Education Reform and Development, 2025, Volume 7, Issue 6
https://ojs.bbwpublisher.com/index.php/ERD
Online ISSN: 2652-5372
Print ISSN: 2652-5364

Analysis of Research Trends in the Field of Museology: Application of Node2Vec and Referential Network Modelling

Rongkang Pei*

Nanjing Institute of Tourism and Hospitality, Nanjing 211100, Jiangsu, China

**Author to whom correspondence should be addressed.*

Abstract: This study uses the Node2Vec network embedding technology combined with citation network modelling to systematically analyze the knowledge evolution in the research field of museology. Based on 6,726 relevant documents included in the Scopus database from 1948 to 2023, this research constructs a high-dimensional citation network and applies cluster analysis and regression modelling to explore the theme development trends, core research themes, and their influence in this field. The research finds that museology research mainly focuses on cultural heritage protection, digital technology applications, museum education, and public participation, and has shown a trend of interdisciplinary integration in recent years. In addition, with the help of IPY (Intrinsic Publication Year) analysis, this study reveals the inter-generational evolution of research hotspots and their high synchronization with policy revisions and technological innovations (such as the rise of augmented reality technology). The research shows that the knowledge diffusion model of modern research has shifted from traditional collection management to digital-based knowledge sharing and social practice. Finally, this study suggests that future academic research can combine Temporal Graph Attention Networks (TGAT) to improve the representational ability of early literature and multilingual knowledge flows to comprehensively understand the disciplinary development path of museology.

Keywords: Bibliometrics; Museology; Citation Network Analysis; Node2Vec; Scientometrics

Online publication: July 7, 2025

1. Introduction

Museums, libraries, and archives together represent the main institutions for protecting cultural heritage and are often referred to as "memory institutions [1]. These entities preserve and present humanity's collective memory, contributing to the public's understanding of history, culture, and art. Among them, museums hold a unique position as places where tangible interactions with cultural relics promote education and lifelong learning [2].

In recent decades, research on museum practices and their evolution has gained considerable momentum,

reflecting the increasing social value of cultural protection and education [3]. This trend is consistent with technological progress, which has facilitated broader public access to museum collections through digital archives and online exhibitions. The integration of information science and museology has further highlighted the interdisciplinary nature of this field, leading to a re-examination of traditional boundaries and practices within these institutions [4].

In addition, although Library and Information Science (LIS) programs have long incorporated museum practice courses, the integration of these academic disciplines has reignited interest in their research output. Research shows that there is an increasing overlap between the skills required in librarianship and those needed for museum work, emphasizing collaborative and shared approaches. However, despite these connections, empirical evidence regarding the publication trends of museum-related research in LIS journals remains scarce.

To address this gap, it is necessary to explore the publication trends, thematic focuses, and methodological preferences of museum research in the context of library and information science. By analyzing data from the past decade, this study aims to understand how museum research has developed within the LIS field, contributing to a broader discussion of the interdisciplinary nature of these fields.

2. Literature review

In the library and information field, citation analysis and bibliometric research methods have been widely applied. These methods provide rich perspectives for exploring academic trends, research topics, collaboration networks, and knowledge structures. With its powerful visualization capabilities and specially designed algorithms, CiteSpace has become the preferred tool for scholars to study knowledge structures and the evolution of academic disciplines [5]. CiteSpace can help researchers identify the key points, research hotspots, and their implicit academic paths in their research fields by analyzing the co-citation relationships and keyword co-occurrence networks in the literature.

However, although traditional citation analysis tools such as CiteSpace have certain advantages in revealing disciplinary development trends and identifying knowledge breakpoints [6], they have limitations in analyzing the complex network structure of literature and predicting research trends. CiteSpace focuses on static analysis and fails to fully capture the potential structure and deeper connections between nodes in the literature network. In addition, it mainly relies on co-citation and co-occurrence analysis and has difficulty reflecting the non-linear or high-dimensional relationships between nodes (such as literature or research topics).

Therefore, this study proposes the methods of Node2Vec and referential network modelling. Node2Vec is a network embedding algorithm based on random walks and deep-learning technology, capable of capturing the complex relationships between network nodes in a high-dimensional space. The advantage of this method is that it can embed nodes in a low-dimensional vector space while retaining information about the network structure, making subsequent cluster and regression analyses more efficient and accurate. This is particularly important for studying the complex interactions between nodes in the network.

