Analysis of Zero-Cited and Highly-Cited Papers from the Perspective of Research Themes: A Case Study in Environmental Science (Post-print)

Authors: Pan Fei, Wang Xiaoyue, Bai Rujiang, Zhou Yanting

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

Abstract

[Purpose/Significance] From a thematic perspective, this study analyzes uncited papers in the field of environmental science, compares differences between uncited and highly cited papers in terms of content and external indicators, and reveals the underlying causes of uncited papers. [Method/Process] First, PLDA topic identification is applied to abstracts of 260 highly cited papers and 907 uncited papers in domestic environmental science from the Web of Science database. Next, topic similarity calculation is employed to identify inter-topic relationships, with topic popularity as an internal indicator and publication date and journal as external evaluation metrics. Finally, a comparative analysis between uncited and highly cited papers is conducted for both identical and distinct topics by integrating thematic content with external indicators. [Results/Conclusions] When research topics are identical, journal impact factor is the primary factor influencing uncited papers; when topics differ, the thematic content of the research is the main reason why papers remain uncited.

Full Text

Preamble

ChinaXiv Collaborative Journal
Volume 62, Issue 20, October 2018

Analysis of Zero-Cited and Highly-Cited Papers from the Perspective of Research Topics: A Case Study of Environmental Science

Pan Fei, Wang Xiaoyue, Bai Rujiang, Zhou Yanting
Institute of Scientific and Technical Information, Shandong University of Technology, Zibo 255049

Abstract

[**Purpose/Significance**] This study analyzes zero-cited papers in the field of environmental science from a topical perspective, comparing differences in article content and external indicators between zero-cited and highly-cited papers to reveal the reasons for the existence of zero-cited papers. [**Method/Process**] First, PLDA topic modeling was applied to identify topics from the abstracts of 260 highly-cited papers and 907 zero-cited papers in domestic environmental science from the Web of Science database. Topic associations were then discovered through topic similarity calculation. Topic popularity served as an internal indicator, while publication time and journal served as external evaluation indicators. Finally, a comparative analysis between zero-cited and highly-cited papers was conducted by combining paper topics with external indicators for both similar and different topics. [**Result/Conclusion**] The results show that under the same research topic, journal impact factor is the main factor affecting zero-cited papers; under different topics, the research topic content itself is the primary reason leading to zero-cited papers.

Classification Number: G250
Keywords: zero-cited, highly-cited, topic recognition, environmental science, comparative analysis, evaluation indicators

1. Introduction

With the development of science, technology, and information technology, the volume of scientific literature continues to increase, and research on citation frequency has attracted significant attention from scholars. In bibliometrics, the basic indicator for measuring paper impact or quality is citation count [1-2]. Compared with highly-cited papers, zero-cited papers receive less attention, and research on them lags far behind that on highly-cited papers. However, the potential value of zero-cited or low-cited literature, if excavated, may far exceed our current imagination [3]. Zero-cited papers refer to a collection of papers published by a country, institution, discipline, journal, or individual in a given year or time period that receive no citations within one or several citation time windows of varying lengths after publication [4]. Scholars are concerned about why zero-cited papers exist and how to better discover their value. This paper addresses the following questions regarding highly-cited versus zero-cited papers: Are highly-cited papers frequently cited because they address hot topics, or because they are published in high-impact journals? Are zero-cited papers uncited due to low topic popularity, or because they are published in low-impact journals? To answer these questions, this study selects data on highly-cited and zero-cited papers in environmental science and employs the PLDA (Parallel Latent Dirichlet Allocation) model for topic identification, combined

with analysis of publication time and journals to evaluate the impact of various indicators on zero-cited and highly-cited papers and identify factors influencing zero-cited papers.

