Research Progress on Evaluation Methods for Academic Journal Impact (Postprint)

Authors: Zhang Huiling, Dong Kun, Xu Haiyun

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

Abstract

[Purpose/Significance] Journal impact evaluation represents a critical component and application domain within scientometrics research, bearing significant importance for optimizing journal quality and facilitating scientific communication. [Method/Process] This paper reviews the research progress on journal impact evaluation methodologies from both domestic and international perspectives, encompassing traditional indicators, impact factor series metrics, the h-index and its derivatives, PageRank-type metrics and their variants, multi-factor comprehensive evaluation approaches, social media-based journal impact indicators, and interdisciplinary journal evaluation methods, thereby summarizing the characteristics and deficiencies of existing academic journal impact evaluation frameworks. [Results/Conclusion] Current evaluation indicators demonstrate insufficient integration, inadequately address citation skewness and self-citation issues, and the viability of emerging social media-based evaluation methodologies remains debatable. Future research should prioritize the deep integration of evaluation indicators, diversify approaches for mitigating citation skewness, standardize social media-based journal evaluation protocols, and systematize cross-disciplinary journal evaluation methodologies.

Full Text

Preamble

Zhang Huiling[1,2], **Dong Kun**[2,3], **Xu Haiyun**[2,4] [1]Shanxi University of Finance and Economics, Taiyuan 030006 [2]Chengdu Documentation and Information Center, Chinese Academy of Sciences, Chengdu 610041 [3]University of Chinese Academy of Sciences, Beijing 100190 [4]Institute of Scientific and Technical Information of China, Beijing 100038

Abstract

[**Purpose/Significance**] Journal impact evaluation constitutes a crucial component and application domain of scientometrics research, holding significant importance for optimizing journal quality and promoting scientific communication. [**Method/Process**] This paper reviews research progress on journal impact evaluation methods both domestically and internationally, examining traditional indicators, impact factor series metrics, H-index and its derivatives, PageRank-like and derivative indicators, multi-factor comprehensive evaluation methods, social media-based journal impact evaluation metrics, and interdisciplinary journal evaluation approaches. We summarize the characteristics and deficiencies of existing academic journal impact evaluation methods. [**Result/Conclusion**] Current evaluation indicators exhibit poor integration, seldom consider citation skewness and self-citation issues, and the feasibility of emerging social media-based evaluation methods remains questionable. Future research should emphasize deep integration of evaluation indicators, diversify methods for addressing citation skewness, standardize social media journal evaluation methods, and systematize cross-disciplinary journal evaluation approaches.

Academic journals showcase research achievements in various fields, representing academic advancement and methodological innovation within disciplines. Academic journal evaluation involves comprehensive assessment of journal quality and influence, representing an important type of scientific evaluation and a vital pathway for knowledge management. Given the vast number and variety of academic journals with uneven article quality, evaluating academic journals essentially constitutes a process of "separating the wheat from the chaff." Without proper evaluation, determining journal quality becomes impossible, and researchers cannot quickly grasp the core ideas at the forefront of their disciplines. The formulation and selection of evaluation methods are equally important; poor methodological validity affects evaluation accuracy and consequently influences public judgment of journal value. In an information environment with numerous academic journals, quantitative analysis methods such as scientometrics have played a significant role in academic journal evaluation. Employing scientometric indicators can avoid the time-consuming nature of peer expert evaluation while increasing objectivity. However, existing indicators each have their own advantages and disadvantages, and the scientific validity of academic journal evaluation needs further improvement.

Academic journal evaluation includes various forms, which can be categorized into quality evaluation, reputation evaluation, and impact evaluation based on different emphases. Quality evaluation primarily concerns publication content and editorial processing. Reputation evaluation focuses more on peer recognition. Impact evaluation assesses how journals influence knowledge innovation in their fields, including knowledge dissemination and creation. This study concentrates on journal impact evaluation, systematically reviewing existing methods both domestically and internationally. We classify current evaluation method systems into three major categories based on methodological characteristics and chronological order: peer review, scientometric methods, and altmetrics. Scientometric methods can be further divided into traditional indicators, impact factor series metrics, H-index and its derivatives, PageRank-like indicators, and multi-factor comprehensive evaluation methods (see Figure 1 [Figure 1: see original paper]). Meanwhile, interdisciplinary journal evaluation represents a challenging and hot topic in journal assessment.

2 Academic Journal Impact Evaluation Methods

2.1 Peer Review Method

Peer review emerged far earlier than scientometric approaches. In 1665, when the Royal Society's *Philosophical Transactions* was established, the prototype of peer review had already formed [1]. Geng Yanhui [2] studied journal peer review methods and types, combining peer review with modern technology to give it a data-driven dimension. Since peer review involves evaluation by authoritative experts and disciplinary leaders, it offers comprehensive perspectives and depth. However, peer review [3] has significant drawbacks: (1) The evaluation process carries subjective bias. (2) The evaluation cycle is long and inefficient. (3) Peer review can involve professional jealousy and discrimination against authors from certain countries or regions. Additionally, some reviewers exhibit unethical behavior, such as reviewing articles outside their expertise merely to "enrich" their resumes or showing favoritism toward friends' submissions. These inherent defects have led to the gradual adoption of quantitative scientometric methods in academia.