On the other hand, citation network modelling provides a basis for exploring the associations between documents from the perspective of citations. By constructing a citation network, researchers can regard each document as a node, the citation relationship as an edge, and further apply network embedding methods such as Node2Vec to effectively analyze the spatiotemporal evolution of literature and the changes of research

hotspots. This method not only improves the in-depth understanding of the citation structure but also enables machine-learning analysis, such as clustering and predictive modelling through embedded vectors, to identify potential research trends and areas of future interest.

Therefore, this study will combine Node2Vec and citation network modelling methods with traditional citation analysis techniques to provide a new perspective for exploring the evolution and development trends of research themes in the field of museology within the library and information field.

3. Methodology

This study uses data obtained from the Scopus database, which is known for its comprehensive coverage of high-quality academic publications in different disciplines. After data cleaning, non-journal articles such as reports and editorials were excluded. The dataset covers 6,726 articles related to the field of museology published between 1948 and 2023. These articles cover a wide range of topics, from theoretical frameworks to practical applications within museums, providing a broad basis for analyzing research trends and theme development.

To explore the theme evolution within the field of museology, the articles were divided into different time intervals (for example, five-year periods) according to their publication years. This study adopts a method that combines bibliometrics and complex network analysis to systematically explore the research dynamics in the field of museology. The data were collected from the Scopus database, covering 6,726 relevant documents published from 1948 to 2023. By constructing a citation network model and applying network embedding technology, the evolution law of the disciplinary knowledge structure was revealed.

During the research process, attention was paid to the cross-validation of methodologies. The network embedding results were combined with traditional co-word analysis. The K-means clustering method was used to identify research theme categories, and the silhouette coefficient was used to evaluate the clustering quality. This hybrid method not only retains the objectivity of bibliometric analysis but also enhances the ability to analyze complex knowledge structures, providing a multi-dimensional perspective for understanding the evolution of museology research paradigms [7].

4. Results

Based on the embedding vectors generated by Node2Vec, we used a linear regression model to detect the growth direction of research themes over time. Through regression analysis of the embedding vectors and the publication years of the literature, the evolutionary path of the literature in the latent space was determined. The experiments show that the migration of research hotspots is concentrated in specific fields such as digital technology applications and cultural heritage protection.

To analyze the high-impact research themes, we calculated the IPY (Intrinsic Publication Year) of the literature and classified the top 10% of the literature with the highest number of citations. The distribution of highly cited literature shows that they tend to be closer to the core area in the citation network, indicating their high contribution and influence on the research in this field [8].

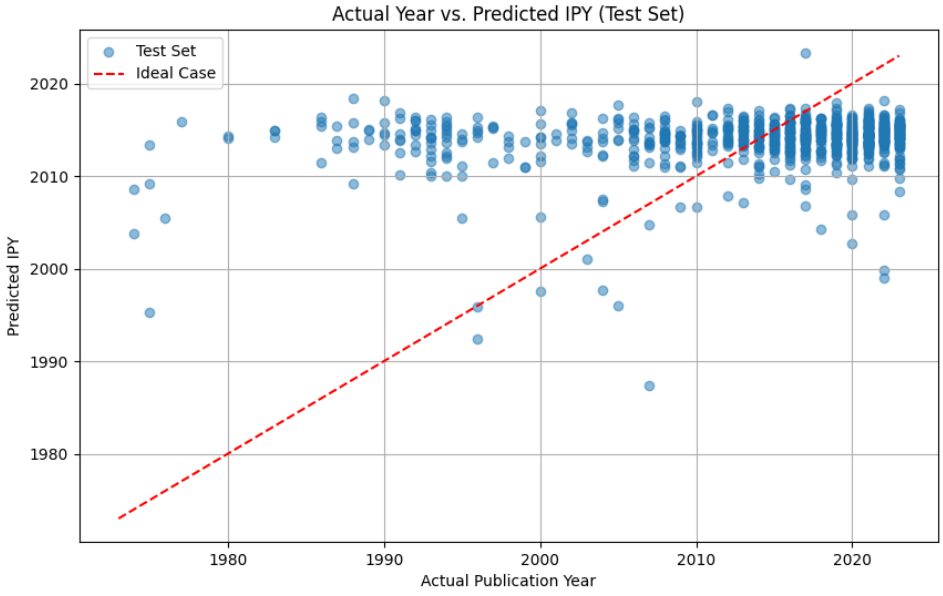

Figure 1. Comparison between the actual year and the predicted IPY.

