Postprint: Research on Altmetrics Academic Quality Evaluation Methods Based on Fairness Testing

Authors: Liu Hongxu, Wang Zheng

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

Abstract

[Purpose/Significance] To investigate the consistency of measurement results among Altmetrics indicators in disciplinary paper evaluation, this study attempts to construct a comprehensive literature evaluation model based on Altmetrics using existing measurement data, thereby establishing novel metrics for the utilization and reuse of evaluation data. [Method/Process] Altmetrics indicators for all articles published in the civil engineering discipline from 2009 to 2016 on PLoS ALMs were collected. Correlation analysis and principal component analysis were employed to partition 15 original indicators into 4 principal components, yielding an Altmetrics evaluation indicator system applicable to specific disciplines. Correlation analysis was performed between individual Altmetrics indicators and comprehensive evaluation indicators, and the "fairness test" method was utilized to explore temporal differences in the correlation of paper evaluation. [Results/Conclusion] The study revealed that after eliminating temporal effects using the "fairness test" method, the correlation of citation indicators increased, whereas indicators related to social media dissemination decreased. However, the Twitter indicator exhibited a contrary trend by increasing, and consistently maintained a high correlation with the F-value.

Full Text

A Study on Altmetrics Academic Quality Evaluation Method Based on "Fairness Test"

Liu Hongxu[1], **Wang Zheng**[2,3] [1]Tongji University Library, Shanghai 200092 [2]School of Public Administration, Northwest University, Xi'an 710127 [3]School of Information Management, Nanjing University, Nanjing 210023

Abstract

[**Purpose/Significance**] This study explores whether Altmetrics indicators produce consistent measurement results in disciplinary paper evaluation, attempting to construct a comprehensive literature evaluation model based on Altmetrics through existing measurement data, thereby creating new metrics for the use and reuse of evaluation data. [**Method/Process**] The study obtains Altmetrics indicators for all articles in civil engineering published on PLoS ALMs from 2009 to 2016. Using correlation analysis and principal component analysis, 15 original indicators are divided into four principal components to establish an Altmetrics evaluation indicator system applicable to specific disciplines. The correlation between single Altmetrics indicators and comprehensive evaluation indicators is analyzed, and the "fairness test" method is employed to explore differences in paper evaluation correlation over time. [**Result/Conclusion**] The findings reveal that after eliminating time effects using the "fairness test" method, the correlation of citation indicators increases while indicators related to social media communication decrease. However, the Twitter indicator shows an opposite trend, increasing instead and maintaining high correlation with the F-value.

---

1. Introduction

In October 2010, J. Priem, together with D. Taraborelli and others, published the "Altmetrics: A manifesto," formally introducing the term "Altmetrics" (Alternative Metrics) [1]. Since 2012, domestic scholars in library and information science and journal publishing have conducted theoretical and empirical explorations of Altmetrics. Domestic scholars have translated it as "alternative metrology," "supplementary metrology," "social media impact metrology," etc., but these translations are ambiguous in the domestic context. For instance, J. Liu argues that Altmetrics aims to replace the single use of citation-based methods for research impact evaluation, entering a diversified stage of research impact assessment, rather than replacing citation methods entirely [2]. Therefore, this paper retains the original term for reliability and simplicity. In 2013, the National Information Standards Organization announced a new two-phase project to research and develop community-based supplementary metrics or recommended practices [3]. In 2016, ProQuest's subsidiary ExLibris announced the integration of Altmetrics into the Summon discovery service [4], significantly improving user experience and promoting content discovery and integration.