2.2 Scientometric Methods

In the mid-to-late 20th century, scientific and technological metrics gained prominence among scholars. Garfield [4] published "Citation Analysis as a Tool in Journal Evaluation" in *Science* magazine in 1972, marking the beginning of widely researched scientific evaluation methods using citation metrics. Scientometric evaluation represents the primary quantitative approach in scientific assessment. This paper focuses on journal impact evaluation metrics, systematically reviewing them across five categories: traditional indicators, impact factor models, H-index models, PageRank-like models, and multi-factor comprehensive evaluation models (see Figure 2 [Figure 2: see original paper]).

2.2.1 Traditional Indicators Traditional indicators were earliest employed in journal evaluation research, primarily using metrics from the *Chinese S&T Journal Citation Reports* [5] such as total citation frequency, immediacy index, cited half-life, funded article ratio, non-self-citation rate, average citation rate, regional distribution, and institutional distribution. Based on data foundations, these can be divided into citation-based indicators and other data-based indicators. Citation-based indicators evaluate various disciplines from temporal characteristics of cited journals, with different citation time points reflecting different features. For instance, total citation frequency counts all citations since a journal's inception, highlighting authoritative journals—higher values indicate greater authority but are also influenced by publication history (older journals tend to have more citations). The immediacy index reflects citation of articles in the publication year, emphasizing immediate response rate, though some articles exhibit "sleeping beauty" phenomena [6] with delayed high-frequency citations. Cited half-life measures the number of years from the current year backward until reaching 50% of total citations, assessing journal aging speed across disciplines, though it's susceptible to journal characteristics (research journals have longer half-lives than popular science journals, not necessarily indicating greater influence). Other data-based indicators include funded articles, regional and institutional metrics, evaluating journal impact from different perspectives. However, traditional indicators fail to account for citation lag effects and journal attributes, yielding one-sided impact measurements.

2.2.2 Impact Factor Series Metrics In the 1970s, based on Bradford's Law, the Impact Factor (IF) emerged. In 1998, Eugene Garfield, director of the Institute for Scientific Information, detailed the impact factor and its development in *Science*, explaining that his original purpose was to evaluate and select journals for *Current Contents*. Garfield's analysis of SCI data revealed that journal article citations peak two years after publication, leading him to adopt a two-year time window. Essentially, impact factor represents the average citation rate of papers published in the previous two years. The calculation formula [7] is:

$$IF = \frac{\text{The number of citations in the current year to articles published in the previous two years}}{\text{The number of citable articles published in the previous two years}}$$

Among numerous journal evaluation indicators, impact factor has become indispensable and represents one of the most representative metrics in JCR. Pendlebury [8] summarized impact factor's advantages in journal evaluation: simple and understandable formula; focus on short-term and recent research; insensitivity to citation errors in individual papers; and ability to construct time series data revealing influence patterns.

However, impact factor's prominent status has attracted extensive research on its limitations. Zhang Jiyu [9] noted that impact factor is influenced by both discipline characteristics and journal publication frequency/capacity, making fair evaluation difficult. Vinkler and Rousseau [10-11] argued that impact factor's formula is overly simplistic, as journal impact cannot be fully measured by average citation rate alone, and Garfield's impact factor fails to reflect multi-dimensional impact. The numerator may inflate due to various citation types, giving high impact factors to journals publishing many review articles.

Researchers have continuously sought improvements, generating multiple modified impact factor metrics, including weighted impact factor, eigenfactor impact factor, recent contribution index, historical impact factor, diachronic/synchronous impact factor, non-parametric statistical methods, and SNIP index. Table 1 summarizes these modified indicators' definitions, characteristics, and limitations.

Table 1. Summary of Impact Factor-Based Journal Impact Evaluation Indicators

Indicator	Proposer	Definition	Characteristics	Limitations
5-year Impact Factor [12]	Garfield, Reuters	Changes time window to 5 years	Compensates for time factor effects	Cannot overcome discipline influence
Cumulative Impact Factor	K. Zyczkowski	Uses 15-year and 7-year windows	Corrects short citation windows	Cannot overcome discipline influence
Weighted Impact Factor	P. Vinkler	Assigns weights to differentiate citation quality	Improves equal treatment of all citations	Data-sensitive, unstable results
Eigenfactor Impact Factor [13-14]	Zhang et al.	Uses eigenvector centrality from adjacency matrix	Improves arithmetic mean representation	Relationship with other metrics unverified
Recent Contribution Index (cci) [16]	P. Vinkler	Compares journal citations to total field citations	Measures short-term impact	Cannot measure long-term impact
Historical Impact Factor [17]	Yu Liping	Calculates citations from past time points	Improves short-term focus	Weak cross-disciplinary application
Diachronic/Synchronous Impact Factor [18]	Ma Ruimin	Cumulative H-index within citation half-life	Improves short-term focus	Weak cross-disciplinary application

Indicator	Proposer	Definition	Characteristics	Limitations
Non-parametric Method [19]	Xu Haiyun	Weights different citation frequency segments	Improves arithmetic mean representation	Ignores time factor
SNIP Index [20]	H.F. Moed	Ratio of average citations to field citation potential	Improves field citation potential consideration	Doesn't exclude self-citations
CiteScore	A.R. Memon	3-year citation average	Expanded coverage	Favors open-access journals

These improvements can be categorized as: (1) compensating for short-term focus, (2) improving arithmetic mean representation, and (3) optimizing equal treatment of citation quality. While some perform well within specific fields, others only partially mitigate impact factor's limitations—improving one drawback while retaining others. For instance, the recent contribution index cannot measure long-term impact or avoid impact factor inflation.

