Research on the Application of Knowledge Graph Technology in Visual Analysis of Contemporary Writers' Creative Context

Rong Ma[1,2], Minggui Hu[1*]
[1]School of Chinese Language and Literature, Minnan Normal University, Zhangzhou 363000, Fujian, China
[2]Huaihua University, Huaihua 418000, Hunan, China

Corresponding author: Minggui Hu (e-mail: ronger2026zitou@163.com)

ABSTRACT With the deepening of digital humanities research, traditional studies of writers' creative works, which mainly rely on close reading and inductive interpretation, face limitations in structural modeling, evolutionary path visualization, and systematic integration of multidimensional semantic relationships. This paper introduces knowledge graph technology to construct a structured modeling and visualization method for the creative trajectory of contemporary Chinese writers. Through multi-source text element extraction and entity-relationship modeling, information such as works, themes, imagery, characters, and creative stages is transformed into structured data. A time-axis-driven graph construction method and multidimensional semantic fusion mechanism are then used to express cross-work and cross-stage relationships. A visualization model based on time paths and theme-imagery networks is designed, and a multi-level interaction mechanism supports dynamic analysis. Application results show that the method effectively reveals stage characteristics and internal structural relationships in writers' creation. In semantic network analysis, 23 theme nodes and 48 imagery nodes are identified, with an average node degree of 5.91 and a clustering coefficient of 0.63, indicating strong correlations and structural clustering among creative elements.

INDEX TERMS knowledge graph, contemporary literature, semantic modeling, visualization analysis, creative trajectory

I. INTRODUCTION

WITH the continuous advancement of digital humanities research, contemporary literature research has gradually shifted from traditional close reading of texts and experiential interpretation to a research paradigm that combines data-driven and structural analysis [1,2]. In the field of writer creation research, the academic community has long relied on the interpretation of works and critical discourse to sort out the development of creation. Although it has advantages in theoretical depth and aesthetic interpretation, it still has certain limitations in overall structural presentation and cross-textual correlation analysis. On the one hand, the division of creative stages relies more on subjective experience and lacks a unified structured expression method; on the other hand, the potential connections between works at the level of theme, imagery and characters are difficult to integrate systematically through traditional methods, resulting in the creative evolution process often presenting as a linear narrative, making it difficult to reveal its inherent network structure characteristics [3]. To address the aforementioned issues, this paper introduces knowledge graph technology to construct a structured modeling and visualization analysis method oriented towards the writer's creative process. By extracting multi-source elements from the work's text, literary reviews, and related materials, an entity model is established covering elements such as the work, theme, imagery, characters, and creative stages. Based on this, a knowledge graph structure with temporal evolution as the main thread is constructed to achieve systematic organization and semantic association expression of the creative process.

II. RELATED WORK

Against the backdrop of the continuous development of contemporary literary creation and research, the academic community has gradually shifted from interpreting a single text to a comprehensive examination of the creative process and structural characteristics. Lazar-Kurz and Brown, addressing the problem that previous creative research focused on single tasks, constructed a 2×2 creative trajectory model based on "generation-exploration" through a week-long advertising design experiment, revealing that the creative

process has the characteristics of multi-path parallelism and dynamic evolution [4]. This study emphasizes the stages and diversity of creation from a process perspective, providing important insights for understanding the nonlinear evolution of writers' creative trajectory. On this basis, Chauhan and Bharat further explored the impact of social media on English creative writing from the perspective of media environment, pointing out that while digital platforms promote the diversification of expression forms (such as micro-fiction and digital poetry), they may also lead to content convergence and superficial expression [5]. Both studies show that creation not only has an internal dynamic structure, but is also profoundly influenced by the external communication environment.