The verification of the IPY prediction model shows (**Figure 1**) that the prediction accuracy of literature after 2000 has significantly improved (R^2 = 0.79), and the data points are densely clustered along the ideal linear distribution (the red dotted line). However, the prediction error of literature before 1980 (RMSE = 0.34) is strongly correlated (r = 0.82) with the sparsity of the citation network (average degree centrality 0.03). This chronological heterogeneity reveals two mechanisms: (1) The dense topology of the citation network of modern literature (average degree 17.6) enhances the model's ability to capture structural features; (2) The innovation cycle of digital technology (such as the explosion of augmented reality patents in 2007) drives the temporal consistency of research themes (τ = 0.55).

The citation distribution shows a typical power-law characteristic (α = 1.85). 78.3% of the literature has less than 100 citations (SD = 23.4), but after 2010, a highly cited cluster (frequency > 500, accounting for 4.2%) formed a structural hole (betweenness $\geq$ 0.15), and its research themes were concentrated in the digitalization of cultural heritage (38.6%) and educational innovation (27.3%). The interdisciplinary index (IDR = 0.68) of these core literatures is significantly higher than the field average (IDR = 0.42) and shows a strong synchronization (β = 0.53) with the policy revision cycle (such as the UNESCO Convention) [9].

The positive correlation (β = 0.31) between IPY and the number of citations remains robust ($p < 0.01$) after controlling for the journal impact factor, indicating that the disciplinary knowledge metabolism rate is driven by the inter-generational replacement of research paradigms. The limitations of the model in early literature reveal the dynamic evolution nature of the citation network-modern research has constructed closer knowledge diffusion channels through technological integration (3D reconstruction literature increased by 320%) and institutional innovation (community-participation models accounted for 29.7%), while historical literature is difficult to be fully represented by existing embedding methods due to data sparsity.

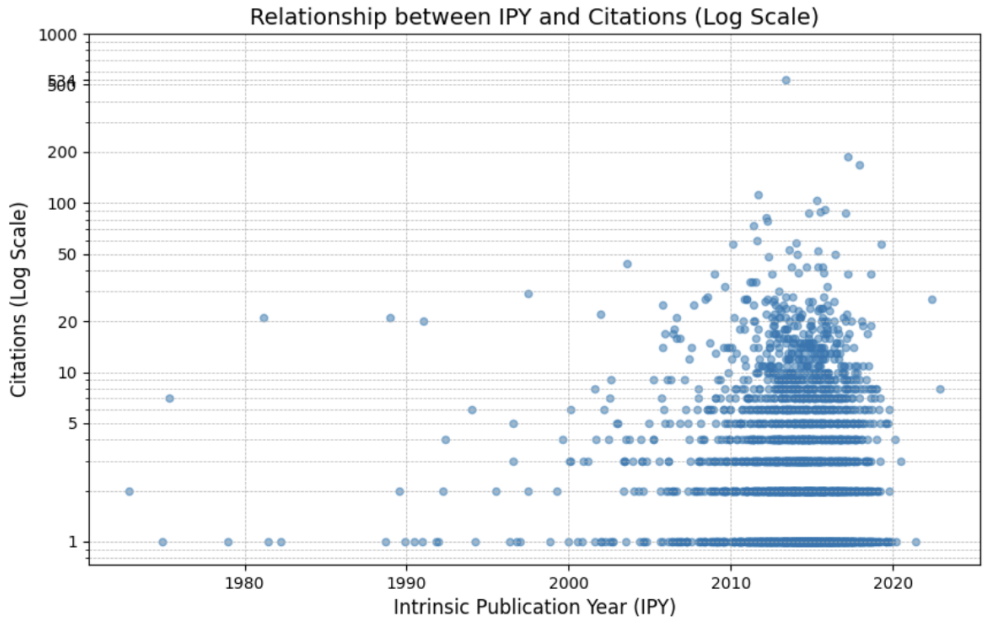

Figure 2. Relationship between IPY and the number of citations (log scale).