Time factors remain a challenge in journal evaluation. Different time windows reflect different evaluation emphases. Glänzel and Moed [22] noted that slower-developing disciplines with later citation peaks cannot be fairly evaluated using a 2-year window. In 2009, Thomson Reuters introduced a 5-year window (IF5) to address IF's limitations. Jacsó [23] compared IF and IF5, finding that IF5 better compensates for IF's defects, particularly for journals with citation peaks beyond 3 years [24-26].

Current evaluation indicators typically use 2-year, 3-year, or 5-year windows. Comparative analysis [25-28] and citation peak patterns [29-32] reveal different evaluation foci: short-term indicators (2-year) emphasize authority and novelty; 3-year indicators (CiteScore, SJR, SNIP) focus on disciplinary characteristics; 5-year indicators (Eigenfactor, H5 index) concentrate on citation balance, encompassing both peak and decay periods for more comprehensive assessment.

2.2.3 H-index and Derivative Evaluation Indicators The H-index was first proposed in 2005 by physicist Jorge Hirsch at UC San Diego to quantify individual research output [33]. Although initially designed for researchers, it was quickly adapted for journal evaluation by Braun et al. [34]. Egghe [35] developed a mathematical model based on Lotka's Law, extending its application beyond individual scientists to information production processes. The H-index incorporates both publication count and citation quantity, offering more comprehensive reflection of journal impact than impact factor alone, with relatively simple calculation. However, it has limitations: insensitivity to both highly and poorly cited articles (only citations to H-core papers increase the index), and potential for significant drops if borderline papers fall out of the H-core. Additionally, H-index grows over time, focusing only on long-term citation accumulation.

To address these issues, scholars have proposed various extensions from different perspectives. These can be categorized into H-index extensions, H-core improvements, and other novel extensions.

(1) H-index Extensions. Academic proposals include G-index, W-index, H-like indices, and P-index (see Table 2).

Table 2. Summary of H-index Extensions

Indicator	Proposer	Definition	Characteristics	Limitations
G-index [36]	L. Egghe	Largest rank g where top g papers have at least g^2 citations	Improves high-citation sensitivity	Ignores time factor
W-index [37-40]	G.J. Woeginger; R. Rousseau	Largest w where w papers have $ $10w citations	Improves high-citation sensitivity	Originally for authors, may not suit journals
H-like index [42]	M. Kosmulski	Largest H(x) where top H(x) papers have $\geq[H(x)^2]$ citations	Improves high-citation sensitivity	Not for short-term/technical journals
P-index [43-45]	G. Prathap	Square root of total citations divided by total publications	Improves high-citation sensitivity	Ignores time factor, citation skewness

These indices aim to improve H-index's insensitivity to highly cited articles, focusing on citation counts to differentiate high-impact literature. However, most ignore temporal factors, remaining insensitive to short-term/technical journals. The W-index, originally for author evaluation, may reduce accuracy when directly applied to journals due to different evaluation factors.

(2) H-core Improvement Indicators. Jin et al. noted that H-index lacks discrimination in evaluating researchers, introducing the "Hirsch core" (H-core) concept [47]—the highly cited paper zone formed by the top H papers in citation ranking. Table 3 summarizes H-core-based improvements.

Table 3. Summary of H-core Improvement Indicators

Indicator	Proposer	Definition	Characteristics	Limitations
R-index [47]	Jin et al.	Square root of total citations in H-core	Measures citation intensity in H-core	Doesn't deeply measure H-core papers
AR-index [47]	Jin et al.	Age-weighted average of H-core citations	Improves time-only-increasing issue	Ignores low discrimination problem
A-index [47]	Jin et al.	Average citations of H-core papers	Improves high-citation insensitivity	Only measures H-core average
m-index [48]	L. Bornmann	Median citations of H-core papers	Improves high-citation insensitivity	Skewed data may cause errors
π-index series [49-50]	P. Vinkler	$0.01 \times C(p\pi)$ where $p\pi = p1/2$	Measures H-core impact	Overly sensitive to single high-cited paper

These H-core-based improvements primarily address H-index's insensitivity to highly cited articles, recognizing citation quantity's importance in journal evaluation. While square roots can increase sensitivity to high-cited articles, excessive sensitivity to single papers can distort results.

Overall, most indices target H-index's insensitivity to highly cited articles or its long-term focus, but each only optimizes one limitation. This demonstrates that neither impact factor nor H-index alone can comprehensively measure journal impact, necessitating integrated multi-factor indicators.

2.2.4 PageRank-like and Derivative Indicators Since 2012, researchers have introduced Google's webpage ranking algorithm into scientometrics, creating various PageRank-like journal impact evaluation methods. PageRank determines webpage relevance/importance through recursive calculation of all incoming links. In academia, journal influence is divided into popularity and prestige [51], with Pinski and Narin first proposing using the principal eigenvalue of journal citation matrices for ranking [51]. Journal prestige evaluation primarily involves assigning different weights to citations. JCR Eigenfactor Score and Scopus's SCImago Journal Rank (SJR) are major prestige metrics, both based on PageRank.

