Research HotSpots and Topic Evolution of Long Tail Theory in China: a CiteSpace Knowledge Graph Analysis

Chunlan Li[a], Qiuying Ma[b*]

College of Economics and Management, Guangxi Normal University, Guilin, Guangxi, China

[a]lcl_lee@126.com
[b*]Corresponding author: 2532218090@qq.com

Abstract. The Long Tail Theory posits that given a sufficiently large platform for storage and distribution, the combined market share of niche or low-demand products can rival or even surpass that of a few best-selling items. . This study aims to analyze the research hotspots and thematic evolution of Long Tail Theory in China, providing a reference for both theoretical research and practical applications in the country. Drawing data from China National Knowledge Infrastructure (CNKI) and employing CiteSpace visualization software, this paper conducts a visual analysis of 1,528 relevant documents. The findings suggest that in term of authorship and institutional distribution, cooperation remains dispersed, and a core research cooperation network has yet to form.. Research hotspots primarily focus on the themes of "online marketing, libraries, and new media"; while future research tredns are anticipated to concentrate on areas such sa "knowledge payment, sharing economy, and online streaming" .

Keywords: Long Tail theory; CiteSpace; visualization analysis; hot topics; research frontier

1 Introduction

The term "Long Tail" was initially coined by Chris Anderson, the editor-in-chief of Wired magazine, in 2004. Since then, numerous scholars have engaged with the Long Tail Theory, leading to a proliferation of research outcomes. Faced with this extensive body of literature, it becomes imperative to efficiently comprehend and identify the forefront and focal points of the Long Tail Theory. Scientific Knowledge Graph, a structured data model for organizing scientific knowledge, offers a solution. It seamlessly integrates traditional bibliometric techniques with modern text mining and visualisation methods, facilitating the effective discovery and utilization of scientific knowledge.For example, through knowledge mapping, researchers can swiftly locate relevant literature, discern emerging research trajectories, and explore the interplay and cross-pollination of ideas across diverse fields[1]. Among the various tools for knowledge graph analysis tools, CiteSpace stands out as one of the most widely employed. Therefore, this study employs CiteSpace to construct a knowledge map spanning the period from 2005-2023, specifically focusing on the Long Tail Theory.

Y. Kuang et al. (eds.), *Proceedings of the 2024 5th International Conference on Education, Knowledge and Information Management (ICEKIM 2024)*, Atlantis Highlights in Computer Sciences 22,
https://doi.org/10.2991/978-94-6463-502-7_84

Through correlation analysis, the aim is to gain comprehensive insights into the current landscape and evolving trends within this domain of research.

2 Literature Sources and Research Tools

2.1 Literature Sources and Processing

CiteSpace software demands data with a balance of broad coverage and high accuracy. With this criterion in mind, the present study conducted a topic search on the CNKI retrieval platform using a Boolean logic combination of "long-tail theory" or "long-tail effect" for the period from 2005-2023. This search yielded a total of 1648 articles. After removing duplicates, conferences, newspaper articles, and other documents not directly relevant to the search topic, a final set of 1528 documents was obtained.

2.2 Research Tools

Developed by Dr Chaomei Chen, a Chinese American, Citespace software offers an intuitive and visual approach to analyzing literature through knowledge mapping . This tool aids in identifying and presenting the research advancements and cutting-edge directions within a particular subject area[2]. Leveraging CiteSpace, the present study generated a knowledge map of issuing authors , institutions, and keyword co-occurrences related to the long-tail theory. This analysis further exploration of research hotspots and frontierss within the topic, offering insights into its development dynamics.

3 Overview of Long Tail Theory Research

3.1 Chronological Analysis of Literature

This section undertakes a chronological examination of the literature pertaining to long tail theory. The literature sample was meticulously tabulated, and the resultant trajectory depicting growth of research on long tail theory is illustrated in Figure 1.