Altmetrics helps fill knowledge gaps that citation indicators cannot express. Not only does Altmetrics encompass more data types than traditional bibliometrics, but it also operates on a larger scale. Although theoretical research and practice on Altmetrics remain in the exploratory stage, several forward-looking issues warrant investigation. In evaluating individual papers, Altmetrics compensates for shortcomings in peer review through its "crowd review" advantage

[7]. Regarding methodological superiority, Altmetrics indicators were not initially comparable with traditional citation metrics: first, they reflect different aspects of literature influence; second, Altmetrics indicators have a complementary relationship with traditional metrics. Altmetrics evaluation tools such as ImpactStory and Plum Analytics both incorporate citation counts from traditional bibliometrics as sources for impact assessment. In March 2017, Altmetric.com also included citation counts from Web of Science as an indicator. By combining with traditional citation metrics, Altmetrics indicators can more accurately assess literature impact.

Current research primarily focuses on validating correlations between Altmetrics indicators and other bibliometric measures, testing whether single Altmetrics indicators outperform comprehensive Altmetrics indicators in evaluating academic paper impact. This is typically done through peer review or other third-party indicators for validation. Existing studies have confirmed that Altmetrics is not isolated and holds potential for future applications [5-6]. However, single indicators remain defective for academic evaluation, as J. Priem notes that individual indicators must be combined, though not all alternative metrics achieve satisfactory results [16]. J. Lin proposed combining altmetrics with peer review, but subsequent implementation yielded limited success [17]. Chinese scholar Yu Houqiang also externalized academic evaluation results in network environments as "degree of engagement" and "depth of utilization" [18], but without providing specific measurement standards.

This study focuses on the consistency of measurement results among Altmetrics indicators themselves, employing principal component analysis and dimensionality reduction to construct a comprehensive literature evaluation model based on Altmetrics through existing measurement data, creating new metrics for evaluation data use and reuse.

2. Literature Review

2.1 Exploring Correlations Between Altmetrics and Traditional Indicators Researchers have conducted multi-angle attempts to verify correlations between Altmetrics and traditional indicators, yielding mixed results. Regarding correlation strength between traditional citation indicators and Altmetrics, K. Yan et al. [8] found that HTML views and PDF downloads follow a long-tail distribution with strong correlation. X. Wang et al. [9] discovered that citation counts correlate more significantly with PDF views than other indicators. For internal correlations among Altmetrics indicators, H. Alhoori et al. [10] found weak correlations between Altmetrics and varying attention across platforms for individual papers. Additionally, the relationship between Altmetrics and citation indicators has received widespread attention. L. Bornmann et al. [11] found low correlation between blogs, microblogs, and citation indicators, while bookmarks performed slightly better. R. Costas et al. [12] identified disciplinary differences, with Altmetrics being applicable only to recently published papers. A. Zuccala et al. [13] found no significant correlation between citation counts and reader comments in monographs. K. Chen et al. [14] verified weak correlations between citations and Altmetrics across different disciplines and regions. M. Thelwall et al. [15] demonstrated varying positive correlations between Altmetrics and citation indicators through specific models. J. Priem [16] argued that single indicators remain defective for academic evaluation and must be combined, though not all alternative metrics achieve satisfactory results.

2.2 Current Status of Article-Level Metrics Research Based on PLOS
Article-level metrics [19-20] and Altmetrics [21-22] are not synonymous, differing in connotation and research scope. The former attempts to measure individual paper impact at the article level, collecting various data sources during measurement, while the latter uses new data sources to measure multiple objects (such as a paper, journal, or individual scholar). Article-level metrics utilizes Altmetrics data sources and traditional sources to redefine paper impact at the article level. Altmetrics, in contrast, deeply examines the characteristics, patterns, and applications of various new data sources, not merely aggregating data sources at the paper level [23].