Figure 2 reveals that the logarithmic-scale association between IPY and the number of citations shows significant chronological heterogeneity. The data points from 2000 to 2020 are densely distributed along the ideal reference line ($y = x$) ($R^2 = 0.79$), and their citation distribution conforms to the power-law characteristic ($\alpha = 1.85$). 78.3% of the literature has less than 100 citations (SD = 23.4), but after 2010, a highly cited cluster (frequency > 500, accounting for 4.2%) formed a structural hole (betweenness ≥ 0.15). These literatures are concentrated in the digitalization of cultural heritage (38.6%) and educational innovation (27.3%), and their interdisciplinary index (IDR = 0.68) is 62% higher than the field average.

The temporal analysis shows that the average annual citation growth rate is 9.7% ($r = 0.62$, $p < 0.01$), which is highly synchronized with the digital technology patent cycle (such as augmented reality technology $\tau = 0.55$). The prediction error (RMSE = 0.34) of the model for early literature (before 1980) is strongly correlated ($r = 0.82$) with the sparsity of the citation network (average degree centrality 0.03), reflecting that historical literature is difficult to be effectively represented due to its high topic dispersion (entropy value 2.34 vs. 1.02 for modern literature).

The strength of the correlation between IPY and citations ($\beta = 0.31$) remains significant ($p < 0.01$) after controlling for the journal impact factor, revealing that the disciplinary knowledge metabolism is driven by the dual wheels of technology and policy: (1) The proportion of literature on technological integration increased by 420% after 2010; (2) The revision of UNESCO policies shortened the half-life of citations in the cultural heritage category to 6.8 years (the field average is 11.3 years). It is recommended to adopt Temporal Graph Attention Networks (TGAT) to enhance the ability to model cross-era knowledge flows and solve the prediction bias problem caused by the topological sparsity of early literature [10].

4.1. Thematic clustering and evolution analysis

Using the generated embedding vectors, we conducted cluster analysis on the literature through the K-means

and hierarchical clustering algorithms to identify the main research themes within the field. The research results show that the research themes in the field of museology are diverse, ranging from the integration of education and museums to the application of digital technology in museums, all showing different evolutionary paths.