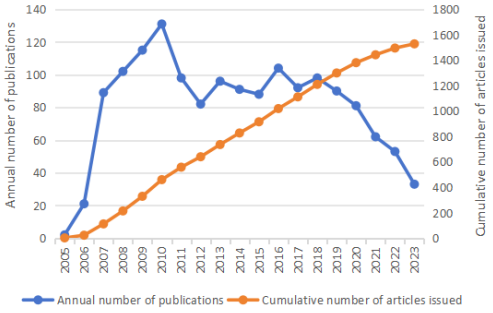

Fig. 1. Trend map of literature growth

According to Figure 1, the research on long tail theory can be delineated into three distinct phases: (1) Rapid growth period (2005-2010): During this phase, the volume of literature experienced a significant surge, escalating from a mere 2 articles in 2005 to 131 articles by 2010. (2) Stable development period (2011-2016): This epoch witnessed the widespread application of long tail theory across various domains. The theoretical framework matured, garnering widespread acknowledgment within the academic community.(3) Temporary decline period (2017-2023): Despite an annual publication volume exceeding 20 articles, this period experienced a transient downturn. Over the course of seven years, 499 papers were published, constituting 32.87% of the total literature. This indicates a sustained interest among scholars in long tail theory, albeit at a slightly diminished level compared to the preceding phase.

3.2 Analysis of Core Authors

Core authors are pivotal in steering and advancing the progression of a research domain[3]. The author Collaboration Network provides a visual representation of these core authors and their collaborative efforts within a specific research field[4]. In this investigation, it conducted an analysis of all the citations of the sample literature employing the author analysis function of CiteSpace. The finding are depicted in Figure 2.

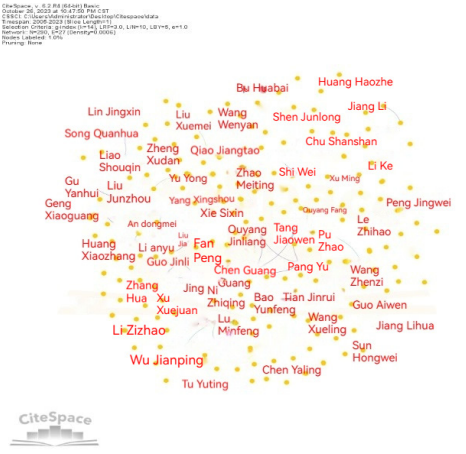

Fig. 2. Author Collaboration Network

Figure 2 shows that the nodes within the network are fragmented and display sparse connections. This This indicates a limited level of collaboration among authors in this research domain, with most operating independently and forming few close partnerships.Therefore, to foster consensus and cohesion in long tail theory research, there is a need for enhanced emphasis on team building in the future endeavors.

3.3 Analysis of Research Institutions

The distribution of research institutions not only offers insights into knowledge generation, dissemination and evolution within a specific field[5], but also facilitates an understanding of inter-institional collaboration, thereby informing the development of novel research partnerships[6]. In this study, leveraging the institution analysis function of CiteSpace, we examine the affiliations of authors and their collaborative networks in this research domain in this research domain, as shown in Figure 3. The distribution of research institution appears dispersed without distinct clustering, with the notable exception of a prominent cooperation network cnetered around the School of Journalism at Renmin University of China. However, the majority of institutions appear to publish articles independently, underscoring a prevalent trend of individualized research pursuits with limited inter-institutional cooperation within this field.

Fig. 3. Institutional cooperation Networks

4 Research HotSpots and Evolutionary Lineage

4.1 Research HotSpots

Keyword Co-occurrence Analysis

Keyword serve as specialized term reflecting the core research content of academic papers, offering concise summaries of articles. Moreover, keywords appearing frequently in the literature within a specific field illuminate the research hotspots[7]. In this investigation, employing the Keyword analysis function of CiteSpace, we visualize and analyze keywords associated with long tail theory literature, with results presented in Figure 4.

In Figure 4, each node represents a keyword, wiht node size indicative of keyword frequency. Connecting lines between nodes donate the degree of association amnong keywords. Notably, the two larger nodes, "long-tail theory" and "long-tail effect", exhibit the highest frequency of occurrence and strongest connectivity with other keywords. This underscores their prominence within the analysed literature and their

central role in the network. In addition, keywords such as "law of two-eighths", "network marketing", "library", "new media" and "niche market" also feature prominently, demonstrating subdstantial connectivity and aggregation. This high-frequency keywords, interconnected with others, delineate the focal points of long tail theory research.