As early as 2009, C. Neylon and S. Wu [24] discussed the feasibility of article-level metrics (ALMs) using PLoS and F1000 as examples, addressing data sources and incentive mechanisms for expert commentary. PLOS's Altmetrics indicators primarily target PLOS publications, with ALMs indicators divided into five categories [25]: 1) Viewed (including PLOS Journals HTML, PDF, XML; PubMed Central HTML, PDF); 2) Cited (including CrossRef, Datacite, Europe PMC, PubMed Central, Scopus, Web of Science); 3) Saved (including CiteULike, Mendeley); 4) Discussed (including PLOS Comments, Facebook, Reddit, Twitter, Wikipedia); and 5) Recommended (including F1000Prime). Although PLOS has relatively few metrics, each data source is authoritative and widely used [26]. Wang Rui found [27] that PLoS ALMs was the earliest tool to implement Altmetrics theory for single-paper impact evaluation. Liu Xiaojuan et al. [28] discovered different dimensions of literature value based on PLOS ALM data, suggesting that evaluation should reasonably construct indicator systems according to their advantages, disadvantages, and applicable scopes.

2.3 Application of Altmetrics in Disciplinary Evaluation and Research Services Altmetrics indicators are not simply correlated or not with traditional citation analysis; disciplinary differences directly affect their correlation [29-30]. Additionally, Altmetrics tools can provide scholars with academic portfolio services, helping them shape more colorful professional images through corresponding evaluation services [31]. Some universities have attempted to combine Altmetrics for diversified disciplinary services. The University of Pittsburgh Library System piloted Altmetrics tools for university research output evaluation, establishing direct communication mechanisms with researchers to form multi-participant impact evaluation schemes [32]. Additionally, MIT's medical school library promoted the concrete application of Altmetric.com in institutional repositories, while the London School of Economics collaborated with library research support services to use Altmetrics to present social impact landscapes for researchers, helping them track online attention dynamics [33].

2.4 Application of "Fairness Test" Method in Paper Impact Evaluation M. Thelwall [15] argued that analyzing Altmetrics indicators and citation indicators requires eliminating time window effects on data sampling, proposing a "fairness test" method to place test data in a relatively fair environment. The fundamental problem with indicator analysis is data bias. For citations, papers from different disciplines may receive citation counts differing by orders of magnitude at the same time. Using appropriately corrected indicators (such as CNCI) can eliminate inter-disciplinary bias, which falls within the scope of "fairness test" processing. Similarly, when measuring impact across time within the same discipline, time effects must be eliminated to improve reliability in longitudinal studies. J. Priem and H. Piwowar [16] examined the distribution of social media events for PLoS articles over time, noting behavioral differences: citations, page views, and Wikipedia citations increase over time, while CiteULike, Mendeley, Delicious bookmarks, and F1000 ratings are relatively unaffected by publication period. Other indicators vary significantly due to data service changes and limitations, highlighting the need to consider indicator stability and issues in longitudinal studies. Most scholars are only interested in the latest articles in their fields, suggesting that papers receive most attention at publication, declining over time.

Specifically, in the "fairness test," each article is compared only with the two articles immediately preceding and following it (in the same journal). Only literature published at roughly the same time can be compared, minimizing citation delay and usage bias. Any slight advantage or disadvantage of articles published after the tested article should be offset by equivalent advantages or disadvantages with previously published articles. The test yields three possible outcomes: 1) Success: score higher than the average of adjacent articles AND citation score higher than adjacent articles' average, OR score lower than average AND citation score lower than average; 2) Failure: opposite of success conditions; 3) Invalid: all other cases. For each comparison pair, a list is created, allowing articles with positive measurement scores to be effectively monitored through this measurement.

F. Radicchi et al. [34-35] adopted the "fairness test" method to provide measurable evaluation criteria for inter-disciplinary comparison of academic journal evaluation indicators from a probability perspective. Using this method, they empirically analyzed that normalized fractional citation indicators are far more fair than journal impact factors for top 10% papers across disciplines. Chen Fuyou et al. [36] used relative journal impact factor and average percentile rank indicators to test the applicability of two discipline-normalized evaluation indicators across disciplines. Wang Rui et al. [37] used 273 high-altmetric papers as samples, applying the "fairness test" to eliminate time effects on citations, finding that papers with high Altmetrics scores also have high citation counts.