In the 1950s, the renowned American information scientist E. Garfield proposed citation analysis methods, which are commonly used for information analysis and scientific evaluation [1]. However, previous indicators have primarily focused on the "head" of the citation distribution curve representing "highly-cited and attention-grabbing papers," while neglecting "low-cited and temporarily unnoticed papers." According to the long-tail theory, zero-cited papers' contribution to science matches that of highly-cited papers. In 2004, British scholar A. Weale et al. [5] proposed that the zero-citation rate could serve as a reverse evaluation indicator for journal quality. Subsequently, T.N. Van Leeuwen and H.F. Moed found a decreasing functional relationship between journal impact factor and zero-citation rate, with a Pearson correlation coefficient of -0.63 [6]. Domestic scholar Tang Xiaoli validated the reasonableness of using zero-citation rate for reverse journal evaluation using economics as an example [7]. Li Meiyu et al. confirmed that zero-citation rate could serve as a reverse indicator of key journal evaluation metrics in library and information science, though disciplinary differences should be considered [8]. By studying the relationship between zero-citation rates and journal evaluation metrics and constructing new research evaluation indicators that integrate zero-citation rates, scholars have expanded journal evaluation standards and affirmed the value of zero-cited papers.

Overall, researchers have studied the causes of zero-cited papers from the data itself, including paper type, national and institutional research strength, funding distribution, disciplinary differences, language, degree of research collaboration, and topic selection. While citation counts are important metrics for research papers, they represent simple statistical analysis without delving into the topic information behind the data. Even when studying topic factors, researchers have only examined textual information such as keywords and high-frequency terms without exploring semantic relationships. This paper selects data on highly-cited and zero-cited papers in environmental science, uses the PLDA model for topic identification, employs topic popularity as an internal evaluation metric, and combines it with external indicators such as publication time and journals to analyze the reasons and patterns behind zero-cited papers.

2. Related Research

Regarding why academic papers receive no citations, many factors beyond the academic quality of the paper itself are involved. Fang Hongling analyzed zero-cited papers published in 2003 in five Chinese core ophthalmology journals and found that some low-cited or even zero-cited papers had high download volumes, and that zero-cited papers covered a wide range of topics [9]. Wei Ruibin et al. argued that the main cause of zero citations was that paper topics were too

narrow and not part of mainstream research fields [10]. Gao Jiping et al., using JCR spectroscopy journals as an example, identified data sources, publication time, and research topics as important factors influencing zero-cited papers [4]. Yang Siluo, using 15 core journals in library, information, and archival science as an example, found that the average zero-citation rate was negatively correlated with average citations per paper and H-index, and positively correlated with comprehensive ranking; zero-citation rates varied significantly across time, journals, and disciplines [11]. Hu Zewen used questionnaire surveys to identify reasons for zero citations, finding that short publication time, low paper quality, unpopular or unoriginal topics, and low journal impact (or quality) were the main causes [12]. Wen Fangfang, comparing zero-cited and highly-cited papers in information science journals, concluded that paper visibility and accessibility, author influence, number of co-authors, and whether topics were novel and popular all affected citation frequency to varying degrees [13]. Du Xinzheng et al. [14] analyzed characteristics of zero-cited papers in *Acta Hydrobiologica Sinica* from six aspects: paper type, content structure, funding distribution, author institution distribution, page count, and language, finding no correlation between page count/language and zero citations, while other indicators could reflect zero-cited paper characteristics. Zhao Yue analyzed topic factors and found that zero-cited papers had dispersed and outdated research topics, though these topics did not deviate from disciplinary research areas or show highly forward-looking characteristics [15]. Kuang Shumei et al. analyzed keywords of papers in library and information science and found that the research similarity between cited and uncited papers gradually decreased, and the relevance of research content to disciplinary hotspots greatly affected the probability of zero citations [16]. Li Heqiong et al. analyzed zero-cited papers in 10 comprehensive surgery journals in 2011, finding little relationship between zero-citation rate and impact factor, that most had 2-5 authors, that first authors were mainly from provincial/municipal hospitals, that most had no funding, and that paper types were primarily clinical studies and case reports [17].

---

3. Research Approach

To further analyze differences and connections between zero-cited and highly-cited papers, this study uses research topics as internal evaluation indicators and publication time and journals as external evaluation indicators. Only through internal-external analysis can we effectively identify characteristics of zero-cited papers, and topic modeling is an effective means to analyze paper internal information. The research approach consists of three steps, as shown in Figure 1 [Figure 1: see original paper].