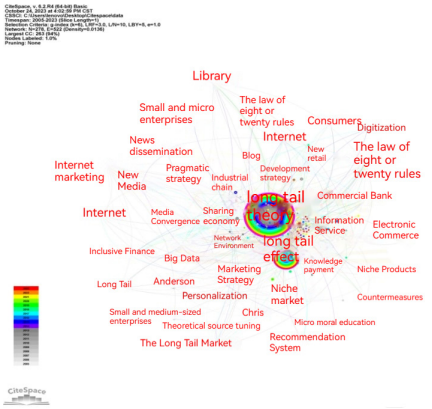

Fig. 4. Keyword co-occurrence knowledge graph

Keyword Clustering Analysis

To deepen our comprehesion of the research hotspot surrunding "Long Tail Theory", this study utilizes the LLR algorithm to conduct clustering based on the keyword co-occurrence knowledge map. Subsequently, we derive the keyword clustering knowledge map, depicted in Figure 5.

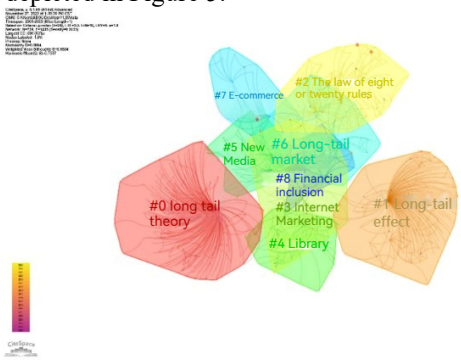

Fig. 5. Knowledge map of keyword clustering

The clustering structure depicted in Figure 5 provides insight into the research hotspots within the realm of long-tail theory. From the clustering outcomes, it is evident that the primary focus of long-tail theory research revolves around nine main categories, including #0 long-tail theory, #1 long-tail effect, #2 the law of twenty-eight, #3 network marketing, #4 libraries, #5 new media, #6 long-tail market, #7 e-commerce, and #8 inclusive finance.

4.2 Evolutionary Trajectory

The keyword time zone map offer a clear depiction of the evolutionary trajectory of a research field over time, highlighting mutual influence relationships[8]. Therefore, this study employs the "Time Zone" function of CiteSpace to visualize and analyze the time zones of keywords, with results presented in Figure 6. It reveals the evolutionary progression of long tail theory research, transitioning from its core concepts of long tail theory and long tail effect to encompass themes such as libraries, online marketing, niche markets, new media, information services, inclusive finance, sharing economy, and self media.

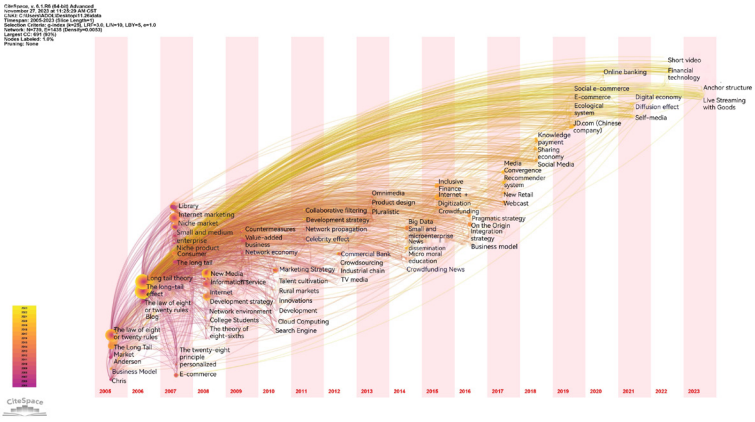

Fig. 6. Keyword time zone mapping

Keyword emergence serves as a valuable indicator of research hotspot and directional trends during a specific timeframe. The intensity of keyword emergence reflects the academic attention directed toward a particular keyword and research hotspot, with higher emergence intensities indicating greater academic interest[9]. The keyword occurrence mapping of literature pertaining to the long tail theory research illustrated in Figure 7. Withen the figure, "Strength" denotes the magnitude of keyword emergence, with hier values indicating increased citations and greater influence withen the literature. Based on the keyword time zone mapping and emergence map, this study delineates the evolutionary trajectory of research topics within the realm of long tail theory into three stages:

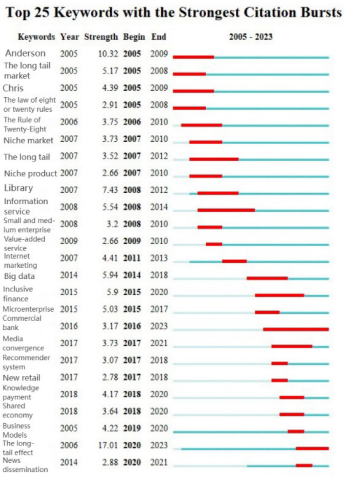

Fig. 7. Long tail theory research keyword emergence map

(1) Early research stage (2005-2010): In addition to the keywords "Anderson and Chris", the keywords that emerged with greater intensity in this phase were long tail market (5.17) and the rule of two-eight (3.75). The Long Tail Market demonstrates the core concepts of the Long Tail Theory through concrete examples, which helps people to understand the meaning of the Long Tail Theory more deeply. However, because the proposal of the long tail theory in that period brought some impact on the traditional two-eight rule, the research value position of scholars at that stage mainly stayed on the dialectical relationship between the long tail theory and the two-eight rule. And people have different opinions about the relationship between the long tail theory and the law of two-eighths, so far there is still no a unified statement, the academic community has appeared two mainstream views. One view is that the long tail theory challenges the traditional law of two-eight by exploiting the Internet's ability to expand product accessibility, thus becoming a major paradigm for economic behaviour and management[10]. Conversely, the other view is that the long tail theory complements and extends the law of two-eight, as it primarily helps to identify niche markets, while the law of two-eight remains dominant in each market[11].

(2) Mid-term development stage (2011-2016): The frontier keywords in this phase mainly include library (7.43), big data (5.94), financial inclusion (5.9) and information service (5.54). The research focus of this phase is the application of the long tail theory in the field of education and culture as well as information technology, which indicates that the research focus of scholars in this period has been transitioned from theoretical exploration to practical application in the field of long tail theory. For example, in the application of libraries, scholars advocate strengthening the construction of digital resource system[12], strengthening the professional training of librarians and expanding the long tail of service mode can effectively strengthen the long tail effect of digital library information service[13]. Based on the above analyses, applying the long tail

theory to real life can help us break the traditional mass market thinking, better meet the diversified individual needs, and create more value.

(3) Innovation research stage (2017-2023): The keywords with greater intensity at this stage include business model (4.22), knowledge payment (4.17), media integration (3.73) and sharing economy (3.64), etc. At the same time, emerging keywords such as "webcasting, self-media and live streaming with goods" emerge in the time zone map at this stage. Among them, "knowledge payment and sharing economy" are part of the business model innovation, and "webcasting and live streaming with goods" are the deep-level innovation of media integration. For example, in terms of webcasting, some scholars have put forward some constructive opinions based on the long tail theory model on the market mechanism and content mechanism of webcasting in two dimensions, which not only reshapes the value perception of webcasting, but also promotes the long-term development of the live broadcasting industry[14]. It can be seen that the long tail theory can help us understand the business logic behind these emerging phenomena, further guide enterprises in marketing, product development and other aspects of decision-making, so as to better help enterprises adapt to market changes, meet consumer demand, and promote the sustainable development of enterprises.

Throughout the three stages of development of the long tail theory, we find that its development course follows the general law, and the whole development course reflects the logical development path from theoretical research to practical transformation and then to continuous innovation. However, scholars usually use the long tail theory to qualitatively analyse a phenomenon in their research, and few scholars combine the long tail theory with specific data to analyse a phenomenon. In the future, scholars can combine the long tail theory with specific data to analyse a phenomenon in depth, in order to verify the feasibility of the theory and put forward more specific and operable suggestions.