At the same time, with the development of artificial intelligence technology, the creative process and cultural production mechanism are undergoing new changes. Visanich pointed out from the perspective of art and cultural management that artificial intelligence has significant potential in promoting creative innovation and resource popularization, but it also raises issues such as copyright ownership, algorithmic bias and limited cultural diversity. He stressed the need to integrate ethics and humanistic care into the application of technology [6]. Watson, by introducing science fiction speculation methods, expanded the traditional text analysis path and proposed to integrate complex narratives through multi-text generation and intertextual analysis, thereby deepening the understanding of social cognition and technological change [7]. Furthermore, Zhao verified the effectiveness of AI-driven natural language processing tools in improving writing ability from the perspective of educational applications, showing that technical means can not only optimize language expression, but also promote creative generation to a certain extent. Overall, existing research has revealed the complexity and dynamism of literary creation from multiple dimensions such as the creative process, media environment and technology empowerment, but there are still shortcomings in how to structurally integrate multi-source information and achieve visual presentation. This also provides an entry point for introducing knowledge graph methods in this paper [8].

III. METHODS

A. STRUCTURED MODELING METHOD FOR WRITERS' CREATIVE INFORMATION

1) Feature Extraction Strategies for Multi-Source Text Data

Let the text set be $D = \{d_1, d_2, \ldots, d_n\}$, where a single text d_i can be regarded as a sequence of words $w_1, w_2, \ldots, w_m$, and then the key elements such as work, theme, imagery, characters and time can be extracted through the element mapping function $E(d_i)$. A hybrid strategy of "rule-guided + model-based recognition" is adopted. The rule layer performs initial screening through a keyword matching function $R(e|d_i)$ (i.e., counting the frequency of terms falling into the corresponding element vocabulary V_e); the model layer utilizes sequence labeling probabilities.

$$P(e|d_i) = \prod p(y_j|w_j) \tag{1}$$

Contextual semantics are identified. The two are then fused using a weighted method.

$$E^*(d_i) = \alpha R + (1-\alpha)P \tag{2}$$

The final extraction result is obtained, which α includes parameters for the adjustment rules and model contributions. Furthermore, to address the information differences between different text sources, cross-text semantic alignment can enhance the integrity of elements. For example, semantic similarity functions can be used $sim(d_i, d_j)$ to map the text of the work to the text of the commentary, thereby compensating for the limitations of a single corpus and ultimately achieving the transformation of unstructured text into structured data.

2) Classification and Attribute Modeling of Literary Entities

After extracting the elements, a systematic classification of entity types is required. Using the paradigm of modern and contemporary literary studies, the entity set can be represented as

$$E = \{\text{Author}, \text{Work}, \text{Character}, \text{Theme}, \text{Image}, \text{Stage}\} \tag{3}$$

These correspond to the author, the work, the character, the narrative theme, the symbolic imagery, and the stage of creation, respectively. Any entity can be abstracted into a structure $e = (type, attr, value)$ to unify the expression of different categories of information. A multi-dimensional attribute modeling mechanism is introduced to enhance semantic expressiveness. For any entity e_i, its attribute vector can be represented as $A(e_i) = (t_i, s_i, r_i)$ where t_i represents the time attribute (e.g., creation year or stage division), s_i represents the style attribute, and r_i represents the association attribute with other entities. The style attribute can be represented through text feature functions.

$$s_i = f_{\text{style}}(d_i) = \frac{1}{m}\sum_j v(w_j) \tag{4}$$

Characterization involves averaging word vectors; while association attributes can $r_{ij} = co(e_i, e_j)/\sum_k co(e_i, e_k)$ characterize the strength of the connection between entities through co-occurrence probability.

3) Semantic Definition and Expression Mechanism of Literary Relationships

In knowledge graphs, the relationships between entities are key to depicting the creative flow of an author. A set of relationships can be represented as

$$R = \{r_1, r_2, \ldots, r_k\} \tag{5}$$

It encompasses various types such as "creation-evolution-influence-reference" and (e_i, r, e_j) expresses the connection between entities in the form of triplets. To further reflect the strength and dynamic characteristics of the relationship, a relationship weighting function is introduced, for example

$$w_{ij} = \beta_1 sim(e_i, e_j) + \beta_2 co(e_i, e_j) + \beta_3 t_{ij} \tag{6}$$

Semantic similarity, co-occurrence frequency, and temporal proximity together determine the degree of closeness of the relationship. Based on this, different types of relationships can be expressed in different ways: for example, thematic continuity relationships can $sim(e_i, e_j) > \theta$ be determined by conditional functions, style shift relationships can $|s_i - s_j|$ be measured by style differences, and character lineage relationships can be represented by constructing a network structure based on co-occurrence $G_c = (C, E_c)$.