Figure 1. Research Framework for Comparative Analysis of Highly-Cited and Zero-Cited Papers

1. **Data Collection and Preprocessing.** First, data sources are identi-

fied. The Web of Science database is selected for data collection, and retrieval formulas are constructed to obtain literature in the discipline. The collected literature is then preprocessed and cleaned, including removing stop words, punctuation, and numbers; extracting word stems; and processing high-frequency terms and meaningless words in the field to retain meaningful terms for subsequent text topic identification.

2. **Topic Identification and Similarity Calculation.** The PLDA model is used to identify topics embedded in the abstracts of highly-cited and zero-cited papers, constructing topic-document and topic-term matrices. Based on the topic-term matrices, topic similarity calculation methods are applied to compute similarity between zero-cited paper topics and highly-cited paper topics, revealing differences and connections.

3. **Feature Extraction and Analysis.** The PLDA topic modeling results provide data on publication time and journals under each topic. Using indicators such as topic popularity, publication time, and journals, the main factors influencing zero-cited papers are identified, providing methods and ideas for further analysis of zero-cited paper causes.

3.1 PLDA Topic Model

The PLDA model is an LDA model based on a Gibbs sampling approximate distribution parallel framework. To ensure accurate topic numbers, perplexity—a commonly used evaluation metric in statistical language models—is employed to determine the optimal number of topics. Lower perplexity indicates better topic identification. D.M. Blei et al. defined perplexity for a topic model with M documents as [18]:

$$\text{perplexity}(D_{\text{test}}) = \exp\left\{-\frac{\sum_{d=1}^{M}\log p(w_d)}{\sum_{d=1}^{M} N_d}\right\}$$

where M is the number of documents in the collection, $p(w_d)$ is the probability of the LDA model generating the d-th document, and N_d is the number of words in the d-th document. When perplexity(D_{test}) is minimized, K topics best express the semantic relationships of the model, thus determining the optimal number of topics.

3.2 Topic Similarity

The relationships among topics identified by the PLDA model can be analyzed through topic similarity. This paper uses cosine similarity to calculate topic similarity and sets a threshold Y. Similarity values range between [0,1], with higher values indicating greater similarity.

Step 1: Build Vector Space Model (VSM). Highly-cited topics and zero-cited topics are described using vectors. In VSM, T(Topic) represents topics,

T(Term) represents topic terms, and W(weight) represents term weights. Topic vectors can be represented by topic terms as Topic $= \{t_1, t_2, t_3, \ldots, t_n\}$, and term weight vectors as TopicVector $= \{w_1, w_2, w_3, \ldots, w_n\}$, where each topic term has a weight.

Step 2: Calculate pairwise topic similarity. The calculation result falls between [0,1]. The topic similarity formula is:

$$\mathrm{Sim}(\mathrm{Topic}_i, \mathrm{Topic}_j) = \cos\theta = \frac{\sum w_k(\mathrm{Topic}_i) \times w_k(\mathrm{Topic}_j)}{\sqrt{\sum w_k(\mathrm{Topic}_i)^2} \times \sqrt{\sum w_k(\mathrm{Topic}_j)^2}}$$

If the similarity exceeds Y, the two topics are considered the same; otherwise, they are considered different. The numerator represents the dot product of two topic vectors, and the denominator represents the product of the magnitudes of the two topic vectors.

3.3 Comparative Analysis Indicators for Zero-Cited and Highly-Cited Papers

Drawing on hybrid judgment indicators proposed in current research [19-20], this paper proposes an indicator system for paper analysis through examination and summarization of textual content and external attributes, including topic popularity, publication time, and publication journals.

3.3.1 Topic Popularity Indicator Paper publication volume and citation volume can serve as evaluation indicators for research popularity [19-20]. Based on the number of papers identified under each topic, topic popularity is defined as the number of papers within each topic—that is, the weight of papers in different topics relative to the total number of papers. Topic popularity can intuitively analyze the attention and influence trends of research topics [21].

$$TH = \frac{X_i}{\sum X_j}$$

where TH represents topic popularity, X_i represents the number of papers under each topic, and $\sum X_j$ represents the total number of papers across all topics.