2.5 Summary Analysis of Altmetrics-related research reveals scholars continuously exploring its connotation and application. However, further in-depth research is needed regarding Altmetrics validity. Many factors such as publication time, discipline, journal, citation motivation, region, and data richness may cause experimental bias, yet few grouping studies exist. Therefore, this study selects a specific discipline and journal to explore disciplinary paper evaluation models. To examine whether combining citation, usage, and Altmetrics indicators outperforms single-type indicators and whether evaluation is credible, this paper proposes applying Altmetrics in specific disciplinary paper evaluation. Since researchers in different disciplines have vastly different tool preferences, current Altmetrics measurement methods are more applicable to STEM fields, while other disciplines (humanities, arts, and social sciences) lack available tools and indicators. Thus, this study selects civil engineering, a specific discipline within engineering, to explore Altmetrics applicability. Additionally, the "fairness test" method is employed to eliminate publication time effects on journal paper impact evaluation, with simplified procedures to maximize analytical reliability.

3. Research Design

To explore consistency among Altmetrics indicator measurement results, this study employs correlation analysis and principal component analysis for dimensionality reduction, constructing a single-discipline comprehensive literature evaluation model based on Altmetrics through existing measurement data to create new metrics for evaluation data use and reuse.

3.1 Sample Selection for Civil Engineering Papers Considering that papers require time from publication to viewing, saving, discussion, and citation, this study retrieves all articles with the theme of Civil Engineering published from 2009 to 2016 on PLoS ALMs (https://www.plos.org/article-level-metrics) on November 3, 2017, selecting Research Article type, yielding 410 articles. Metrics Data for these articles is downloaded. Initial data review reveals all articles have 0 Web of Science citation counts, which the authors speculate relates to PLoS database coverage [38]. Therefore, all articles are matched by DOI identifier in the Web of Science Core Collection to obtain Web of Science 180-day usage counts, Web of Science usage counts from 2013 to present, and Web of Science citation counts to date. Additionally, over 90% of data in the Metrics Data table is 0, lacking analytical value, so these fields are deleted. The final data included for analysis is shown in Table 1 . The code column is used for indicator system construction, where WOSShortUse refers to article usage counts on Web of Science in the most recent 180 days, and WOSLongUse refers to usage counts on Web of Science from 2013 to present. Statistical analysis software uses Stata SE 12.

3.2 Comprehensive Score Calculation Method According to Altmetrics scoring standards, the system emphasizes public and online impact but does not consider download and view counts. Combining calculation methods from altmetric.com and PLoS itself, this study designs an Altmetrics formula applicable to specific disciplinary academic environments—a bibliometric algorithm integrating Altmetrics for civil engineering.

3.2.1 Principal Component Analysis The feasibility of principal component analysis is further tested using KMO (Kaiser-Meyer-Olkin) and SMC (squared multiple correlations) measures (see Table 3). KMO sampling adequacy measures the strength of relationships between variables by comparing correlation coefficients and partial correlation coefficients. The KMO test result is 0.8025, indicating principal component analysis can effectively reduce data. SMC, the squared multiple correlation coefficient of a variable with all other variables (the coefficient of determination in multiple regression), shows high values, indicating strong linear relationships suitable for principal component analysis.

The first four components account for 67.743% of total variance, with subsequent components contributing progressively less. Following principal component retention criteria (eigenvalues > 1), the first four principal components are retained, with cumulative variance contribution approaching 80%. Therefore, 15 original indicators are divided into four principal components (see Table 4). Rotating the loading matrix yields corresponding eigenvectors (see Table 5).

From the component structure after principal component analysis, PLOSvi, PLOSPDFd, PLOSXMLd, PMCPDFD, Mendeley, Scopus, PMCEuCt, and WOSCit have high loadings on the first component, indicating it captures these metrics' information. Facebook and Twitter have high loadings on the second component. WOSSU and WOSlU load highly on the third component, reflecting Web of Science indicators. MCviews, Figshare, and Wikipe have high loadings on the fourth component. In summary, four principal components can adequately reflect all indicator information, replacing the original 15 indicators.

