Post-prints of Long-term Literature Research in Humanities and Social Sciences Journals

Authors: Zhu Shiqin, Jiang Xinwei

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

Abstract

[Purpose/Significance] Long-term effective literature refers to publications that continue to be cited for many years after publication or maintain relatively high citation frequencies. Investigating the influence of this slowly obsolescing long-term effective literature contributes to a comprehensive evaluation of literature's value throughout its entire life cycle.

[Method/Process] This study selects citation age data of journal articles published between 2012 and 2016 from nine disciplines in source journals (2016–2017) indexed in the CSSCI database, and employs the synchronous observation method to analyze scientific literature obsolescence and long-term effective literature.

[Results/Conclusions] Citations with older citation ages in journals originate from a large collection of low-frequency cited literature. Long-term effective literature constitutes a small proportion and can be categorized into three types: growth type, mature type, and decline type, which undergo certain transformations over time. Journals with substantial long-term effective literature do not necessarily possess high impact factors. Except for psychology, each discipline has exactly one journal with significant long-term influence. The keywords of long-term effective literature resemble current disciplinary research hotspots but trace back to older time periods, thereby reflecting the origins of current hotspots while playing a role in predicting future research trends.

Full Text

Preamble

Long-Utility Literature in Humanities & Social Sciences Journals
Zhu Shiqin & Jiang Xinwei

Institute of Science and Technology Information, East China University of Science and Technology, Shanghai 200237, China
ChinaXiv Cooperative Journal, Vol. 62, No. 19, October 2018

Abstract

[**Purpose/Significance**] Long-utility literature refers to publications that continue to be cited for many years after release, either continuously over multiple years or with relatively high citation frequency. Studying the influence of these slowly aging documents contributes to a more complete evaluation of literature value across its entire lifecycle. [**Method/Process**] This study examines citation age data from journal articles published between 2012-2016 across nine disciplines in the CSSCI source journal database (2016-2017 edition), analyzing scientific literature obsolescence and long-utility literature through synchronic observation methods. [**Result/Conclusion**] The findings reveal that older citations in journals originate from a large collection of low-frequency cited literature. Long-utility literature constitutes a small proportion and can be divided into three types: growing, mature, and declining, with transformations occurring among these types over time. Journals with substantial long-utility literature do not necessarily have high impact factors. Except for psychology, each discipline has exactly one journal with significant long-utility influence. The keywords of long-utility literature resemble current disciplinary research hotspots but trace back to more distant time periods, reflecting the origins of current hotspots while potentially predicting future research trends.

Introduction

Since C.F. Gosnell first proposed the concept of literature obsolescence in his 1943 doctoral dissertation [1], related research has spanned over 70 years, yielding multiple mature evaluation metrics. Traditional journal literature evaluation primarily calculates citation frequency within specific timeframes. Classic indicators include J.D. Bernal's "literature half-life," adapted from physics [2]; D.J. Price's Price Index [3]; and E. Garfield's impact factor, adopted by JCR in 1975 [4]. These metrics emphasize recent literature in their definitions, resulting in relatively short observation periods and susceptibility to the stock effect of highly-cited documents. Consequently, metrics have continuously evolved: J.E. Hirsch proposed the H-index in 2005 [5]; L. Egghe developed the G-index based on the H-index [6]; and Yu Liping introduced the historical impact factor to eliminate total citation frequency effects [7]. While these developing metrics enrich the evaluation system from different perspectives, they remain limited to highly-cited literature, feature short observation windows, and suffer from historical data stock effects, lacking evaluation for older publications.

Literature obsolescence research employs two primary dimensions: diachronic observation and synchronic observation. Diachronic observation studies obsolescence through continuous monitoring of a literature's complete lifecycle, while

synchronic observation improves data accuracy from a citation perspective. Xie Haixiu and Wang Hongxin [8] utilized grey relational analysis and synchronic observation methods, arguing that synchronic observation offers stability suitable for disciplinary literature aging research. R. Vernon [9] first applied lifecycle theory to products, viewing literature aging as a process from publication through growth, peak, and decline phases (see Figure 1 [Figure 1: see original paper]). Each phase length varies according to annual citation frequency trends post-publication, with boundaries marked by citation peaks: the period before peak citation counts represents the growth phase, the adjacent period around the peak constitutes the maturity phase, and the period of declining citations after the peak marks the decline phase. Domestic researchers Wei Ruibin et al. [10] integrated lifecycle theory in their National Social Science Fund project, proposing concepts of cited lifecycle and citation proportion metrics. Subsequently, Zhang Zhongwen et al. [11] employed survival analysis using Kaplan-Meier estimators to analyze literature survival rates, introducing concepts such as "citation lifespan," "citation resurrection," and "survival citation counts."