3.3.2 Publication Time Indicator The publication time indicator is a fundamental factor in analysis, primarily examining changes in the number of papers under topics across different years. It reflects topic development trends—whether emerging, growing, or declining [22]—and shows researcher attention to papers.

3.3.3 Publication Journal Indicator Scientific journals are the primary carriers of papers, and journal quality also affects citation counts. This paper compares journals of zero-cited and highly-cited papers to determine journal impact. SPSS exponential regression is used for analysis. The exponential model formula is:

$$Y = \beta_0 e^{\beta_1 x}$$

where Y represents the number of papers published in each journal under each topic, x represents the chronological order of papers, and β_0 and β_1 are constants.

---

4. Data Analysis

4.1 Data Source and Preprocessing

4.1.1 Data Source Web of Science offers authoritative and comprehensive advantages; therefore, this study collected data from the Web of Science database, selecting environmental science as the research discipline. **Database:** SCI-EXPANDED; **Search query:** TI = "environ*"; **Time span:** 2006-2015; **Document types:** article and review; **Subject categories:** environmental science & environmental science ecology; **Search country:** Peoples R China; **Search language:** English; **Search date:** June 20, 2017; **Search results:** 15,002 papers.

Based on the definition of zero-cited papers, this study treated papers with zero citations between 2006-2015 as zero-cited papers. Highly-cited papers were defined using Thomson Reuters' ESI tool [23] as papers ranked in the top 1% of citations in each discipline over the past 10 years. The final dataset included 907 zero-cited papers and 260 highly-cited papers.

As shown in Figure 2 [Figure 2: see original paper], both highly-cited and zero-cited papers in environmental science show growth trends. Highly-cited papers numbered 0 in 2006, reaching 58 papers 10 years later, indicating good development in environmental science during this period. Zero-cited papers numbered 19 in 2006, increasing to 464 in 2015—24 times the 2006 figure, with growth exceeding that of highly-cited papers. Reasons include: (1) papers typically reach citation peaks one to two years after publication, resulting in more zero-cited papers; (2) as paper volume increases, irrelevant scholars produce unrelated papers that are difficult to locate, leading to non-citation.

4.1.2 Data Preprocessing The SATI 3.2 tool was used to extract abstracts from highly-cited and zero-cited papers. Python was then used for text data preprocessing and cleaning, including removing punctuation and numbers, converting case, stop word processing, and stemming. Further cleaning removed

topic-unrelated words and high-frequency domain terms such as “environment,” “environmental,” “china,” “Elsevier,” “right,” and “paper” to improve topic identification accuracy. Results are shown in Figure 3 [Figure 3: see original paper].

4.2 Comparative Topic Analysis of Zero-Cited and Highly-Cited Papers

4.2.1 Experimental Parameter Settings and Results Analysis After text processing, topic identification requires careful parameter setting, particularly the number of topics (No. of topic) and terms per topic (No. of words per topic), which greatly affect accuracy. This study experimented with the relationship between topic number and perplexity for highly-cited papers. Given the small abstract text volume, the estimated topic number was under 15; No. of topic was set from 2 to 12 in increments of 2. The relationship is shown in Figure 4 [Figure 4: see original paper]. When No. of topic is 10, the trend stabilizes. While fewer topics yield lower perplexity, this causes overfitting. Therefore, the final topic number was set to 10, with 15 terms per topic. Other parameters: Alpha 0.5; Beta 0.1; Iterations 2000.

Similarly, experiments on zero-cited paper abstracts showed perplexity stabilizing at 12 topics.

Topic identification results for zero-cited and highly-cited papers are shown in Tables 1 and 2 . Different topics represent different research content.