Denoting the four principal component factors as F1, F2, F3, and F4, their expressions are:

F1 = 0.779 *V1 + 0.925*V2 + 0.296 *V3 + 0.522*V4 + 0.629 *V5 + 0.892*S1 + 0.167*S2 + 0.165*S3 + 0.528*S4 + 0.083*D1 + 0.249*D2 + 0.272*D3 + 0.861*C1 + 0.777*C2 + 0.875*C3

F2 = 0.427 *V1 + 0.049*V2 + 0.035 *V3 + 0.009*V4 - 0.128 *V5 - 0.056*S1 + 0.059*S2 + 0.163*S3 + 0.093*S4 + 0.788*D1 + 0.841*D2 + 0.167*D3 - 0.265*C1 - 0.288*C2 - 0.248*C3

F3 = -0.055 *V1 + 0.043*V2 - 0.293 *V3 - 0.379*V4 - 0.424 *V5 + 0.242*S1 - 0.12*S2 + 0.769*S3 + 0.588*S4 - 0.131*D1 - 0.109*D2 - 0.028*D3 + 0.052*C1 - 0.175*C2 + 0.065*C3

F4 = -0.01*V1 - 0.121*V2 - 0.176*V3 + 0.559*V4 + 0.358*V5 + 0.024*S1 + 0.421*S2 + 0.175*S3 + 0.309*S4 - 0.216*D1 - 0.164*D2 + 0.41*D3 - 0.296*C1 - 0.246*C2 - 0.27*C3

3.2.2 Determining the Indicator System In principal component analysis, each component's variance contribution rate represents its information content. Using these contribution rates as weights, the four principal components' weights are calculated as 0.5512, 0.1788, 0.1467, and 0.1233, respectively (from Table 4). The weighted average of the four principal components yields the comprehensive academic impact:

F = 0.5512F1 + 0.1788F2 + 0.1467*F3 + 0.1233*F4

This model allows calculation of comprehensive scores for each paper.

4. Analysis of Paper Academic Impact Based on the "Fairness Test" Method

4.1 Correlation Analysis Between Indicators and Altmetrics Comprehensive Score Using the comprehensive impact calculation formula F, the comprehensive impact of all literature is computed. Correlations between each indicator and the Altmetrics comprehensive score are calculated and tested at the 1% significance level, with results shown in Table 6 .

The results show Altmetrics indicators have weak correlations (0.5-0.6) with Scopus citation indicators, PMC Europe Citations indicators, and Web of Science citation indicators. Single Altmetrics indicators have only weak relationships with traditional citation-based metrics. Building an Altmetrics evaluation system based on multi-source data indicators (such as open review data, attention and readership counts in reference management tools, and recommendation and click counts on social networks) addresses this issue.

Although citation counts remain an important indicator for judging academic impact, correlation analysis reveals that the PLOSviews indicator among view metrics correlates with the comprehensive score at 99.26%. Therefore, if other metrics correlate with PLOSviews, they may be associated with academic impact. However, simply correlating a paper's Altmetrics indicators with PLOSviews is unscientific due to social networks' short-term explosive propagation characteristics. Information on Twitter often declines rapidly over time, causing Altmetrics indicators based primarily on social network information to decrease over time. Meanwhile, papers introducing emerging frontiers often receive higher Altmetrics scores. Citation indicators show the opposite pattern, increasing over time without significant developmental trends. These differences require adjustment for more accurate evaluation using diverse data.

4.2 Eliminating Time Effects Using the "Fairness Test" Method For PLOS civil engineering literature, correlation analysis between Altmetrics scores and original indicators shows low correlation, necessitating experimental design to verify relationships between Altmetrics indicators and view counts. Since all papers come from PLOS within the same discipline with dense distribution across issues, this study selects three papers as a comparison group, using the two papers before and after each as references. The selection rule uses the two nearest published papers in the entire sample, significantly reducing workload. This new method eliminates time window effects, ensures consistent data sampling intervals for each group, guarantees data calculation independence, and reduces computational load.