(1) JCR Eigenfactor. To address impact factor and H-index's equal treatment of all citations, Bergstrom et al. proposed Eigenfactor Score [52], arguing that citations from better journals carry more weight than those from average journals. The basic assumption: highly cited journals confer high impact. It calculates eigenvector centrality in citation networks, avoiding isolated nodes [57-58]. The algorithm excludes self-citations, considering both citation quantity and quality [59-61]. Liu Yanhua [53] detailed Eigenfactor's definition and principles, noting it excludes self-citations to avoid measurement bias and clearly shows citation differences between journals. However, Eigenfactor cannot fully characterize journal impact due to differences between journals and webpages, shows weak disciplinary sensitivity, and has poor result discrimination.

(2) SJR Index. The SJR index [54], developed by a Spanish research group using Scopus data, employs a PageRank-like algorithm. It calculates journal importance through inter-journal citation relationships, interpreting journal A's citations to journal B as "votes" from A to B. Journal B's score from A equals A's score (representing A's importance) multiplied by vote count. Zhao Xing [55-56] analyzed SJR's advantages (broader source journals, more accurate than Eigenfactor) and disadvantages (calculates impact from all article types, not just research articles, allowing low-quality review articles to inflate scores).

Both Eigenfactor and SJR inherit PageRank's advantages, using network analysis to measure impact—higher network citations indicate greater influence. However, they share defects: high network citations from low-quality articles are deemed high-impact, and network randomness/variability affects measurements.

**(3) Authority Factor. The 2011 Chinese S&T Journal Citation Reports [62] introduced a new metric—Authority Factor. Addressing citation type diversity, the Chinese Institute of Scientific and Technical Information proposed PrestigeRank algorithm based on PageRank, considering both citation type diversity and quantity to reflect journal authority. The algorithm introduces virtual nodes for missing references to form complete citation networks, preventing bias from incomplete data. Su Cheng et al. [63] empirically analyzed Authority Factor against total citations, finding it can differentiate citation importance, 弥补了 impact factor's focus on quantity over quality, and provides more comprehensive authority measurement.

2.2.5 Multi-factor Comprehensive Evaluation Methods Recent years have seen numerous journal impact indicators, but each has limitations and rarely considers citation distribution skewness. Research has shifted from improving individual metrics to developing multi-factor comprehensive models that account for citation distributions. Table 4 details new multi-factor indicators: HIF index, PRP index, and f(x) index.

Table 4. Summary of Multi-factor Comprehensive Evaluation Indicators

Indicator	Proposer	Definition	Characteristics	Limitations
HIF Index [64]	Li Chao	HIFi = WH × PiH + WIF × PiIF	Combines H-index and IF advantages	Doesn't exclude self-citations
PRP Index [65-66]	P. Vinkler	PRP(i,j,F) = 100 × (r(i,j,F)-1)/(P-1)	Cross-disciplinary evaluation	Doesn't consider citation quality
f(x) Index [67]	Shao Zuoyun	f(x) = log x (a>1)	Divides citations into weighted intervals	Assumes equal citation weights
I3 Index [68]	L. Leydesdorff	I3(i) = Σf(Xi) × Xi	Considers percentile rankings	Doesn't address citation skewness
CI Index [69-70]	Wu Junhong et al.	Ic = √(2 - $(1\text{-IF})^2$ + $(1\text{-TC})^2$)	Combines normalized IF and citations	Ignores long-term citations

These models primarily treat citation distribution as the research object, dividing it into intervals with different weights. The I3 index partitions citations into high, low, and H-core zones, combining H-index advantages with impact factor weighting to avoid skewness-induced bias while reducing single-dimension limitations. The HIF index integrates impact factor's short-term sensitivity with H-index's long-term feasibility. The CI index reflects both citation quantity and quality. Thus, multi-factor methods consider both citation skewness and timeliness, providing comprehensive multi-angle evaluation that 弥补 single-metric limitations.

2.3 Social Media-Based Journal Impact Evaluation Models

Social media-based evaluation uses social networks to measure academic impact, leveraging user behaviors (comments, likes, shares) on platforms like Facebook and Twitter. This approach, also called altmetrics, was first termed by J. Priem on Twitter in 2010 to supplement traditional citation-based metrics. Altmetrics indicators include usage, captures, mentions, social media activity, and citations [71]. Influenced by altmetrics, companies like Altmetric.com, ImpactStory, and Plum Analytics have emerged, with Altmetric.com becoming a primary tool [72].

Evaluation media selection is crucial. Academic search engines [73], social media platforms [74-75], and traditional citation databases constitute main evaluation channels. Kousha [76] used Google Scholar data for journal evaluation. Zahedi [77] found positive correlation between Mendeley readership and citation counts. Thelwall [78] evaluated ten social web services including Twitter. Shu Fei [79] studied Chinese papers' international influence via Twitter user behavior. Ji Fang evaluated journal impact through WeChat public platform interactions. Lin [80] examined Wikipedia's influence on journal evaluation. These studies demonstrate social media's applicability and validity for evaluating papers, scholars, and journals.