Zero-cited paper topics fall into three categories:
1. **Environmental pollutant research:** Topic 0 studies pollutant toxicity to aquatic organisms and plants through exposure concentrations in water.
2. **Natural ecosystem evaluation research:** Using mathematical, physical, spatial, landscape ecological modeling, and monitoring mechanisms to assess ecosystem status and changes holistically, analyzing causes of ecosystem changes and environmental impacts on both environment and humans to inform regional ecological management policies. Topics 1, 2, 3, 4, 6, and 11 address ecosystem evaluation involving climate, geology, hydrology, etc., with Topic 11 offering a new landscape-based evaluation perspective.
3. **Pollution treatment methods:** Topics 5, 7, 8, 9, and 10 analyze pollution causes and apply microbial-biological technologies to treat pollutants in air, water, and soil.

Highly-cited paper topics fall into two categories:
1. **Environmental pollutant research:** Topics 1 (antibiotic pollution), 3 (potential dangers of nanomaterials), 4 (effective utilization of polluted materials), and 6 (e-waste and heavy metal pollution treatment) examine environmental pollutants from different angles.
2. **Environmental pollutant treatment research:** Topics 0 and 2 study pollutant treatment using new materials like graphene and carbon nanotubes; Topic 7 approaches pollution reduction from ecological/green energy perspectives; Topics 5, 8, and 9 treat pollutants through biological methods, degrada-

tion, and photocatalytic technologies.

Zero-cited paper topics are more extensive, including ecosystem evaluation, while highly-cited topics focus on specific technical measures for toxic pollutant management.

4.2.2 Topic Popularity Topic popularity reflects research directions in the field. Different topics have different popularity levels, which represent attention levels. Higher popularity indicates higher attention. The popularity threshold was set at 0.1; values above 0.1 indicate high-popularity topics. Among topics below 0.1, the bottom three were classified as low-popularity, with others as medium-popularity.

Table 3 shows topic popularity for zero-cited and highly-cited papers. For zero-cited papers, Topics 4, 9, and 10 are high-popularity; Topics 0, 1, 3, 6, 8, and 11 are medium-popularity; Topics 2, 5, and 7 are low-popularity. For highly-cited papers, Topics 0, 1, 7, and 9 are high-popularity; Topics 3, 6, and 8 are medium-popularity; Topics 2, 4, and 5 are low-popularity.

4.2.3 Topic Similarity Python's Gensim toolkit was used to calculate similarity among 22 topics identified from zero-cited and highly-cited papers. A threshold was set to determine similarity degrees, with results shown in Table 4 .

Similarity calculations revealed that among highly-cited topics, Topic 9 had the lowest similarity with zero-cited topics. HT9 focuses on photocatalyst materials, particularly nanomaterials like graphitic carbon nitride (g-C3N4), while zero-cited topics concentrate on microbial treatment and ecosystem evaluation. Among zero-cited topics, Topic 11 had the lowest similarity with highly-cited topics. ZT11 covers ecosystem evaluation (urban, agricultural, forest, marine, landscape ecology) not addressed in highly-cited topics.

To enable meaningful comparison, a similarity threshold of 0.2 was set; values above 0.2 indicate related topics, below 0.2 indicate no relationship. Table 4 shows four similar topic pairs: ZT2/HT4 (effective utilization of polluted materials), ZT4/HT5 (water pollution control and environmental microbial technology), ZT7/HT8 (pollutant treatment technology), and ZT10/HT0 (environmental adsorption, focusing on adsorption techniques and materials). Remaining topics are dissimilar.

4.3 Same-Topic Feature Comparison

4.3.1 Topic and Publication Time Comparison Four topic pairs showed similarity. Analysis of their research topics and publication times is shown in Table 5 .

(1) HT0 and ZT10 Feature Analysis
Conclusion 1: High topic popularity. Both address pollutant adsorption re-

search with high topic popularity.
Conclusion 2: Sustainable publication timeline. Papers on this topic were cited across all 10 years, with the most publications in 2014. ZT10 had few zero-cited papers in 2006-2011, with growth beginning after 2011. As a high-popularity topic in highly-cited papers, environmental adsorption attracted researcher attention, generating many results and leading to papers without citations yet.