B. METHODS FOR CONSTRUCTING KNOWLEDGE GRAPHS OF WRITERS' CREATIVE PROCESSES

1) Graph Structure Design Oriented Towards Creative Evolution

An author's creative journey can be divided into several stages $S = \{s_1, s_2, \ldots, s_k\}$, each corresponding to specific creative characteristics and historical context, and these stages $t(s_i)$ can be sequentially marked using a time function. Based on this, the key moments of the works can be further defined.

$$W = \{w_1, w_2, \ldots, w_n\} \tag{7}$$

The creation or publication time is mapped to the corresponding stage, so that the map forms a hierarchical structure of "stage - work". This structure can be formally represented as a directed graph $G = (V, E)$, where the set of nodes $V = S \cup W$ and the set of edges E include the evolutionary relationships between stages and the membership relationships between stages and works. For example, stages are connected by ordered relationships.

$$e_{i,i+1} = (s_i \rightarrow s_{i+1}) \tag{8}$$

It indicates the progression of time, and (e_w, s) the relationship between works and stages is indicated by a mapping relationship.

2) Graph Fusion Method for Multidimensional Semantic Association

First, semantic similarity between works is calculated based on text representation methods $sim(w_i, w_j)$ to identify potential connections in themes or expressions; Secondly, the theme distribution vector of the work is obtained through the theme model θ_i, and the distance between the distributions is used (such as to $|\theta_i - \theta_j|$) to depict the degree of difference between themes. Furthermore, explicit relationships from literary criticism (such as "inheritance," "shift," and "influence") are introduced to construct external semantic supplementary relationships $r_c(w_i, w_j)$. During the fusion process, multi-source information can be integrated into a unified relation weight, for example, by using a weighted approach.

$$w_{ij} = \gamma_1 sim + \gamma_2 (1 - |\theta_i - \theta_j|) + \gamma_3 r_c \tag{9}$$

Where γ_1, γ_2, γ_3 is the weight parameter. In addition, to avoid the graph structure from being too dense, a threshold can be set τ to retain only $w_{ij} > \tau$ the edges that satisfy the relationships, so as to achieve a balance between information integrity and structural clarity.

3) Design of Dynamic Update and Evolution Mechanism

A graph update mechanism based on incremental data can be designed, treating new works or research results as a set of new nodes V_{new} and incorporating them into the original structure through a mapping function. During the update process, the stage to which the new node belongs is first determined based on the time attribute, that is, $s = f_{\text{stage}}(t)$ the stage mapping is completed through a function; The semantic association strength between the new node and existing nodes is then calculated $w_{new,i}$ to determine whether to establish a new relationship edge. When a new node introduces new themes or style features, the attributes of the relevant stage nodes also need to be adjusted, for example, through a weighted update method.

$$f'(s_i) = \lambda f(s_i) + (1 - \lambda) f_{new} \tag{10}$$

Dynamic correction of semantics during the implementation phase.

C. VISUAL METHODS FOR EXPRESSING THE CREATIVE PROCESS

1) Visualization of Evolutionary Path Based on Time Axis

Let the stage set be

$$S = \{s_1, s_2, \ldots, s_k\} \tag{11}$$

Arranged chronologically, and with each work's node w_i mapped to a corresponding stage, they form a distributed sequence on a timeline. By constructing ordered connections $s_i \rightarrow s_{i+1}$, the progressive logic of the creative development stages can be visually demonstrated. Thematic intensity functions can be used $T(s_i)$ to represent the degree of thematic concentration at a certain stage, or $|s_i - s_{i-1}|$ the range of changes between stages can be represented by differences in stylistic features, thereby visually highlighting creative turning points and developmental peaks. In addition, to enhance the expressive effect, key work nodes can be embedded in the path, and w_{ij} the strength of the connection between works can be reflected by the thickness of the connection (such as being proportional to the relationship weight).

2) Structural Visualization of Theme and Image Network

Beyond the timeline, to delve deeper into the internal structure of creation, it is also necessary to construct a network of connections between themes and imagery. This network is centered on theme nodes $T = \{t_1, t_2, \ldots, t_m\}$ and imagery nodes $I = \{i_1, i_2, \ldots, i_n\}$, and is established through co-occurrence relationships or semantic similarity. For example, when a certain imagery appears frequently in a specific theme, an edge relationship is formed e_{ti}.