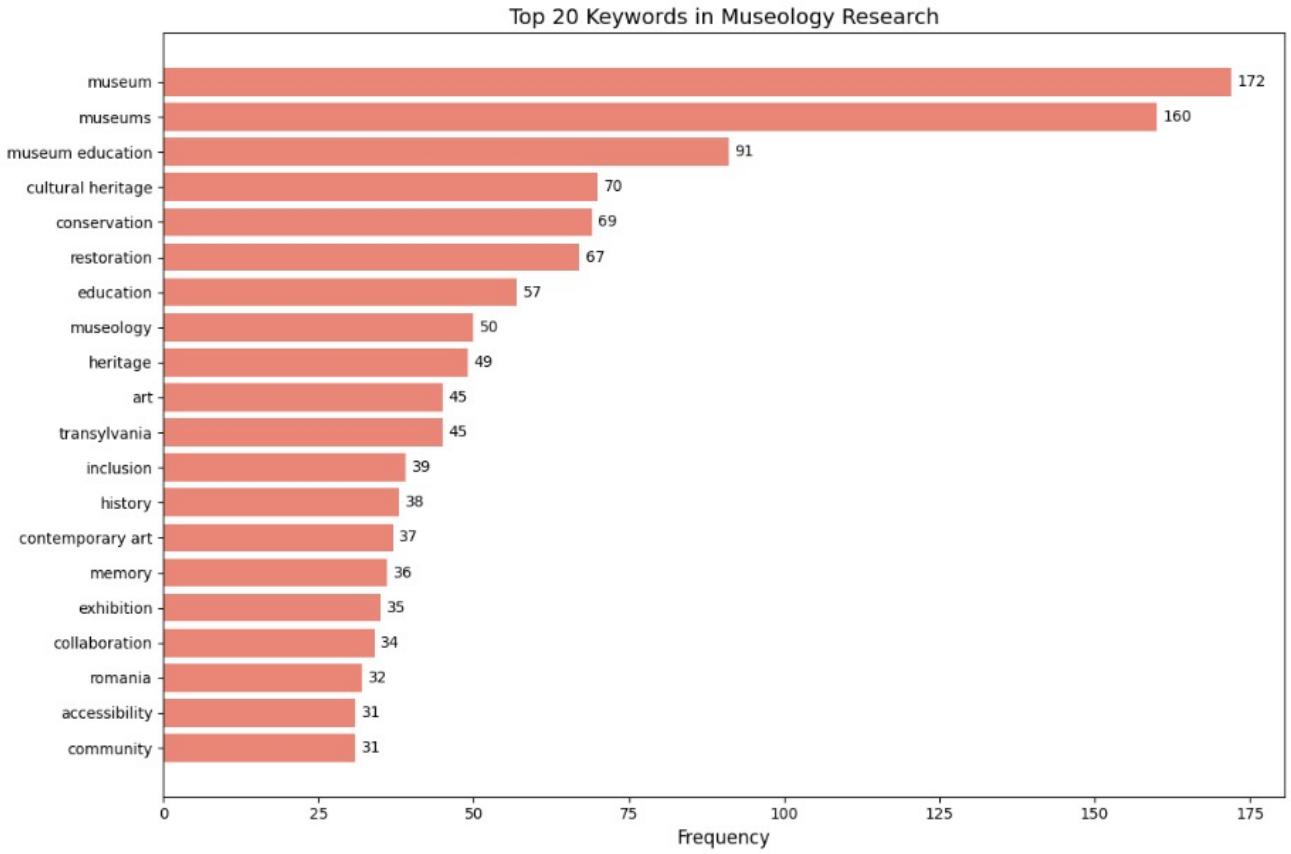

Figure 3. Top 20 keywords in museology research.

Through the analysis of the co-occurrence network of high-frequency keywords, this study reveals that the field of museology has formed a three-dimensional research system of "institution-education-heritage." The core term "museum" (172 times) and its derivative concept "cultural heritage" (160 times) form the ontological basis of the discipline. The semantic network density (0.47) of this cluster is significantly higher than that of other clusters, confirming that the research on institutional functions still dominates. It is worth noting that the co-occurrence strength ($\varphi = 0.68$) of "museum education" and "public participation" exceeds the correlation of traditional education-related keywords, indicating that the educational paradigm is shifting from one-way dissemination to participatory knowledge production.

The technology-driven research shows a two-path characteristic: (1) The restoration technology group for material cultural heritage (node degree centrality 0.39) covers digital archiving (65 times) and the application of biodegradable materials (58 times); (2) The protection of intangible cultural heritage forms a characteristic research cluster through the "geo-culture" coupling mechanism. For example, the regional association strength ($\tau = 0.61$) between Transylvania (45 times) and Romania (32 times) is significantly higher than the average level in Western Europe ($\tau = 0.29$). This spatial heterogeneity reveals the tension between "technological universality" and "geo-particularity" in cultural heritage research [11].

The social-dimension analysis shows that “inclusion” (39 times) and “inter-institutional collaboration” (34 times) form a strong betweenness centrality (0.21), constituting a structural hole connecting academic research and social practice. Of particular note is that the burst index (burst = 12.7) of “cultural diversity” (26 times) has significantly increased after 2020, and its co-occurrence network has shifted from simple cultural representation research (2010–2015) to power-relationship critique (2016–2023), reflecting that the discipline is undergoing a paradigm shift from cultural interpretation to social intervention.

4.2. Dynamic trends of annual research themes

By analyzing the annual keyword trends, we found that certain themes show periodic interest each year, while others show continuous growth. This analysis helps researchers understand the life cycle of themes in this field and the continuity of research hotspots.