Applying the "fairness test" method to 410 civil engineering papers from PLOS, each paper is compared with its immediately preceding and following papers in the sample. Since the earliest and latest published papers lack complete comparison pairs, 408 comparison groups are obtained for calculation and comparison.

4.3 Results Analysis Let the Altmetrics comprehensive score of a single paper be A, and a specific indicator be X. Each paper in the sample is denoted as p, with reference papers m and n (where m is published before n). Comparison results are divided into three groups as shown in Table 7 :

- **Success**: Ap > 1/2(Am + An) AND Xp > 1/2(Xm + Xn), OR Ap < 1/2(Am + An) AND Xp < 1/2(Xm + Xn)
- **Failure**: Ap > 1/2(Am + An) AND Xp < 1/2(Xm + Xn), OR Ap < 1/2(Am + An) AND Xp > 1/2(Xm + Xn)
- **Invalid**: All other cases

Since Altmetrics comprehensive scores retain three decimal places, invalid comparison results are reduced. The final comparison shows no invalid results across groups. Table 8 presents Altmetrics scores and comparison results, showing success counts far exceeding failure counts. PLOSviews has the highest success rate at 96.81%, validating previous correlation analysis results. Other test items show stable results, with most success rates between 56%-82%.

Comparing Tables 6 and 8 reveals differences in correlations with the comprehensive score after "fairness test" correction. Before correction, indicators with correlations above 0.5 (from high to low) are: PLOSsvi - PLOSPDFd - Mendeley - WOSCit - Scopus - PMCEuct. After correction, the ranking becomes: PLOSviews - PLOSPDFdownloads - PMCviews - Twitter - Mendeley - Scopus - PMCPDFDownloads. After eliminating time window interference, indicators with low correlation with the comprehensive score (such as Facebook, PMCEuropeCitations, Wikipedia) decline most, while those with high correlation (such as PMCviews, Twitter) increase significantly, demonstrating that eliminating time effects amplifies differences between indicators and the F-value.

Among indicators with increased correlation after testing, PMCviews among view metrics increases most, as papers can only gain high attention through being viewed. Citation indicators in Web of Science show increased correlation, suggesting traditional citation-based impact evaluation remains meaningful. The increased WOSShortUse among save metrics indicates short-term usage is also an important impact indicator, with its importance becoming prominent after eliminating time factors.

In contrast, social media communication-related indicators (such as Facebook, Wikipedia) show decreased correlation after testing, consistent with their explosive propagation characteristics. These indicators generally have low correlation with the F-value, though notably, the Twitter indicator increases instead, maintaining high correlation with the F-value. This suggests users have lower stickiness to Twitter than Facebook, with Twitter having broader and freer propagation that still shows short-term explosive patterns, while Facebook's friend-based communication has higher stability and less temporal volatility.

5. Conclusion and Discussion

This study's innovations include: 1) constructing a comprehensive evaluation indicator model through principal component analysis of ALMs indicators for civil engineering papers in PLOS; 2) applying the "fairness test" to examine correlations between single Altmetrics indicators and comprehensive Altmetrics scores, finding that after eliminating time window interference, PMCviews among view metrics increases most, as papers require viewing to gain attention.

The study has limitations: 1) PLoS ALMs only processes PLOS publications and articles stored in PubMed Central, limiting its scope; 2) current natural science fields emphasize journal papers, while many humanities and social sciences outputs (such as poetry, records, artworks) are difficult to evaluate quantitatively, requiring qualitative indicators like peer comments. Future research could explore evaluation systems for humanities and social sciences using Book Citation Index (BCI) and other indicators.

Altmetrics serves as an evaluation filtering tool but not the only one. Although Altmetrics research is flourishing, it remains immature with recognition to be improved, sometimes requiring correction through methods like "fairness test." Analysis of PLoS ALMs data reveals imperfect integration of Altmetrics indicators in existing tools, with substantial missing data constraining Altmetrics development. The analysis of paper impact should consider not only correlations between Altmetrics indicators and impact but also the importance and quantification degree of different indicators.