Current literature obsolescence research focuses on aging patterns, models, and specific disciplinary trends, with limited investigation into the influence evaluation of slowly aging literature. Traditional metrics emphasize high citation scenarios. For literature in the aging phase, key influence indicators like citation frequency are significantly lower than during peak periods. However, considering their longer publication history, this study argues they can effectively reflect influence—specifically, long-term influence. V. Larivière et al. [12] confirmed that researchers increasingly rely on older literature. This study integrates lifecycle theory with traditional bibliometric methods, employing synchronic observation using CSSCI data from nine disciplines' journals (2012-2016) to analyze citation age distributions, calculate risk rates of literature aging through survival analysis, and define this slowly aging yet continuously cited literature as "long-utility literature" for analysis, thereby supplementing journal evaluation systems.

Data Processing

Data were sourced from CSSCI source journals (2016-2017 edition), selecting citation age data for nine disciplines' journal articles published 2012-2016, downloaded in February 2017. Some data and journals were excluded for two reasons: (1) Inconsistent journal founding dates prevented meeting the required retrospective timeframe, and some journals had missing database data, compromising stability and validity; (2) A small number of cited works predated 1975, concentrated in the 1930s or early post-liberation period, creating gaps exceeding 10 years during academic turbulence. These citations were removed. Using synchronic observation with the cited literature's publication year as baseline (Year 0), the maximum retrospective range was set at 40 years. Final selected disciplines and journal counts appear in Table 1 .

During screening, CSSCI data records contained minor errors in four areas:

punctuation errors in titles, typographical errors, missing subtitles, and missing authors. These errors caused computational programs to split single publications into multiple entries due to title discrepancies, drastically reducing screening efficiency, creating duplicate counts, and fragmenting citation statistics, thereby affecting long-utility literature identification.

Long-Utility Literature Research

3.1 Definition of Long-Utility Literature

Bibliometric concepts commonly include "classic literature" and "highly-cited literature." Wang Jingshan [13] defined classic literature as possessing importance, standardization, and timelessness. This study's focus—literature maintaining certain citation volumes during the aging phase—shares similarities with classic literature regarding time span and value but differs in scope. Classic literature achieves high short-term utility and becomes hotspots; even when superseded, their paradigms remain representative. In contrast, long-utility literature encompasses broader scope, reflecting sustained utility during aging rather than short-term hotspots, thus supplementing journal evaluation systems.

This study examines literature cited multiple years after publication, integrating lifecycle theory with time and frequency dimensions to propose the "long-utility literature" concept: publications that, after considerable time, remain continuously cited over multiple years or maintain relatively high citation frequency.

Long-utility literature requires precise temporal thresholds defining "considerable time" and sufficient continuous citation. Literature meeting the temporal threshold but lacking sufficient continuous citation frequency is termed "duration literature." Long-utility literature is a subset of duration literature (Long-utility literature Duration literature).

3.2 Temporal Threshold for Long-Utility Literature

This study employs Cox regression, a semi-parametric estimation in survival analysis (see Equation (1)), to verify factors influencing literature aging.

$$h(t) = h_0(t)exp(\beta_{1x}1 + \beta_{2x}2 + \beta_{3x}3 + ...\beta_{px}p)$$

Using synchronic observation, we grouped data by year, compiling citation age distribution tables for each journal's 2012-2016 publications. With publication year as Year 0, we observed citation age distributions retrospectively. For example, a 2000 publication cited in 2013 was recorded as Year 13, while the same 2000 publication cited in 2016 was recorded as Year 16. Citation counts for each Year N were aggregated across groups, reflecting each journal's 2012-2016 citation age distribution. This distribution calculated cumulative citation frequency, citation age peaks, Price Index, and disciplinary half-life. These

traditional bibliometric indicators, combined with journal disciplines as influencing factors, were modeled using Stata. Regression analysis plotted overall humanities and social sciences literature aging risk rates (Figure 2 [Figure 2: see original paper]), showing risk rates beginning to rise from Year 10, approaching maximum near Year 30 at approximately 0.6—indicating a 60% probability that journals not experiencing aging events by Year 30 will receive no citations that year.

Plotting duration literature by ascending citation frequency (Figure 3 [Figure 3: see original paper]) reveals consistent patterns across all disciplines: under synchronic observation, most literature cited fewer than 1-2 times after 10 years, demonstrating minimal long-term utility. At the inflection point, curves approach 90-degree transitions, rising sharply. Long-utility literature utility thresholds were calculated based on average citation frequencies around these inflection points.

3.3 Utility Threshold for Long-Utility Literature

Given disciplinary variations in citation volume, half-life, and extreme values, fixed thresholds are inappropriate. Using duration literature citation frequency distributions, this study established discipline-specific citation frequency thresholds. Table 2 presents the calculated utility thresholds.