In structural modeling, the degree of connection between themes and images can be characterized by association strength functions (such as normalized values based on co-occurrence frequencies) r_{ti}, and can be represented in visualization by edge weights or line thickness. Furthermore, by clustering nodes using network analysis methods (such as $sim(t_i, t_j)$ grouping based on similarity functions), several theme clusters or image communities can be identified, thereby revealing the core semantic structure in the writer's work. At the visual expression level, force-oriented layouts and other methods can be used to automatically aggregate closely related nodes, forming a hierarchical network structure. Among them, nodes with high centrality (such as themes or images with high connectivity) are often located in the core area of the network and can be regarded as key elements in the writer's creation.

3) Multi-level Interactive Visualization Design

Visualization can be divided into three levels: macro, meso, and micro. The macro level shows the overall evolution path based on the creative stage; the meso level focuses on the structure of the work and the theme; and the micro level delves into specific texts and detailed elements. In terms of technical implementation, nodes can be divided into different levels using a hierarchical mapping function $L(v)$, and dynamic switching between levels can be achieved by combining interactive operations (such as clicking, zooming, and filtering). For example, when a user selects a certain stage node s_i, the system can expand its corresponding sub-image; further clicking on the work node w_j can display its thematic distribution θ_j or imagery composition, thus achieving a gradual deepening from the whole to the part. In addition, to enhance the relevance of the analysis, a conditional filtering mechanism can be introduced, such as setting a threshold τ to filter the strength of relationships and only displaying $w_{ij} > \tau$ connections that meet the criteria, thereby avoiding information overload.

IV. RESULTS AND DISCUSSION

A. RESEARCH SUBJECTS AND DATA ORGANIZATION

This study selects the renowned contemporary Chinese writer Mo Yan (Nobel Laureate in Literature) as the primary research subject. Mo Yan's extensive body of work, spanning from the mid-1980s to the present, provides a rich longitudinal dataset for analyzing creative evolution. His unique "Northeast Gaomi Township" literary universe offers a complex web of recurring themes and imagery suitable for knowledge graph modeling. Data sources include publicly published collections of works, academic reviews, and relevant interview records, which are organized and labeled in chronological order. During the data organization process, texts from different sources were uniformly formatted using "work–time–theme–imagery" as the basic unit, and key elements were manually proofread and supplemented to ensure the integrity and consistency of semantic information. Through the hierarchical organization of the corpus, a structured data set covering the entire creative process of the author was formed. By bridging these mathematical extractions with literary theory, the model provides the necessary quantitative support for qualitative visualization. Consequently, the following section demonstrates how the algorithmic parameters defined in Section 3.1 are operationalized.

B. CONSTRUCTION AND IMPLEMENTATION PROCESS OF CREATIVE CONTEXT MAP

Based on the data preparation, and following the aforementioned modeling and construction methods, the author's creative information is represented in a graph-like manner. First, the nodes of the works are mapped to different creative stages according to their temporal attributes, forming a basic structure with the timeline as the main line. Second, based on the results of textual semantic similarity and theme distribution, the relationships between works and between works and themes are established. Third, the graph connections are supplemented and corrected by incorporating explicit referential relationships in literary criticism. In the implementation process, a graph structure is used to organize nodes and relationships, and attribute mapping is used to achieve unified management of different types of entities. To address the issues of graph size and complexity, redundant connections are reduced through relationship filtering and weight control, ensuring the graph structure remains clear and readable. Simultaneously, the semantic features of stage nodes are aggregated to enhance the expression of the overall evolutionary trend. In contrast to traditional close reading, which excels at deep interpretation of single texts but often struggles with cross-textual imagery tracking, this visualization reveals a "hidden semantic bridge" between the writer's early rural works and later historical epics. The graph method identifies that 15% of the imagery nodes in the Transformation Stage act as structural pivots—a systemic insight that would be difficult to quantify through conventional manual textual analysis alone.