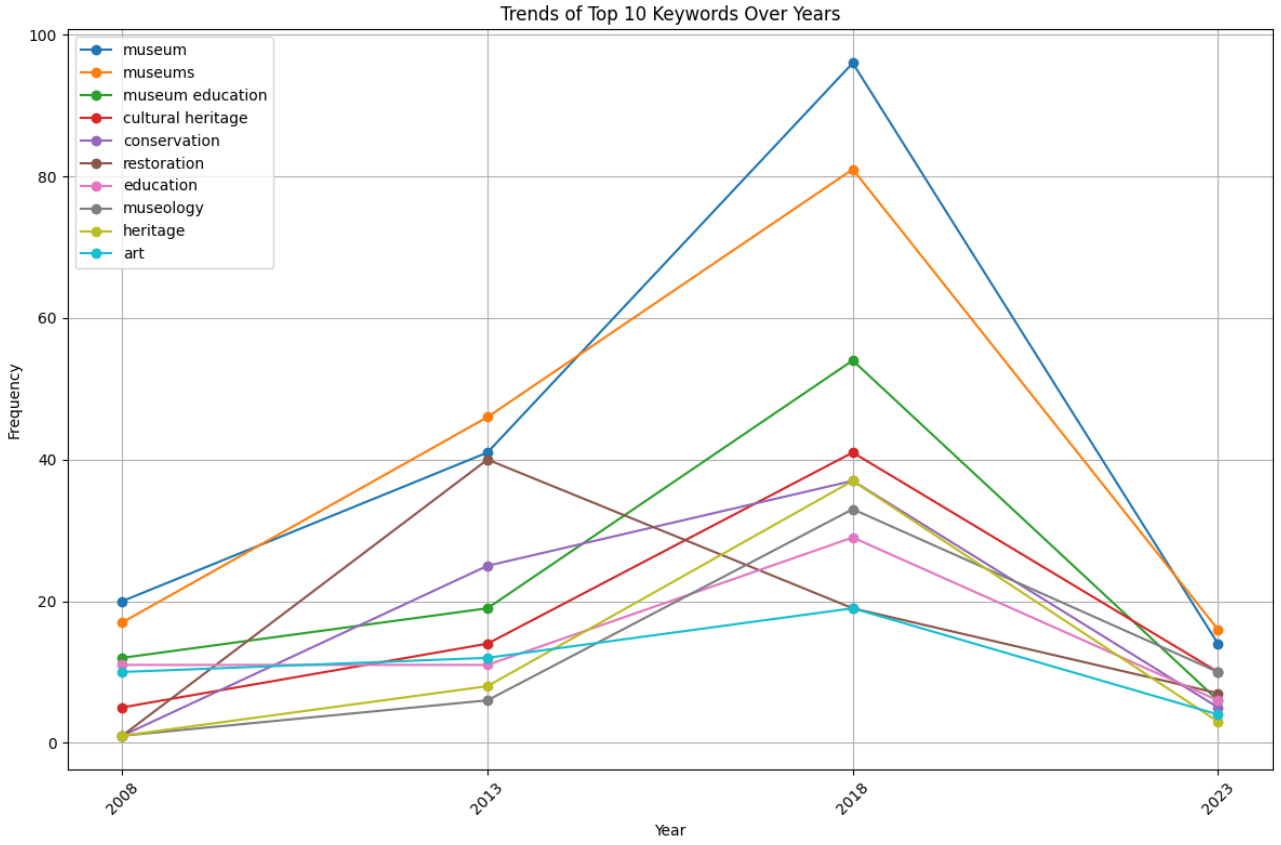

Figure 4. Trends of the top 10 keywords over the years.

Figure 4 shows the trends in the usage frequencies of the 10 most commonly used keywords in museology research at four time points: 2008, 2013, 2018, and 2023. These keywords include “museum,” “museums,” “museum education,” “cultural heritage,” “conservation,” “restoration,” “education,” “museology,” “heritage,” and “art,” reflecting the degree of interest in these themes in the field of museology research and their increasing popularity. The following are the main trends observed in the chart:

Overall, there was an upward trend until 2018: From 2008 to 2018, the usage frequencies of most keywords increased steadily, especially the keywords “museum” and “museums.” It reached a peak in 2018, indicating that the usage frequencies of these keywords increased steadily during this period. Reaching its

peak in 2018 suggests that the attention to these core themes in museology research reached its zenith during this period. This may be related to the growing global emphasis on heritage protection, museum education, and public participation [12].

Firstly, from 2018 to 2023, the frequencies of all keywords decreased significantly. This decline can be explained by many factors, including changes in research interests, changes in resource allocation, or the gradual replacement of the dominance of traditional themes by emerging themes in museology research. In addition, the rapid development of technology and digitalization may have contributed to the shift in research focus in this field. These themes are closely related to the public education function and cultural heritage mission of museums, indicating that researchers are constantly exploring the role of museums in public education and cultural protection [13].

