

# Big Data-Driven Analysis of Stylistic Evolution in Ancient Chinese Literature: A Dynamic Time-Space Coupling Model

Qingshan Ding

 <http://orcid.org/0009-0003-0354-1304>

Zhengzhou College of Finance and Economics, China

Received: September 23rd, 2025 | Accepted: February 9th, 2026

## ABSTRACT

This study pioneers a digital humanities approach to systematically examine the millennium-spanning evolution of Chinese literary style. By constructing a cross-generational corpus from the pre-Qin through the Ming-Qing periods and developing a multidimensional quantitative index system, it integrates linguistic feature extraction with social network analysis to overcome the limitations of traditional chronological methods. Methodological innovations include dynamic time-window processing and an enhanced latent Dirichlet allocation model, which reveal key transition patterns between poetry and prose. Findings identify significant “style interpenetration” during the Tang-Song transformation and underscore the previously underestimated normative influence of imperial examination reforms, whose effects surpassed those of dynastic transitions. These insights challenge conventional literary historiography while providing new perspectives on the dynamic interplay between literary production and social structures.

## KEYWORDS

Big Data Analysis, Ancient Literary Works, Style Evolution, Stylistic Infiltration

## INTRODUCTION

Research on the stylistic evolution of ancient Chinese literature has long faced the dual challenges of historical depth and spatial scope (Ming, 2024). Traditional textual scholarship relies on close reading and empirical induction (Kemppanen, 2024). Although it has achieved significant progress in textual dating and genre analysis, it remains difficult to systematically reveal patterns of evolution across millennia. In recent years, innovations in digital technology have provided new perspectives for literary research (Cheng et al., 2023). Western academic circles have applied statistical methods to classical text analysis since the 1950s (Matricciani, 2025). For example, Gao Benhan's quantitative study of *A Dream of Red Mansions* initiated interdisciplinary exploration (Yuan, 2022). With the maturation of latent Dirichlet allocation (LDA) topic modeling, social network analysis, and related technologies, researchers can capture period-specific stylistic features through the analysis of “content-free words,” confirming that literary style follows quantitative patterns of intergenerational inheritance and regional diffusion (Kim, 2022).

However, gaps remain in the existing research. First, the quantitative system of stylistic elements remains underdeveloped, as most studies focus on surface features such as vocabulary and sentence

DOI: 10.4018/IJICTE.402720

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

structure while overlooking deeper patterns, including stylistic infiltration and rhetorical evolution (AlKhuzaey et al., 2024). Second, long-term continuous analyses are largely confined to a single genre and lack a multidimensional perspective for cross-stylistic comparison. More importantly, traditional quantitative methods struggle to capture the nonlinear characteristics of stylistic evolution and offer limited explanatory power for complex issues such as the response mechanisms to major historical events and the diffusion pathways of regional schools (Swargiary, 2024).

This study proposes a methodological framework that combines a dynamic time window with an improved topic model, aiming to overcome three limitations of traditional research. First, a multidimensional stylistic index system is constructed that incorporates prosodic features, narrative patterns, and other latent elements into quantitative analysis. Second, a time–space coupling analytical model is established to simultaneously track the temporal evolution and spatial diffusion of stylistic elements. Third, a lag coefficient of literary response is designed to reveal the dynamic relationship between social change and stylistic transformation. This innovation can verify classical theories such as “the theory of reform in Tang and Song Dynasties” and “the theory of style replacing decline,” while potentially uncovering evolutionary patterns overlooked by traditional methods.

## LITERATURE REVIEW

Research on the evolution of ancient Chinese literary style has undergone three methodological shifts. Early scholarship, represented by Fu Xuancong's *Imperial Examination and Literature in the Tang Dynasty*, established a paradigm of chronological analysis from the perspective of social history by examining the relationship between writers' lives and institutional changes (Fu, 2002). In the 21st century, Li Huiyi's team pioneered the introduction of the LDA topic model into classical text analysis and revealed the intergenerational transmission patterns of image communities in Song *Ci* through the statistical analysis of “words without content,” marking the formal application of quantitative methods in stylistic research (Li, 2015). Although this approach captures diachronic shifts in high-frequency vocabulary, it struggles to quantify deeper stylistic elements such as structural patterns and rhetorical strategies.