To enrich evaluation media diversity, based on Kantar Media CIC's social platform classification and considering journal dissemination feasibility, we identify seven suitable platform categories (excluding gaming and consumer platforms) as shown in Table 5 .

Table 5. Social Media Platforms Suitable for Journal Evaluation

Category	Platforms	Evaluation Behaviors	Advantages	Limitations
Blog	Sina Blog, Blogbus, Blogger, Bola	Comments, likes, shares	Professional	Narrow evaluation scope
Microblog	Sina Weibo, Tencent Weibo, Twitter	Comments, likes, shares, bookmarks	Broad reach	Low professionalism
Encyclopedia	Baidu Baike, MBAlib, Hudong	Edit counts	Single evaluation angle	False information possible
Q&A	Zhihu, Baidu Zhidao, Answers	Comments, follows, bookmarks	Professional	Narrow scope
Social Bookmarking	QQ Bookmarks, Delicious	Likes, comments, bookmarks	Broad reach	Low professionalism
Forum	Tianya, Mop, BigBoards	Comments, bookmarks	Broad reach	False information possible
Social Network	Douban, Renren, Facebook	Comments, bookmarks	Broad reach	False information, low professionalism

Each platform has distinct characteristics. Microblog and social network platforms have broad reach but low professionalism and potential false information. Q&A and forum platforms have uneven quality, 容易产生偏差. Encyclopedia platforms only allow evaluation through edit counts, offering limited perspective. Current research focuses heavily on Twitter, Facebook, and Mendeley, with less attention to Q&A, forums, and social bookmarking platforms—areas warranting future investigation.

Social media evaluation 弥补 traditional metrics' focus on quantity over quality, incorporating user perspectives for more authentic and comprehensive assessment.

2.4 Interdisciplinary Journal Evaluation

Interdisciplinary journal evaluation has become a critical challenge. Most journals are organized by discipline, but cross-disciplinary research has created many interdisciplinary fields where articles belong to multiple categories, each with different core journals and evaluation standards. Applying uniform standards to interdisciplinary and single-discipline journals violates the principle of comparable comparison, making suitable interdisciplinary evaluation methods imperative.

Interdisciplinary journal research involves identification methods and evaluation indicators. Interdisciplinary journals exhibit characteristics of disciplinary aggregation, balance, and diversity [81], requiring more than traditional metrics. Identification methods consider information entropy, Euclidean distance, and network indicators [82], proposing diversity metrics (NS, COC [83]), balance metrics (SE, GC [84]), and aggregation metrics (BC, ND). However, most target articles rather than journals, with few interdisciplinary identification indicators at journal level.

Interdisciplinary evaluation indicators include impact factor percentiles [85], PR8 index [86], I3 index, and Pnew [87], which normalize across disciplines to avoid skewness differences. These not only 修正 traditional indicators but must also incorporate interdisciplinary characteristics.

3 Research Deficiencies and Future Prospects

3.1 Research Deficiencies

Overall, journal impact evaluation systems are transitioning from breadth to depth with richer methods, yet fundamental issues remain unresolved:

(1) Poor integration among numerous derivative indicators. Current methods include peer review, scientometrics, and altmetrics, or quantitative, qualitative, and mixed approaches. While abundant indicators exist (citation-based, network-based, social media-based), their multiplicity and different foci create randomness in selection, yielding different impact rankings. Evaluation indicators should shift from depth to integration, using fewer metrics to measure multiple impact dimensions.

(2) Most indicators ignore citation skewness and self-citation. Citation skewness refers to non-power-law distributions and long-tail phenomena. Using only mean citation counts for non-power-law distributions is inappropriate. Emerging field journals may have low citations but high potential value. Most indicators also fail to address self-citation, which artificially inflates measurements. Table 6 summarizes deficiencies in some indicators.

Table 6. Summary of Deficiencies in Some Journal Impact Evaluation Indicators

Indicator	Deficiency
f(x) Index	Assumes equal citation weights, doesn't exclude self-citations
SNIP Index	Doesn't exclude self-citations
Historical Impact Factor	Ignores citation skewness
π-index Series	Ignores citation skewness, doesn't exclude self-citations

(3) Feasibility of social media evaluation remains uncertain. While social media evaluation's interactivity enriches the system and incorporates user perspectives, user professionalism varies, causing potential injustice. Some users evaluate based on preference rather than content understanding, reducing authenticity. Selecting optimal platforms and qualified users requires further research.

3.2 Future Research Prospects

Current evaluation systems are maturing, emphasizing user experience and evaluation depth while modifying early indicators. However, inherent defects persist, suggesting several directions:

(1) Deep integration of existing indicators. Future indicators must deeply analyze each metric's focus and limitations for targeted improvement. Since different metrics reflect different impact dimensions (citation frequency, network structure, user behavior), developing scientific, applicable comprehensive indicators is crucial. Integration requires understanding mathematical characteristics between indicators to find appropriate fusion mechanisms beyond simple combination.

