] Burt R S. Structural Holes: The Social Structure of Competition[M]. Harvard University Press, 1992: 15-29.

[22] Chen C, Horowitz M, et al. Towards an explanatory and computational theory of scientific discovery[J]. Journal of Informetrics, 2009(3): 191-209.

[23] Chen G, Liu X. Emerging trends in regenerative medicine: A scientometric analysis using CiteSpace[J]. Expert Opinion on Biological Therapy, 2012(5): 593-608.

[24] CWTS Leiden Ranking[EB/OL]. [2023-03-07]. https://www.leidenranking.com/information/fields/.

[25] Introducing Citation Topics in InCites[EB/OL]. [2023-03-07]. https://clarivate.com/blog/introducing-citation-topics/.

[26] Dimensions: A new scientific data platform[EB/OL]. [2023-03-07]. http://blog.sciencenet.cn/blog-1792012-1113138.html.

[27] Pirolli P. Information Foraging Theory: Adaptive Interaction with Information[M]. Oxford: Oxford University Press, 2007: 3-29.

[28] Kleinberg J. Bursty and hierarchical structure in streams[C]//Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2002.

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