The breakthrough of digital humanities technology has provided new pathways for research (Novikova & Novikov, 2024). Wang Zhaopeng's team developed a mapping platform for the chronology of Tang and Song literature that realized spatiotemporal coupling analysis of literary events for the first time; however, a quantitative correlation model linking stylistic elements to spatial distribution has yet to be established (Wang & Jiang, 2020). Similarly, the Harvard University China Biographical Database has advanced the reconstruction of personal relationship networks but has not resolved the challenge of distinguishing stylistic imitation from innovation (Di, 2024).

Current scholarship faces a methodological dilemma. First, the quantitative index system remains limited (Song, 2024). Existing studies emphasize surface features such as vocabulary frequency and sentence length but lack effective tools to measure deeper patterns, including stylistic interpenetration, such as the influence of *lǚ* and *fú* on parallel prose, and rhetorical evolution, including intergenerational differences in allusion density (Chen et al., 2025). Second, the spatiotemporal analytical model is fragmented (AlDajani & Bocanegra, 2024). The spatial diffusion of stylistic elements still relies on qualitative description, and quantitative tools such as a “style propagation rate model” have not been developed. Third, a causal inference mechanism is lacking, as research on the effects of major historical events on literary style largely remains at the level of time-series-based empirical judgment (Lin et al., 2024).

In recent years, methodological innovation has demonstrated two major shifts (Kim & Bruhn, 2024). At the data level, cross-cultural corpora developed for literary research provide a strong foundation for comparative studies related to the content and context of *The Cambridge History of Chinese Literature* (Chang, 2010). At the algorithmic level, the improved sliding-window LDA model can identify abrupt nodes within the gradual evolution of *ci* style in the Southern Song through

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/article/big-data-driven-analysis-of-stylistic-evolution-in-ancient-chinese-literature/402720](http://www.igi-global.com/article/big-data-driven-analysis-of-stylistic-evolution-in-ancient-chinese-literature/402720)

## Related Content

---

### Inadequate Infrastructure and the Infusion of Technology into K-12 Education

Gregg Asher (2009). *Encyclopedia of Distance Learning, Second Edition* (pp. 1155-1156).

[www.irma-international.org/chapter/inadequate-infrastructure-infusion-technology-into/11891](http://www.irma-international.org/chapter/inadequate-infrastructure-infusion-technology-into/11891)

### Third-Level Digital Divide in English Teaching and Learning

Zhonggen Yu (2018). *International Journal of Information and Communication Technology Education* (pp. 68-80).

[www.irma-international.org/article/third-level-digital-divide-in-english-teaching-and-learning/205622](http://www.irma-international.org/article/third-level-digital-divide-in-english-teaching-and-learning/205622)

### Certain about Uncertainty: Strategies and Practices for Virtual Teamwork in Online Classrooms

Stella F. Shields, Gisela Gil-Eguiand Concetta M. Stewart (2004). *The Distance Education Evolution: Issues and Case Studies* (pp. 116-141).

[www.irma-international.org/chapter/certain-uncertainty-strategies-practices-virtual/30303](http://www.irma-international.org/chapter/certain-uncertainty-strategies-practices-virtual/30303)

### From R&D Project to Virtual Universities

Harald Haugenand Bodil Ask (2009). *Encyclopedia of Distance Learning, Second Edition* (pp. 1032-1039).

[www.irma-international.org/chapter/project-virtual-universities/11873](http://www.irma-international.org/chapter/project-virtual-universities/11873)

### Self and Peer E-Assessment: A Study on Software Usability

Rosalina Babo, Joana Rocha, Ricardo Fitas, Jarkko Suhonenand Markku Tukiainen (2021). *International Journal of Information and Communication Technology Education* (pp. 68-85).

[www.irma-international.org/article/self-and-peer-e-assessment/277379](http://www.irma-international.org/article/self-and-peer-e-assessment/277379)