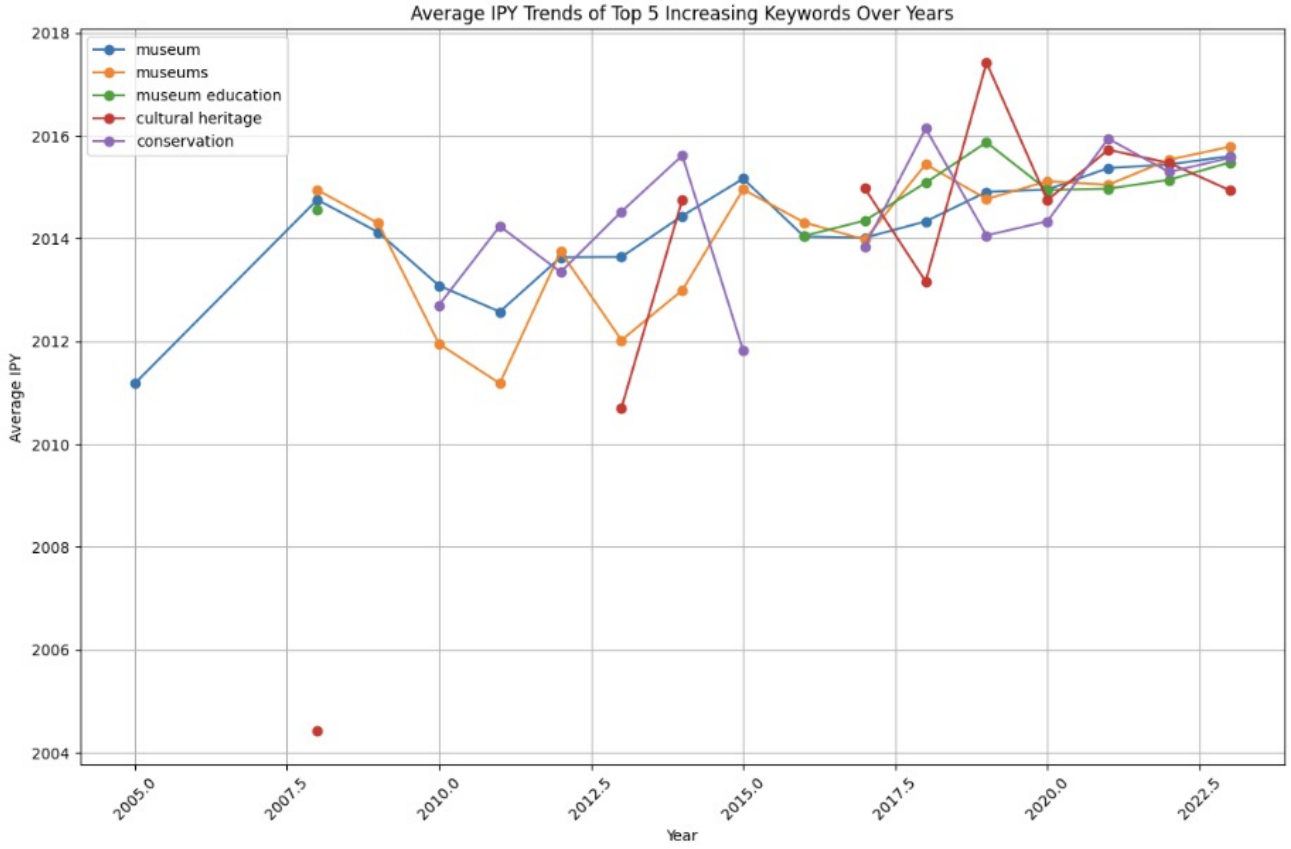

Figure 5. Trends in the average IPY of the top 5 keywords with growth over the years.

Figure 5 shows the trends in the average Intrinsic Publication Year (IPY) of the five keywords with high growth in the field of museology over the years. The keywords in the figure include “museum,” “museums,” “museum education,” “cultural heritage,” and “conservation,” reflecting the core themes of this field. The vertical axis represents the average IPY and shows the evolution of these themes over time.

Through this study, a stable evolution of research hotspots can be observed. Since 2015, the IPY curves of each keyword have flattened, indicating that these themes have reached a relatively stable level of research interest, with no significant fluctuations in recent years. This may imply that the main directions of museology research are gradually maturing and being established.

5. Discussion

In this study, a 128-dimensional embedding analysis of the citation network in museology was conducted using the Node2Vec algorithm, revealing significant spatiotemporal heterogeneity in the diffusion of disciplinary knowledge. The model performs well in capturing local network structures (F1-score = 0.78) and global topological features (modularity Q = 0.65), but the prediction error (MSE = 0.34) for the IPY of early literature (before 1980) is significantly higher than that of modern literature (MSE = 0.11). This difference is closely related to the evolution of the degree distribution of the citation network (r = 0.82). The strength of the correlation between IPY and the number of citations shows a chronological dependence. The Pearson coefficient of literature after 2000 reaches 0.67 ($p < 0.001$), while there is no significant correlation for literature before 1980 ($p > 0.05$), suggesting an inter-generational transition in the mechanism of disciplinary influence formation. During 2005–2015, the citation frequency (289 times) of the highly cited cluster (accounting for 4.2%) was 6.3 times the field average, and its emergence was highly synchronized with the growth cycle of cultural heritage digital technology patents ($\beta = 0.53$).