(2) Diversification of citation skewness improvement methods. Current methods primarily involve partitioning and normalizing citation distributions. Given different indicator foci, diverse skewness handling methods are essential. Solutions require selecting normalization methods based on skewness characteristics, using interval weights to reduce high-citation zone influence, and prioritizing effectiveness over simplicity.

(3) Standardization of social media evaluation methods. Social media evaluation 弥补 traditional metrics' citation focus but faces boundary issues: should evaluators include only researchers or also scholars and the public? How to handle false data and integrate results across platforms? Future research should: (1) use digital cloud technology to build group evaluation platforms with clear evaluator qualifications, (2) establish evaluation rules and filter invalid data through database matching, (3) normalize results across platforms using weights, and (4) analyze indicator correlation and validity to establish robust social media metrics.

(4) Systematization of cross-disciplinary evaluation methods. With flourishing interdisciplinary research, cross-disciplinary journal evaluation has become important. Current approaches remain preliminary, facing two main challenges: identification methods and evaluation indicator systems. Most identification methods operate at article level; accurate, efficient journal-level identification is needed. Future research should borrow from interdisciplinary studies, applying diversity, balance, and aggregation characteristics to journal identification and evaluation system development.

References

[1] ZUCKERMAN H, MERTON RK. Patterns of evaluation in science: institutionalisation, structure and functions of the referee system. Minerva, 1971, 9(1): 66-100.

[2] GENG Yanhui, WANG Lixin, ZHU Xiaohua. Review of journal peer review research. Bianjizhiyou (Editor's Friend), 2015(5): 48-52.

[3] WAN Hao, TAN Zongying, ZHU Xiangli. Comparative analysis of the roles of peer review and bibliometrics in research evaluation. Library and Information Service, 2017(1): 134-152.

[4] GARFIELD E. Citation analysis as a tool in journal evaluation. Science, 1972, 178(4060): 471-479.

[5] PAN Yuntao. 2011 Edition Chinese S&T Journal Citation Reports: Core Version. Beijing: Scientific and Technical Documentation Press, 2011.

[6] POLCHINSKI J. Dirichlet branes and ramond-ramond charges. Physical review letters, 1995, 75(26): 4724-4727.

[7] GARFIELD E. Journal impact factor: a brief review. CMAJ: Canadian medical association journal, 1999, 161(8): 979-980.

[8] PENDLEBURY DA. The use and misuse of journal metrics and other citation indicators. Archivum immunologiae et therapiae experimentalis, 2009, 57(1): 1-11.

[9] ZHANG Jiyu. Construction of academic journal influence and its evaluation index system. Journal of Shaanxi Normal University (Philosophy and Social Sciences Edition), 2010(5): 70-76.

[10] SCHUBERT A, GLÄNZEL W. Statistical reliability of comparisons based on the citation impact of scientific publications. Scientometrics, 1983, 5(1): 59-73.

[11] ROUSSEAU R. Why am I not cited or why are multi-authored papers more cited than others? Journal of documentation, 1992, 48(1): 79-80.

[12] GARFIELD E. Long-term vs. short-term journal impact: does it matter? Physiologist, 1998, 41(3): 113.

[13] ZHANG C, LIU X, XU Y, et al. Quality-structure index: a new metric to measure scientific journal influence. Journal of the American Society for Information Science & Technology, 2014, 62(4): 643-653.

[14] ZYCZKOWSKI K. Citation graph, weighted impact factors and performance indices. Scientometrics, 2010, 85(1): 301-315.

[15] XU Haiyun, FANG Shu, FU Xijin. Research on journal impact factor based on eigenvector centrality weighting. Information Studies: Theory & Application, 2011, 34(11): 108-112.

[16] VINKLER P. Introducing the current contribution index for characterizing the recent, relevant of journals. Scientometrics, 2009, 79(2): 409-420.

[17] YU Liping. Historical impact factor: a new stock evaluation index for academic journals. Library and Information Service, 2015, 59(2): 89-92.

[18] MA Ruimin. Research on academic journal influence evaluation—new practice based on diachronic perspective. Chinese Journal of Scientific and Technical Periodicals, 2014(11): 1397-1403.

[19] XU Haiyun, FANG Shu. Research on non-parametric statistical methods for journal influence evaluation. Library and Information Service, 2013, 57(5): 107-113.

[20] MOED HF. Measuring contextual citation impact of scientific journals. Journal of informetrics, 2009, 4(3): 265-277.

[21] SILVA JATD, MEMON AR. CiteScore: a cite for sore eyes, or a valuable, transparent metric? Scientometrics, 2017, 111(1): 1-4.

[22] GLÄNZEL W, MOED H. Journal impact measures in bibliometric research. Scientometrics, 2002, 53(2): 171-193.

[23] JACSÓ P. Five-year impact factor data in the journal citation reports. Online information review, 2009, 33(3): 603-614.

[24] ZHAO Xing. Analysis of JCR five-year impact factor. Journal of Library Science in China, 2010, 36(3): 120-126.

[25] WANG Juan, ZHENG Chunhou. Empirical analysis of JCR 5-year impact factor. Library and Information Service, 2012, 56(4): 143-147.

[26] SALAS, GRAFMAN J. Cortex 2009 5-year and 2-year Impact Factor: 4.1. Cortex, 2010, 46(9): 1069.