(2) HT5 and ZT4 Feature Analysis
Conclusion 1: Different topic popularity. Both address water pollution control and environmental microbial technology. HT5 is low-popularity in highly-cited papers, while ZT4 is a high-popularity topic.
Conclusion 2: Recent publication time. Publications concentrate in the last 5 years, indicating this short-term research frontier is novel and advanced but not yet systematic, producing many zero-cited papers.

(3) HT4 and ZT2 Feature Analysis
Conclusion 1: Low topic popularity. Both focus on effective utilization of polluted materials but with low popularity, as treatment technology research difficulty increases (e.g., anaerobic technology, carbon nanotechnology), requiring specialized equipment and conditions.
Conclusion 2: Sustainable publication timeline. Both cited and uncited papers appear across all 10 years, with high research difficulty and few breakthrough technologies leading to long publication cycles.

(4) HT8 and ZT7 Feature Analysis
Conclusion 1: Low topic popularity. Both address pollutant treatment technology but with low popularity.
Conclusion 2: Different publication timelines. ZT7 has papers across all 10 years, while HT8 citations only increased significantly in the last 5 years, indicating a potentially promising topic.

4.3.2 Journal Comparison Papers under the same topic were sorted by time, and counts of papers in the same journals were analyzed using IBM SPSS Statistics 24.0. Each paper served as the independent variable (X, horizontal axis), and the number of papers in the same journal as the dependent variable (Y, vertical axis). Scatter plots were used to preliminarily judge trends, with exponential functions for analysis. Results are shown in Figure 5 [Figure 5: see original paper].

Curve trends: Zero-cited paper curves change noticeably, concentrating in older journals, while highly-cited paper counts in the same journals remain stable over time. HT5 shows positive growth, indicating recent concentration.

Distribution patterns: Zero-cited papers are relatively dispersed across many journal types, while highly-cited papers are concentrated and regular, appearing in fewer journal types.

Journal productivity: Among zero-cited topics, the top 5 journals by paper count are *Fresenius Environmental Bulletin* (41), *Environmental Progress & Sustainable Energy* (17), *Sustainability* (14), *Journal of Environmental Sciences* (13), and *Frontiers of Environmental Science & Engineering* (13), with impact factors of 0.425, 1.672, 0, 2.937, and 1.716, respectively. These are Q3/Q4 ranked, lower-impact journals. Among highly-cited topics, the most productive journals are *Journal of Hazardous Materials* (33), *Environmental Science & Technology* (20), *Science of the Total Environment* (7), *Water Research* (6), and *Energy & Environmental Science* (4), with impact factors of 6.065, 3.751, 4.900, 6.942, and 5.715, respectively. These are Q1 ranked, high-quality, high-impact journals.

Conclusion: Under the same topic, journal impact factor is the main factor affecting zero citations. Highly-cited papers appear in high-impact, top-ranked journals representing research frontiers. Researchers following these topics often publish in lower-impact journals, creating "replication projects" of work already published in high-level journals, resulting in zero-cited papers.

4.4 Different-Topic Feature Comparison

Zero-cited and highly-cited papers also have their own distinct topics: zero-cited topics include ZT0, ZT1, ZT3, ZT5, ZT6, ZT8, ZT9, ZT11; highly-cited topics include HT1, HT2, HT3, HT6, HT7, HT9. These are compared by topic popularity and publication journals.

Topic popularity: Zero-cited papers show ZT9 as high-popularity, ZT0, ZT1, ZT3, ZT6, ZT8, ZT11 as medium-popularity, and ZT5 as low-popularity. Highly-cited papers show HT1, HT7, HT9 as high-popularity, HT3, HT6 as medium-popularity, and HT2 as low-popularity. Highly-cited papers concentrate on high-popularity topics representing research frontiers, while zero-cited papers focus on medium-popularity topics with lower attention and less grasp of current research directions.

Journal analysis: As shown in Figure 6 [Figure 6: see original paper], zero-cited paper curves change rapidly, concentrating in tail-end journals with few papers per journal. Highly-cited paper curves change slowly, with small citation count gaps across journals. Zero-cited papers are relatively concentrated in a few journal types, while highly-cited papers are evenly distributed across journals.