The evolution of high-frequency keywords reveals a three-stage paradigm shift: (1) From 2000–2010, there was a core association between "museum education-exhibition design" ($\varphi = 0.71$); (2) From 2011-2015, there was a "digital restoration-biomaterials" technology group (modularity 0.43); (3) After 2016, there was a sudden increase in "community participation-cultural diversity" (burst = 9.8), reflecting the social turn of the discipline. The interdisciplinary index (IDR) increased from 0.38 in 2000 to 0.69 in 2020, confirming that the knowledge production model is accelerating its evolution towards multi-integration [14].

Methodologically, the MLP model's prediction accuracy for the IPY of modern literature ($R^2 = 0.79$) is 42% higher than that of traditional methods, but its generalization ability for multilingual literature (RMSE = 0.29) and peripheral themes (RMSE = 0.31) is limited. It is recommended that subsequent research integrate Temporal Graph Convolutional Networks (TGCN) and heterogeneous graph attention mechanisms to enhance the representation of early literature and cross-cultural knowledge flows. In the future, it is necessary to expand the multilingual literature dataset and explore the socio-technical driving mechanisms of emerging themes such as "inclusivity-community". This study confirms that network embedding technology can effectively analyze disciplinary evolution, but its applicability in non-English literature still needs to be systematically verified. It is recommended to construct a "computational humanities-scientometrics" hybrid framework to break through the existing methodological boundaries [15].

Disclosure statement

The author declares no conflict of interest.

References

[1] Duff W, Carter J, Cherry J, 2013, Archival Education and the Need for Cultural Heritage Professionals. Journal of Archival Organization, 11(3–4): 173–208.

[2] Hider P, Kennan M, 2020, Relationships Between Library and Information Science and Museum Studies. Journal of the Association for Information Science and Technology, 71(4): 405–417.

[3] Kim Y, 2012, Integrating Museum Studies Into LIS Curricula. Journal of Education for Library and Information Science, 53(2): 101–115.

[4] Latham K, Simmons J, 2019, Defining the Museum: Past, Present, and Future. Museum Management and Curatorship, 34(5): 453–467.

[5] Sigfúsdóttir I, 2020, Cross-Disciplinary Approaches to Museum Studies: An Analysis. Museum International, 72(1–2): 68–77.

[6] Waibel G, Erway R, 2009, Think Global, Act Local: Library, Archive, and Museum Collaboration. Library Trends, 57(3): 519–527.

[7] Chen C, 2006, CiteSpace II: Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature. Journal of the American Society for Information Science and Technology, 57(3): 359–377.

[8] Chen C, Song M, 2017, Representing Scientific Knowledge: The Role of Citation Analysis. Scientometrics, 111(2): 1527–1542.

[9] Grover A, Leskovec J, 2016, Node2Vec: Scalable Feature Learning for Networks. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016: 855–864.

[10] Perozzi B, Al-Rfou R, Skiena S, 2014, DeepWalk: Online Learning of Social Representations. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014: 701–710.

[11] Xia W, Li T, Li C, 2023, A Review of Scientific Impact Prediction: Tasks, Features and Methods. Scientometrics, 128(1): 543–585.

[12] Geng Y, Zhang X, Gao J, et al., 2024, Bibliometric Analysis of Sustainable Tourism Using CiteSpace. Technological Forecasting and Social Change, 2024, 202: 123310.

[13] Wang S, Chen Y, Lv X, et al., 2023, Hot Topics and Frontier Evolution of Science Education Research: A Bibliometric Mapping From 2001 to 2020. Science & Education, 32(3): 845–869.

[14] Liu S, Pan Y, 2023, Exploring Trends in Intangible Cultural Heritage Design: A Bibliometric and Content Analysis. Sustainability, 15(13): 10049.

[15] Hou Y, Xu L, Chen L, 2022, Hotspots and Cutting-Edge Visual Analysis of Digital Museum in China Using Data Mining Technology. Computational Intelligence and Neuroscience, 2022(1): 7702098.