[27] ZHANG Liyi, YE Yan. Comparative analysis of Google H5-index and CiteScore index for journal quantitative evaluation—from knowledge diffusion perspective. Journal of Intelligence, 2017, 36(11): 175-179.

[28] HAO Ruoyang. Research on H5 index and its comparative analysis with impact factor. Information Science, 2017(6): 120-124.

[29] LINE M, SANDISON A. "Obsolescence" and changes in the use of literature with time. Journal of Documentation, 1974, 30: 283-350.

[30] EGGHE L, ROUSSEAU R. Aging, obsolescence, impact, growth, and utilization: definitions and relations. Journal of the American Society for Information Science, 2000, 51(11): 1004-1017.

[31] YU Liping, WANG Yan. Research on temporal distribution patterns of academic journal citation frequency—taking library, information and documentation science journals as examples. Information Science, 2016, 34(5): 159-162.

[32] FANG Hongling. Research on citation peak years and influencing factors of foreign journal papers—taking SSCI-indexed library and information science journals as examples. Chinese Journal of Scientific and Technical Periodicals, 2015, 26(11): 1200-1204.

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

[34] BRAUN T, GLÄNZEL W, SCHUBERT A. A Hirsch-type index for journals. Scientometrics, 2006, 69(1): 169-173.

[35] EGGHE L, ROUSSEAU R. An informetric model for the Hirsch-index. Scientometrics, 2006, 69(1): 121-129.

[36] EGGHE L. Theory and practice of the g-index. Scientometrics, 2006, 69(1): 131-152.

[37] WOEGINGER GJ. An axiomatic characterization of the Hirsch-index. Mathematical social sciences, 2008, 56(2): 224-232.

[38] WOEGINGER GJ. Asymmetry axiom for scientific impact indices. Journal of informetrics, 2008, 2(4): 298-303.

[39] ZHANG CT. The e-index, complementing the h-index for excess citations. PloS one, 2009, 4(5): 5429.

[40] YE FY, ROUSSEAU R. Probing the h-core: an investigation of the tail-core ratio of rank distributions. Scientometrics, 2010, 84(2): 431-439.

[41] KOSMULSKI M. A new Hirsch-type index saves time and works equally well as the original h-index. ISSI Newsletter, 2006, 2(3): 4-6.

[42] HUANG Huang. Brief analysis of HS index—a new journal evaluation indicator. Sci-Tech Information Development & Economy, 2013(21): 116-118.

[43] PRATHAP G. Is there a place for a mock h-index? Scientometrics, 2010, 85(2): 561-565.

[44] PRATHAP G. The iCE approach for journal evaluation. Scientometrics, 2010, 84(1): 153-165.

[45] WANG Chao, PENG Chaoqun. Effectiveness analysis of P-index in evaluating Chinese SCI journals. Chinese Journal of Scientific and Technical Periodicals, 2016(8): 893-899.

[46] AN Jing, LI Haiyan, XIA Xu. Correlation analysis between A-index and h-index—new indicators for journal evaluation. Science and Technology Management Research, 2012, 32(15): 264-266.

[47] JIN B, LIANG L, ROUSSEAU R, EGGHE L. The R- and AR-indices: complementing the h-index. Chinese science bulletin, 2007, 52(6): 855-863.

[48] BORNHMANN L, MUTZ R, DANIEL HD. Are there better indices for evaluation purposes than the h index? A comparison of nine different variants of the h index using data from biomedicine. Journal of the American Society for Information Science and Technology, 2008, 59(5): 830-837.

[49] VINKLER P. The π-index: a new indicator for assessing scientific impact. Journal of information science, 2009, 35(5): 602-612.

[50] VINKLER P. The πv-index: a new indicator to characterize the impact of journals. Scientometrics, 2010, 82(3): 461-475.

[51] PINSKI G, NARIN F. Citation influence for journal aggregates of scientific publications: theory, with application to the literature of physics. Information processing & management, 1976, 12(5): 297-312.

[52] BERGSTROM C. Measuring the value and prestige of scholarly journals. College and Research Libraries News, 2007, 68(5): 314-316.

[53] LIU Yanhua, HUA Weina. Eigenfactor—a new indicator for journal evaluation. Journal of Intelligence, 2010(7): 122-126.

[54] KRAJNA T. SCImago Journal & Country Rank. Polimeri, 2009, 29(4): 258-262.

[55] YANG Kang, LIU Mingzheng, ZHANG Xu. Research on SJR index and its comparative analysis with impact factor. Journal of Intelligence, 2009(11): 27-30.

[56] ZHAO Xing, GAO Xiaoqiang, TANG Yu. Comparison between SJR, impact factor, and h-index, and extension proposals for SJR. Journal of Academic

Libraries, 2009(2): 80-84.

[57] YU Liping, LI Shouwei. Characteristics and application analysis of normalized Eigenfactor. Chinese Journal of Scientific and Technical Periodicals, 2016, 27(9): 990-993.

[58] ZHAO Xing. Characteristic study of Eigenfactor—a new journal citation evaluation indicator—based on empirical analysis of Chinese journals. Information Studies: Theory & Application, 2009, 32(8): 53-56.

[59] BERGSTROM CT, WEST JD, WISEMAN MA. The eigenfactor metrics. Journal of the American Society for Information Science & Technology, 2008, 59(13): 2186-2188.