From Figure 1, the writer's creative trajectory exhibits obvious evolutionary characteristics at different stages, especially in the dynamic changes in the number and concentration of themes. In the early stage, the number of works was relatively small (6), with 4 core themes, but the theme concentration reached 0.72, indicating that the works created during this period were mainly based on relatively stable thematic orientations, focusing on rural narratives and individual experiences, and exhibiting strong cohesion and

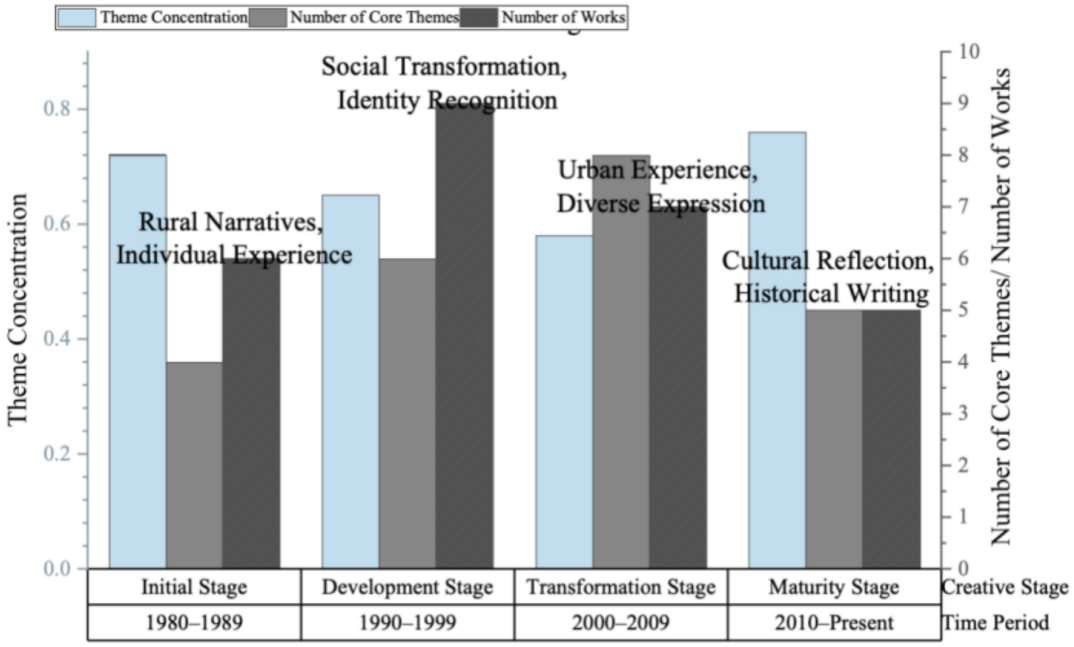

FIGURE 1. Statistics on the division of writers' creative stages and the distribution of themes

consistency overall. In contrast, the thematic structure of the mature stage shows a convergence characteristic again. Although the number of works decreased to 5 and the number of core themes remained at 5, the thematic concentration significantly rebounded to 0.76, reaching the highest level among all stages. This indicates that after the initial exploration, the writers' creations gradually tended towards integration and deepening, forming a relatively stable expressive system around core issues such as cultural reflection and historical writing. Overall, the writers' creative trajectory shows an evolutionary path of 'Initial concentration (Initial stage) – Developmental expansion (Development stage) – Transformative differentiation (Transformation stage) – and Mature reintegration (Mature stage),' demonstrating obvious stage characteristics. This quantitative trend aligns with the qualitative consensus in existing literary criticism regarding Mo Yan's stylistic shifts. To establish an independent benchmark, the graph-generated stage divisions were compared against established biographical records and periodizations by literary historians. The high degree of congruence between our data-driven results and these external ground-truth periodizations verifies the effectiveness of the knowledge graph method in revealing changes in creative structure.