This empirical study finds weak correlation between Altmetrics measures and traditional citation measures, making it impossible to determine whether Altmetrics can replace traditional evaluation methods for paper assessment in this specific case. Future research will focus on improving Altmetrics credibility through multiple method testing and validation.

References

[1] Priem J, Taraborelli D, Groth P, et al. Altmetrics: a manifesto [EB/OL]. [2018-05-21]. http://altmetrics.org/manifesto.

[2] Liu J. Metrics and beyond @ SpotOnLondon2012 [EB/OL]. [2017-11-02]. https://www.altmetric.com/blog/metrics-and-beyond-spoton-london-2012/.

[3] NISO. NISO to develop standards and recommended practices for Altmetrics [EB/OL]. [2017-11-02]. http://www.niso.org/news/pr/view?item_{key}=72efc1097d4caf7b7b5bdf9c54a165818

[4] Proquest. Altmetric data now available in the Summon Discovery Service [EB/OL]. [2017-11-02]. http://www.proquest.com/about/news/2016/Altmetric-data-now-available-in-the-Summon-Discovery-Service.html.

[5] Yu Houqiang, Qiu Junping. Application of altmetrics in library literature services [J]. Journal of Intelligence, 2014, 33(9): 163-166.

[6] Brody T, Harnad S, Carr L. Earlier Web usage statistics as predictors of later citation impact [J]. Journal of the Association for Information Science & Technology, 2006, 57(8): 1060-1072.

[7] Wan JK, Hua PH, Rousseau R, et al. The journal download immediacy index (DII): experience using a Chinese full-text database [J]. Scientometrics, 2010, 82(3): 555-566.

[8] Yan KK, Gerstein M. The spread of scientific information: insights from the web usage statistics in PLoS article-level metrics [J]. PLoS ONE, 2011, 6(5): e19917.

[9] Wang X, Liu C, Fang Z, et al. From attention to citation, what and how does altmetrics work? [EB/OL]. [2017-11-02]. https://arxiv.org/ftp/arxiv/papers/1409/1409.4269.pdf.

[10] Alhoori H, Furuta R. Do altmetrics follow the crowd or does the crowd follow altmetrics? [C]//Proceedings of the 14th ACM/IEEE-CS joint conference on digital libraries. Piscataway: IEEE Press, 2014: 375-378.

[11] Bornmann L. Alternative metrics in scientometrics: a meta-analysis of research into three altmetrics [J]. Scientometrics, 2015, 103(3): 1123-1144.

[12] Costas R, Zahedi Z, Wouters P. Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective [J]. Journal of the Association for Information Science and Technology, 2015, 66(10): 2003-2019.

[13] Zuccala AA, Verleysen FT, Cornacchia R, et al. Altmetrics for the humanities: comparing goodreads reader ratings with citations to history books [J]. Aslib Journal of Information Management, 2015, 67(3): 320-336.

[14] Chen K, Tang M, Wang C, et al. Exploring alternative metrics of scholarly performance in the social sciences and humanities in Taiwan [J]. Scientometrics, 2015, 102(1): 97-112.

[15] Thelwall M, Haustein S, Larivière V, et al. Do altmetrics work? Twitter and ten other social web services [J]. PLoS ONE, 2013, 8(5): e64841.

[16] Priem J, Piwowar HA, Hemminger BM. Altmetrics in the wild: using social media to explore scholarly impact [EB/OL]. [2017-11-02].

https://arxiv.org/abs/1203.4745.

[17] Lin J. A case study in anti-gaming mechanisms for altmetrics: PLoS ALMs and DataTrust [EB/OL]. [2017-11-02]. http://altmetrics.org/altmetrics12/lin/.

[18] Yu Houqiang. Conversion of altmetrics indicators [EB/OL]. [2017-11-02]. http://blog.sciencenet.cn/blog-441629-893201.html.