3.4 Screening Long-Utility Literature

Based on the definition, within the 5-year synchronic observation window, long-utility literature must be published over 10 years prior (e.g., pre-2006 publications cited in 2016) and exceed the discipline-specific utility threshold calculated in Table 2.

Two additional criteria were established: (1) High citation frequency—exceeding the discipline-specific threshold within the defined timeframe; (2) Continuous citation—being cited across different years within the 5-year observation period, reflecting sustained and forward-looking value as origins for subsequent research. Combining these criteria, long-utility literature was classified into three types: Type A satisfies both conditions (high frequency and continuous 5-year citation); Type B satisfies only continuous citation (5-year continuity but low frequency); Type C satisfies only high frequency (high citation but not continuous across all 5 years).

3.5 Type Transformation in Long-Utility Literature

Literature entering the aging phase with minimal but persistent citations becomes duration literature upon reaching the aging threshold (10 years). Over time, continuously cited low-frequency literature transforms into Type B, while high-frequency literature recently surpassing 10 years becomes Type C. Type C literature maintaining high citation frequency over time upgrades to Type A; otherwise, it ceases to be long-utility. Type A literature losing high citation

frequency but maintaining annual citations downgrades to Type B; failure to maintain continuous citation removes it from long-utility status. Figure 5 [Figure 5: see original paper] illustrates these transformation trends. Accordingly, Types A, B, and C correspond to mature, declining, and growing literature, respectively.

3.6 Relationship Between Long-Utility Literature and Impact Factor

Examining journals containing long-utility literature reveals minimal disciplinary influence differences. However, when incorporating quantity statistics, substantial inter-journal differences emerge within management, sociology, and psychology (Table 3). Table 4 lists the top 5 journals by long-utility literature count per discipline (psychology and philosophy have fewer than 5). Comparing top-ranked journals with subsequent rankings shows that, except for psychology, the leading journal's long-utility literature count at least doubles the second-ranked journal's count—approaching 4× in management, sociology, and philosophy, and 8× in economics. This substantial gap indicates that, excepting psychology, the top-ranked journal holds extremely important positions in their disciplines. Psychology features two high-influence journals with concentrated long-utility literature, while other journals follow patterns similar to other disciplines. Excepting library and information science, second and third-ranked journals also show significant quantitative differences from fourth and fifth-ranked journals, forming a second tier.

Impact factor represents a journal's average citation rate for recent two-year publications. Comparing 2015 impact factors from the VIP Database with long-utility literature counts reveals no significant correlation. Journals with the most long-utility literature never have the highest impact factors. In management, education, library and information science, and political science, the highest-impact journals are not among the top 5 long-utility literature journals. Some low-impact journals also contain long-utility literature.

This demonstrates that impact factor, focusing on recent publications, evaluates near-term influence but inadequately assesses duration literature's long-term impact. As a dynamic metric, impact factor varies annually. Long-utility literature counts, based on synchronic observation, represent time-sectional results that also change across years but avoid cumulative effects from complete lifecycle observation. Thus, incorporating long-utility literature counts into journal evaluation provides more comprehensive influence assessment.

Knowledge Network Analysis of Economics Long-Utility Literature

Due to limited long-utility literature quantities, this study selected the economics discipline (with the largest volume) for knowledge network analysis. As CSSCI cannot retrieve pre-1998 literature details, synchronic observation was employed via CNKI to obtain and screen basic information for CSSCI

economics long-utility literature published 1976-2016. CiteSpace analyzed keyword co-occurrence; Table 5 presents keywords appearing $ $3 times, sorted by year. These keywords concentrate primarily in 1999-2003 (25 keywords), with "economic growth" appearing most frequently, followed by "listed companies," "corporate governance," and "human capital" (each >10 occurrences). This reflects that economic growth provides the research backdrop, within which corporate governance—especially listed company management—constitutes a research hotspot.

Ma Feicheng et al. [14] conducted bibliometric analysis of 1998-2011 economics literature, with a timeframe closely matching this study's keyword concentration. Comparison reveals that long-utility literature keyword co-occurrence results closely match Ma's findings from all CSSCI economics literature, but with earlier temporal nodes. For instance, Ma identified "state-owned enterprise reform" and "listed companies" as 1998-2000 hotspots, while this study traces the node event to 1995. Similarly, Ma located "human capital" research in 2001-2003, while this study identifies 1996 as the emergence point. Despite lower citation counts than peak-period literature, long-utility literature keywords closely match overall literature patterns while predating them, suggesting long-utility literature marks research hotspot origins and may predict future trends.