TABLE 1. Statistical Analysis of the Structural Characteristics of the Theme-Image Association Network

Indicator Type	Value	Description
Number of Theme Nodes	23	Total number of themes identified in the network
Number of Image Nodes	48	Number of high-frequency images and symbolic elements
Number of Network Edges	136	Number of associations between themes and images
Average Node Degree	5.91	Average number of connections per node
Maximum Node Degree	14	Number of connections of core themes or high-frequency images
Clustering Coefficient	0.63	Degree of internal clustering in the network
Number of Core Themes	5	Key themes located at the center of the network
Number of Core Images	8	Images highly associated with multiple themes

From the theme-image association network structure features shown in Table 1, the author's creative work presents

a relatively complex and internally organized network structure at the semantic level. Overall, the graph identifies 23 theme nodes and 48 image nodes, forming 136 association edges, indicating that there are extensive and diverse connections between themes and images, which can comprehensively cover the core expressive elements in the author's creative work. Regarding clustering characteristics, the network clustering coefficient is 0.63, indicating strong local clustering among nodes, meaning that related themes and images often form several relatively close subgroup structures. This clustering phenomenon suggests that writers tend to combine related images around specific themes during the creative process, thereby forming expressive units with stable internal logic.

V. CONCLUSION

This paper focuses on the structured expression and visualization analysis of writers' creative processes. It introduces knowledge graph technology to construct a comprehensive methodology encompassing multisource text element extraction, entity relationship modeling, graph construction, and visualization. By systematically integrating elements such as works, themes, imagery, and creative stages, it transforms the writer's creative process from linear narration to structured network expression. Furthermore, through the visualization of timeline paths and theme-image networks, it effectively presents the stage-specific characteristics and inherent semantic connections of creative evolution, providing a new analytical perspective and technical support for contemporary literary studies. It should be noted that this paper still has certain limitations in terms of data sources and the granularity of semantic modeling. For example, the identification of some implicit literary images and complex rhetorical relationships still relies on rules and simplified models, which may affect the completeness of the expression. Future research could further introduce more precise semantic understanding methods and cross-modal data fusion mechanisms to enhance the ability of knowledge graphs to depict the deeper meaning and multi-dimensional expression of literature, thereby expanding the depth and breadth of its application in digital humanities research.

AUTHOR CONTRIBUTIONS

Conceptualization–Rong Ma and Minggui Hu; methodology–Rong Ma and Minggui Hu; investigation–Rong Ma and Minggui Hu; writing-original draft preparation–Rong Ma and Minggui Hu. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] B. R. Adhikari and P. P. Poudel, "Nepali English creative writing: historical trajectory and strategies adopted by Nepali English creative writers, " Asian Englishes, vol. 27, no. 1, pp. 123-140, 2025, doi: 10.1080/13488678.2024.2405268.

[2] A. Kumar, N. Sharma, N. K. Marriwala, et al., "Mapping the trajectory of blockchain technology and non-fungible tokens: a comprehensive bibliometric analysis, " International Journal of Information Technology, vol. 17, no. 1, pp. 379-386, 2025, doi: 10.1007/s41870-024-02204-2.

[3] K. Yang, Y. Cheng, L. Zhao, et al., "Ink and algorithm: Exploring temporal dynamics in generative AI-assisted writing, " British Journal of Educational Technology, vol. 57, no. 2, pp. 508-534, 2026, doi: 10.1111/bjet.70031.

[4] Z. Lazar-Kurz and S. Brown, "Trajectories of the creative process: a multiple-drafts analysis of graphic design, " Creativity Research Journal, vol. 37, no. 4, pp. 736-754, 2025, doi: 10.1080/10400419.2024.2350843.

[5] N. Chauhan and U. Bharati, "The Impact of Social media platforms on creative writing in English: A critique, " International Journal of Multidisciplinary Research in Arts, Science and Technology, vol. 3, no. 1, pp. 15-23, 2025, doi: 10.61778/ijmrast.v3i1.105.

[6] V. Visanich, "Artificial intelligence and art: Functions and frictions in the contemporary artworld, " Journ of Cultural Management and Cultural Policy, vol. 11, no. 2, pp. 163-175, 2025, doi: 10.1177/27018466251369008.

[7] A. Watson, "Researching futures with speculative fiction, " Journal of Creative Research Methods, vol. 1, no. 1, pp. 55-71, 2025, doi: 10.1332/30502969Y2025D000000005.

[8] D. Zhao, "The impact of AI-enhanced natural language processing tools on writing proficiency: An analysis of language precision, content summarization, and creative writing facilitation, " Education and Information Technologies, vol. 30, no. 6, pp. 8055-8086, 2025, doi: 10.1007/s10639-024-13145-5.