[19] Liu Chunli. Research on selective article impact metrics based on PLoS API [J]. Library and Information Service, 2013, 57(7): 89-95.

[20] Gu Liping. Article-level metrics research: application case analysis [J]. New Technology of Library and Information Service, 2013(11): 1-7.

[21] Qiu Junping, Yu Houqiang. The proposal process and research progress of altmetrics [J]. Library and Information Service, 2013, 57(19): 5-12.

[22] You Qingbin, Tang Shanhong. Supplementary metrology and its application prospects [J]. Information Theory and Practice, 2013, 36(12): 6-10.

[23] Liu Chunli. Article-Level Metrics (ALM): development process, characteristics, indicators, and applications [J]. Library Journal, 2016, 35(2): 63-69.

[24] Neylon C, Wu S. Article-level metrics and the evolution of scientific impact [J]. PLoS Biology, 2009, 7(11): e1000242.

[25] PLOS. Article-Level Metrics [EB/OL]. [2018-05-21]. https://www.plos.org/article-level-metrics.

[26] Song Lingling. Research on paper impact evaluation indicators based on PLoS ALMs: taking aerospace medicine as an example [J]. Information Studies: Theory & Application, 2016, 39(9): 30-35.

[27] Wang Rui, Hu Wenjing, Guo Wei. Comparison of common Altmetrics tools [J]. New Technology of Library and Information Service, 2015, 30(12): 18-26.

[28] Liu Xiaojuan, Song Wanzi. Usability analysis of altmetrics indicators based on PLOS ALM [J]. Library and Information Service, 2016, 60(4): 93-101.

[29] Lapinski S, Piwowar H, Priem J. Riding the crest of the altmetrics wave: how librarians can help prepare faculty for the next generation of research impact metrics [J]. College & Research Libraries News, 2013, 119(18): 3359-3366.

[30] Wang Wenxia, Liu Chunli. Differences in evaluation indicator models for paper impact across disciplines [J]. Library and Information Service, 2017, 61(13): 108-116.

[31] Piwowar H, Priem J. The power of altmetrics on a CV [J]. Bulletin of the American Society for Information Science & Technology, 2013, 39(4): 10-13.

[32] Rodgers E, Barbro S. A look at altmetrics and its growing significance to research libraries [EB/OL]. [2017-11-02]. http://eprints.rclis.org/25481/1/Altmetrics.pdf.

[33] Nieves G, Maria-Isabel D, Antonio C. What role do librarians play in altmetrics? [EB/OL]. [2017-11-02]. https://deepblue.lib.umich.edu/handle/2027.42/99709.

[34] Radicchi F, Castellano C. Testing the fairness of citation indicators for comparison across scientific domains: the case of fractional citation counts [J]. Journal of Informetrics, 2012, 6(1): 121-130.

[35] Leydesdorff L, Radicchi F, Bornmann L, et al. Field-normalized impact factors (IFs): a comparison of rescaling and fractionally counted IFs [J]. Journal of the Association for Information Science and Technology, 2013, 64(11): 2299-2309.

[36] Chen Fuyou, Yang Liying, Ding Jielan. Methodology and empirical research on comparing academic influence of journals across disciplines [J]. Library and Information Service, 2013, 57(23): 85-89.

[37] Wang Rui, Hu Wenjing, Guo Wei. Academic impact of scientific papers with high Altmetrics indicators [J]. Library and Information Service, 2014, 58(21): 92-98.

[38] PLOS. Public Library of Science (PLOS) citation Info [EB/OL]. [2018-05-22]. http://www.lagotto.io/plos/#citationInfo.

[39] Altmetrics support. How is the altmetric attention score calculated? [EB/OL]. [2017-11-02]. https://help.altmetric.com/support/solutions/articles/6000060969-how-is-the-altmetric-score-calculated-.

Author Contributions:

Liu Hongxu: Proposed research topic, wrote and revised paper; Wang Zheng: Provided revision suggestions.

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