Figure 6 [Figure 6: see original paper] presents CiteSpace clustering results: larger circles and fonts indicate higher keyword frequency, while similar colors denote closer associations. The clustering reveals two primary research dimensions under the economic growth backdrop: (1) enterprise operation management, market policy, and resources; (2) management techniques, evaluation metrics, and methodologies. High-centrality keywords cluster around 2000-2002, near the median year for the three long-utility literature types.

Conclusion

This study integrates lifecycle theory with synchronic observation of nine CSSCI disciplines, using survival analysis to establish a 10-year threshold for long-utility literature and screening literature based on continuous citation years and citation volume. Key findings include: (1) Researchers cite duration literature extensively but non-concentratedly, with most receiving only 1-2 citations, meaning older citations originate from large collections of low-frequency cited literature. (2) Except for psychology, each discipline has exactly one journal with significant long-utility influence, exhibiting a "one superpower + multiple strong players" hierarchical structure. (3) Qualified long-utility literature is extremely rare, classifiable into three types (growing, mature, declining) that transform over time, reflecting journal aging patterns. (4) Economics long-utility literature analysis using CiteSpace demonstrates that long-utility literature effectively supplements evaluation of older publications. Keyword co-occurrence patterns resemble current disciplinary hotspots but originate earlier, reflecting hotspot sources while potentially predicting future trends.

This study is limited to nine randomly selected CSSCI disciplines; science and engineering disciplines require further investigation. The 10-year threshold derives from this dataset's regression and may not be absolute for single-discipline or other-year studies. Synchronic half-life observation does not capture complete aging lifecycles and varies with selected years, but this dynamic characteristic—similar to impact factor—makes long-utility literature valuable for supplementing post-peak lifecycle evaluation and achieving more complete literature value recognition.

References

[1] GOSNELL C F. The rate of obsolescence in college library book collections as determined by an analysis of the three select lists of books for college libraries[D]. New York: New York University, 1943.
[2] BERNAL J D. The transmission of scientific information: a user's analysis[C]//Proceedings of the international conference on scientific information. Washington: National academy of science, 1959: 85-95.
[3] PRICE D J. Citation measures of hard science, soft science, technology, and nonscience[M]//NELSON C E. Communication among scientists and engineers. Massachusetts: Heath Lexington Books, 1970: 3-22.
[4] 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.
[5] HIRSCH J E. An index to quantify an individual's scientific research output[J]. Proceedings of the National Academy of Sciences of the United States of America, 2005, 102(46): 16569-16572.
[6] EGGHE L. Theory and practise of the g-index[J]. Scientometrics, 2006, 69(1): 131-152.
[7] YU L. Historical impact factor: a new stock evaluation indicator for academic journals[J]. Library and Information Service, 2015, 59(2): 89-92.
[8] XIE H, WANG H. Grey relational analysis of diachronic and synchronic observation of literature obsolescence[J]. Journal of Intelligence, 1997, 16(6): 24-26.
[9] VERNON R. International investment and international trade in the product cycle[J]. The quarterly journal of economics, 1966, 80(2): 190-207.
[10] WEI R, TIAN D, WU Y. Comparative study of journal citations based on cited lifecycle[J]. Library and Information Service, 2014, 58(1): 79-85.
[11] ZHANG Z, XU T, DONG X, et al. Definition and application of survival citation counts for academic journal articles[J]. Acta Editologica, 2015, 27(4): 316-319.
[12] LARIVIÈRE V, ARCHAMBAULT E, GINGRAS Y. Long-term variations in the aging of scientific literature: from exponential growth to steady-state science (1900-2004)[J]. Journal of the Association for Information Science and Technology, 2008, 59(2): 288-296.
[13] WANG J, WANG J. Conceptual analysis of classic literature[J]. Library and Information, 2006(1): 103-105, 115.

[14] MA F, LIU X. Structure, evolution and hotspot detection of knowledge networks: bibliometric analysis of CSSCI economics literature (1998-2011)[J]. Information Science, 2014, 32(7): 3-8.
[15] CHEN C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature[J]. Journal of the Association for Information Science and Technology, 2006, 57(3): 359-377.

Author Contributions:
Zhu Shiqin: Proposed research ideas and methods, designed the paper framework, revised the manuscript.
Jiang Xinwei: Processed and analyzed data, wrote the paper.

Author Affiliations:
Zhu Shiqin (ORCID: 0000-0001-5646-9287), Deputy Director, Associate Research Librarian, Master's Supervisor, E-mail: shqzhu@ecust.edu.cn
Jiang Xinwei (ORCID: 0000-0001-8061-7972), Master's Student

Received: February 6, 2018
Revised: April 24, 2018
Pages: 93-100
Responsible Editor: Wang Chuanqing

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