Journal productivity: Zero-cited papers appear most frequently in *Fresenius Environmental Bulletin* (66), *Environmental Science and Pollution Research* (27), *Environmental Earth Sciences* (27), *Polish Journal of Environmental Studies* (18), and *Journal of Coastal Research* (14), with impact factors of 0.425, 2.741, 1.569, 0.793, and 0.915. Highly-cited papers appear most in *Environmental Science & Technology* (23), *Journal of Hazardous Materials* (22), *Energy Policy* (14), *Energy & Environmental Science* (12), and *Science of the Total Environment* (11), with impact factors of 6.198, 6.065, 4.140, 29.518, and 4.9. Highly-cited journal impact factors exceed those of zero-cited papers.

Conclusion: Zero-cited papers are concentrated in low-impact, low-visibility journals, while highly-cited papers are evenly distributed in high-impact, high-visibility journals. Beyond journal impact, highly-cited papers have superior topic selection—HT1, HT7, HT9 are hot topics with novel content attracting high attention. Zero-cited topics like ZT0, ZT1, ZT3, ZT5, ZT6, ZT8 are medium-popularity with low attention. Therefore, under different topics, topic content is the main reason for zero-cited papers.

---

5. Conclusion

This study addresses the limitation of previous research that primarily used bibliometric indicators without deep textual analysis. Using the PLDA model to identify topics in highly-cited and zero-cited paper abstracts, topic similarity calculation reveals relationships between topics. Comparative analysis of same-topic and different-topic scenarios using topic popularity, publication time, and journal indicators further reveals reasons for zero-cited papers.

Under same topics, publication journals are the main factor affecting zero citations. Highly-cited papers appear in high-impact, top-ranked journals representing research frontiers. Researchers following these topics often publish in lower-impact journals, creating "replication projects" that result in zero-cited papers.

Under different topics, topic selection is the main cause of zero-cited papers. Highly-cited papers appear in high-impact journals with hot, novel topics of research value and guidance. Zero-cited papers typically select medium-popularity topics with mature research systems but lacking innovation, concentrating in low-impact journals with limited influence.

In summary, zero-cited papers are not worthless. We should explore and excavate their value rather than ignore them due to citation limitations. Due to Web of Science's limited coverage and scope, sample data acquisition and analysis may contain some bias. The results serve as a reference, with some conclusions requiring validation by domain experts.

References

[1] Garfield E. Citation indexes for science: a new dimension in documentation through association of ideas[J]. Science, 1955, 122(3159): 108-111.
[2] Garfield E. Citation analysis as a tool in journal evaluation journals can be ranked by frequency and impact of citations for science policy studies[J]. Science, 1972, 178(4060): 471-479.
[3] Zhu Mengjiao, Wu Yishan. Zero-citation phenomenon: a literature review[J]. Information Studies: Theory & Application, 2013, 36(8): 111-116.