5 Conclusions

Utilizing bibliometric method and CiteSpace visualization software, this study examines the research landscape, hotspots and evolution of long tail theory using Chinese literature from the China Knowledge Network as the research samples. The findings yield the following conclusions: (1) Research profile of long tail theory: the trajectory of research interest in long tail theory exhibits both historical shifts and epochal changes. Despite the presence of numerous authoritative authors and institutions, collaboration remains diffuse, lacking the formation of a cohesive core research consortium. (2) Research hotspots and evolution: The hotspots of long tail theory research focus on themes such as "library, network marketing and new media", while the trajectory of research trends emphasizes the application of long tail theory in domains like "knowledge payment, sharing economy and webcasting".

Moving forward, several avenues for advancing research and development in Long Tail Theory are apparent: (1) Enhanced Collaboration: There exists a need to bolster cross-institutional and interdisciplinary cooperation to foster deeper insights and synergies. (2)Focus on Emerging Themes: Future research endeavirs should concentrate

on exploring the instersection of long tail theory with emerging fields such as new media and business model innovation. (3)Methodological Diversification: In addition to theoretical analyses, the integration of more empirical research methodologies is warranted to enrich understanding and validate theoretical constructs.

References

1. Chen, Y., Chen, C.M., Liu, Z.Y., et al. Methodological functions of CiteSpace knowledge graph[J]. Research in Science, 2015, 33(2): 242-253.DOI:10.16192/j.cnki.1003-2053.2015.02.009.
2. Hou, J.H., Hu, Z.G. Review and prospect of CiteSpace software application research[J]. Modern Intelligence, 2013, 33(04): 99-103.DOI:10.3969/j.issn.1008-0821.2013.04.022.
3. Wu, H.J., Dan, X.Q., Shu, Y., et al. National Wetland Parks in China:Current Situation, Challenges and Countermeasures[J/OL]. Wetland Science, 2015, 13(3): 306-314.DOI:10.13248/j.cnki.wetlandsci.2015.03.006.
4. Wang, G.W. Visual analysis of the progress of library space research in China[J]. Library Work and Research,2022,(02):64-76.DOI:10.16384/j.cnki.lwas.2022.02.009.
5. Sun, H.S. Research on scientific output and co-operation status of domestic library and intelligence research institutions[J]. Journal of Intelligence, 2012, 31(2): 67-74.DOI:10.3969/j.issn.1002-1965.2012.02.014.
6. Zhang, Z.Y., Tang, Z.L., Zhang, Z.H., et al. Bibliometric analysis of national wetland park research based on CiteSpace[J]. Journal of Ecology ,2023, (22):1-10.DOI:10.20103/j.stxb.202212043502.
7. Hu, Z.G., Chen, C.M., et al. From citation-based to quotation-based - a new method to count the total number of citations[J]. Library and Intelligence Work, 2013, 57(21): 5-10.DOI:10.7536/j.issn.0252-3116.2013.21.001.
8. Xiong, H.X., Li, J.L. Knowledge mapping analysis of archival research in China in the past ten years based on CSSCI [J/OL]. Archival Research, 2020(3): 16-24.DOI:10.16065/j.cnki.issn1002-1620.2020.03.003.
9. Liu, Y.F., Wang, X.X., Yuan, Y.J. Research on hot spots and theme evolution of library career research in China[J/OL]. Library Work and Research, 2021(4): 65-72+97.DOI:10.16384/j.cnki.lwas.2021.04.011.
10. He, C.Y. The "Long Tail Theory" that Subverts the "Law of Two or Eight"[J]. China Urban Finance, 2008(1): 59.
11. Jin, H.W. The Long Tail Theory is just a supplement to the Law of Eight[J]. China Small and Medium-sized Enterprises, 2007(1): 70-73.
12. Xia, Y. Long Tail Theory and its Implications for Digital Library Information Services[J]. Business Times, 2013(7): 98-100.DOI:10.3969/j.issn.1002-5863.2013.07.045.
13. Hou, L.L. Innovation of university library search service based on long tail theory[J]. Digital Library Forum, 2015(2): 60-64.DOI:10.3772/j.issn.1673-2286.2015.02.010.
14. Huang, X.B. Mobile Internet+Socialisation:A study of webcasting under the perspective of long tail theory[J]. New Media Research,2017,3(04):75-78.DOI:10.16604/j.cnki.issn2096-0360.2017.04.034.