[60] DAVIS PM. Eigenfactor: does the principle of repeated improvement result in better estimates than raw citation counts? Journal of the American Society for Information Science and Technology, 2008, 59(5): 830-837.

[61] FERSHT A. The most influential journals: impact factor and eigenfactor. Proceedings of the National Academy of Sciences of the United States of America, 2009, 106(17): 6883-6884.

[62] JIANG Yongqing, QI Ping. Research on academic journal influence evaluation methods. China Soft Science, 2017(3): 178-185.

[63] SU Cheng, PAN Yuntao, MA Zheng, et al. Authority Factor: a new journal evaluation indicator. Acta Editologica, 2010, 22(4): 369-373.

[64] LI Chao. Mechanism and practice of "HIF index" in evaluating academic impact of scientific and technical journals. Information Studies: Theory & Application, 2011(7): 44-48.

[65] VINKLER P. Comparative rank assessment of journal articles. Journal of informetrics, 2013, 7(3): 712-717.

[66] LI Hua, ZHANG Xuemei, LIU Wenzhi. PRP indicator: a cross-disciplinary journal evaluation method and comparison with journal H-index. Journal of Intelligence, 2015(8): 60-64.

[67] SHAO Zuoyun, LI Xiuxia. f(x) index: a new indicator for evaluating journal academic influence. Chinese Journal of Scientific and Technical Periodicals, 2015(11): 1194-1199.

[68] LEYDESDORFF L. Alternatives to the journal impact factor: I3 and the top-10% (or top-25%?) of the most-highly cited papers. Scientometrics, 2012, 92(2): 355-365.

[69] WU Junhong, SUN Xiukun, SUN Juan, et al. Comparative study of journal influence index and impact factor in evaluating international journals. Acta Editologica, 2017, 29(5): 500-504.

[70] WU Junhong, XIAO Hong, ZHANG Yan, et al. Research on evaluation indicators for international influence of scientific and technical journals. Acta

Editologica, 2015, 27(3): 214-218.

[71] YOU Qingbin, TANG Shanhong. Altmetrics and its application prospects. Information Studies: Theory & Application, 2013, 36(12): 6-10.

[72] ZHAO Rongying, WANG Shaozhen, CHEN Zhiyi. Exploration of altmetrics and its analysis tools. Information Studies: Theory & Application, 2015, 38(6): 29-34.

[73] ZHANG Yang, WU Juanping, LANG Linfang. Research on journal evaluation based on different network data sources. Chinese Journal of Scientific and Technical Periodicals, 2017, 28(2): 176-183.

[74] YU Yisheng, ZHAO Yuehua. Journal influence evaluation indicator based on Twitter attention—taking top international library and information science journals as examples. Library and Information Service, 2016, 60(8): 99-105.

[75] JI Fang, ZHANG Xiaheng. Research on influence of academic journal WeChat public platform—empirical analysis based on 5 CSSCI source journals. Journal of Intelligence, 2016, 35(4): 147-151.

[76] KOUSHA K, THELWALL M. Sources of Google Scholar citations outside the science citation index: a comparison between four science disciplines. Scientometrics, 2008, 74(2): 273-294.

[77] ZAHEDI Z, COSTAS R, WOUTERS P. How well developed are altmetrics? A cross-disciplinary analysis of the presence of 'alternative metrics' in scientific publications. Scientometrics, 2014, 101(2): 1491-1513.

[78] THELWALL M, HAUSTEIN S, LARIVIERE V, et al. Do altmetrics work? Twitter and ten other social web services. PloS one, 2013, 8(5): 64841.

[79] SHU Fei, HAUSTEIN S, QUAN Wei. Study on the impact of Twitter on international attention to Chinese papers. Library Tribune, 2017, 37(6): 55-60.

[80] LIN J, FENNER M. Altmetrics in evolution: defining and redefining the ontology of article-level metrics. Information standards quarterly, 2013, 25(2): 20.

[81] XU Haiyun, YIN Chunxiao, GUO Ting, et al. Review of interdisciplinary research. Library and Information Service, 2015, 59(5): 119-127.

[82] STIRLING A. A general framework for analysing diversity in science, technology and society. Journal of the Royal Society Interface, 2007, 4(15): 707-719.

[83] PORTER AL, CHUBIN DE. An indicator of cross-disciplinary research. Scientometrics, 1985, 8(3/4): 161-176.

[84] SHANNON CE. A mathematical theory of communication (parts I and II). Bell system technical journal, 1949, 27(3): 379-423.

[85] YU Liping. Characteristic study of "impact factor percentile" indicator. Library and Information Service, 2016, 60(10): 103-107.

[86] LIU Xueli, WEI Yahui, SHENG Lina, et al. Journal PR8 index: a new interdisciplinary journal evaluation indicator and its empirical study. Library and Information Service, 2017, 61(11): 116-123.

[87] CHEN Weijing, ZHANG Yue, LIN Meifang. Research on P-index for journal evaluation from citation distribution perspective. Library and Information Service, 2017, 61(17): 122-130.

Author Contributions: Zhang Huiling: Research conception and paper writing Dong Kun: Research framework design and paper revision Xu Haiyun: Paper revision guidance

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