[4] Gao Jiping, Pan Yuntao, Wu Yishan. Analysis of factors causing zero-cited papers: a case study of national, institutional, and topic distribution of zero-cited papers in spectroscopy[J]. Science & Technology Review, 2015, 33(8): 112-119.
[5] Weale AR, Mick B, Lear PA. The level of non-citation of articles within a journal as a measure of quality: a comparison to the impact factor[J]. BMC medical research methodology, 2004, 4(1): 14-14.
[6] Van Leeuwen TN, Moed HF. Characteristics of journal impact factors: the effects of uncitedness and citation distribution on the understanding of journal impact factors[J]. Scientometrics, 2005, 63(2): 357-371.
[7] Tang Xiaoli, Wu Qunfang, Wang Jimin. Research on the relationship between zero-citation rate and journal impact: a case study of economics[J]. Library and Information Service, 2014, 58(19): 100-104.
[8] Li Meiyu, Wang Shuo, Zheng Dejun. Correlation analysis between zero-citation rate and key journal evaluation indicators in Chinese journals: a case study of library and information science[J]. Chinese Journal of Scientific and Technical Periodicals, 2015, 26(4): 399-404.
[9] Fang Hongling. Characteristics analysis of zero-cited papers in five Chinese core ophthalmology journals[J]. Chinese Journal of Scientific and Technical Periodicals, 2014, 25(7): 945-948.
[10] Wei Ruibin, Wang Yan. Research on zero-citation phenomenon in foreign library and information science journals[J]. Journal of Intelligence, 2015(7): 29-33.
[11] Yang Siluo, Cheng Aijuan. Analysis of zero-citation phenomenon in library, information and archival science journals[J]. Journal of the China Society for Scientific and Technical Information, 2015, 34(3): 247-256.
[12] Hu Zewen, Wu Yishan, Yuan Junpeng. Knowledge mapping analysis of zero-citation research literature: historical development, main bodies, and high-frequency topics[J]. Information Science, 2016, 36(3): 85-91.
[13] Wen Fangfang. Analysis of causes and influencing factors of zero-cited papers in Chinese information science: based on comparison between zero-cited and highly-cited papers[J]. Information Studies: Theory & Application, 2016, 39(4): 27-31.
[14] Du Xinzheng, Ye Wenjuan, Yu Qian. Characteristics analysis of zero-cited articles in *Acta Hydrobiologica Sinica*[J]. Acta Editologica, 2016(S1): 106-108.
[15] Zhao Yue, Xiao Xiantao. Research on zero-citation phenomenon in library and information science based on topic factor analysis[J]. Chinese Journal of Scientific and Technical Periodicals, 2017, 28(7): 641-646.
[16] Kuang Shumei, Wu Yu, Han Yi. Analysis of content factors affecting zero citations in research papers[J]. Evaluation and Management, 2017, 15(3): 36-39.
[17] Li Heqiong, Zhang Xiaowei, Fu Xianbo, et al. Characteristics analysis of zero-cited papers in 10 comprehensive surgery journals[J]. Chinese Journal of Minimally Invasive Surgery, 2018, 18(4): 304-307.
[18] Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation[J]. Journal of machine learning research, 2003(3): 993-1022.

[19] Tu Y, Seng JL. Indices of novelty for emerging topic detection[J]. Information processing & management, 2012, 48(2): 303-325.
[20] Huang Lucheng, Tang Yueqiang, Wu Feifei, et al. Research on emerging topic identification methods based on multi-attribute measurement of literature[J]. Science of Science and Management of S.& T., 2015(2): 34-41.
[21] Liu Ziqiang, Wang Xiaoyue, Bai Rujiang. Construction and empirical research of a multi-dimensional topic evolution analysis model[J]. Information Studies: Theory & Application, 2017, 40(3): 92-98.
[22] Bai Rujiang, Leng Fuhai. Research on k-clique community knowledge innovation evolution methods[J]. Library and Information Service, 2013, 57(17): 86-94.
[23] Thomson Reuters. Citation thresholds. [2017-06-25]. http://www.sciencewatch.com/about/met/threshold

Author Contributions

Pan Fei: Data collection, analysis, and paper writing
Wang Xiaoyue: Research proposition and framework design
Bai Rujiang: Paper structure determination and detailed revision
Zhou Yanting: Text data processing

English Abstract

[Purpose/significance] This paper analyzes zero-cited papers in the field of environmental science from the perspective of research topics, comparing differences in article content and external indicators between zero-cited and highly-cited papers to reveal reasons for the existence of zero-cited papers. **[Method/process]** Firstly, the PLDA model was used to identify topics from the abstracts of 260 highly-cited papers and 907 zero-cited papers in domestic environmental science from the Web of Science database. Topic relevance was then found through topic similarity calculation. With topic popularity used as an internal indicator, and time of publication and the journals used as external evaluation indicators, a comparison analysis of zero-cited papers and highly-cited papers was made by combining topical content of the papers with external indicators. **[Result/conclusion]** The experimental results show that under the same research topic, the influence of the journal is the main reason that influences the citation of the paper; under different topics, the topic is the main reason leading to zero-cited papers.

Keywords: zero-cited, highly-cited, topic recognition, environmental science, comparative analysis, evaluation indicators

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