

Postprint of Research on Identification of Genetic Concepts in Chinese Painting Theory Based on Digital Technology

Authors: Niu Liang, Xu Jiajun, Xiang Wei

Date: 2025-08-14T14:22:39+00:00

Abstract

[Purpose/Significance] Genetic concepts serve as a crucial basis for determining cultural trajectories. Identifying the genetic concepts embedded in Chinese painting theory significantly contributes to understanding the aesthetic orientation of Chinese painting and its positioning within the broader context of Eastern art.

[Method/Process] This study first employs the BLAST method to construct a citation network of Chinese painting theory texts. Subsequently, based on an improved Kuhn gene measurement index, genetic concepts of painting theory are extracted. Finally, a co-word network is utilized to investigate the semantic relationships among these genetic concepts.

[Results/Conclusion] The research findings indicate that the “Six Principles” advocated by Xie He, the landscape techniques promoted by Wang Wei, and the literati painting philosophy championed by Su Shi constitute important genetic concepts in the development of Chinese painting. This study offers an exploratory approach to the quantitative identification of cultural genes.

Full Text

Research on the Identification of Genetic Concepts in Chinese Painting Theory Based on Digital Technology

Niu Liang¹, Xu Jiajun¹, Xiang Wei² ¹College of Economics and Management, China Jiliang University, Hangzhou 311222, China ²School of Art, Zhejiang International Studies University, Hangzhou 311222, China

Abstract

[Purpose/Significance] Genetic concepts serve as crucial determinants of cultural trajectories. Identifying the genetic concepts embedded in Chinese painting theory is essential for understanding the aesthetic orientation of Chinese painting and its position within the broader landscape of Eastern art. **[Method/Process]** This study first employs the BLAST method to construct a citation network of Chinese painting theoretical texts. It then extracts key genetic concepts from painting discourse using an improved Kuhn gene measurement metric, and finally examines the semantic relationships among these concepts through co-word network analysis. **[Result/Conclusion]** The findings reveal that Xie He' s "Six Principles," Wang Wei' s landscape painting techniques, and Su Shi' s literati painting philosophy constitute pivotal genetic concepts in the evolution of Chinese painting. This research provides an exploratory framework for the quantitative identification of cultural genes.

Keywords: BLAST method; Kuhn gene measurement; Chinese painting theory; genetic concepts

Dawkins posits that human cultural development is governed by genes that carry "genetic codes" determining civilization' s trajectory [?]. When applied to knowledge domains, this notion becomes "knowledge genes," referring to indivisible knowledge units within literature that express complete conceptual ideas and form the foundation of all knowledge management activities [?]. Liang Zhen-tao et al. introduced knowledge genes into medical informatics, supplementing and expanding previous literature-level studies from a micro perspective to enable more intuitive observation of medical knowledge gene transmission [?]. Sun Xiaoling et al. utilized a knowledge gene discovery algorithm to explore the application and dissemination of graphene as an informational carbon nanomaterial across scientific and technological fields, thereby investigating the relationship between science and technology [?]. However, gene identification methods have not yet been applied to painting theoretical concepts. This paper attempts such an application.

Chinese painting theory represents the accumulated exploration of China' s painting aesthetics and technical methods. Its embedded concepts are vital for understanding Chinese painting' s aesthetic system and artistic positioning. Yet these concepts have evolved under the influence of external environments and painting practices, manifesting in phenomena such as the Northern-Southern School division and the shift from "formal likeness" to "spiritual resonance." Among these transformations, the painting concepts that have been preserved, transmitted, and maintain dominant positions constitute genetic concepts. Chinese painting' s genetic concepts not only advance the construction of China' s painting system but also contribute significantly to Chinese painting' s positioning within Eastern art. Therefore, identifying these genetic concepts in Chinese painting theory and examining their role in Chinese painting history is crucial.

1. Theory and Methods

Since concepts are established and reinforced through writing, if painting theoretical texts from different dynasties exhibit citation relationships, concepts propagate through these networks. The longer the transmission chain, the more likely a concept qualifies as a genetic concept. Consequently, identifying genetic concepts in painting theory becomes a matter of recognizing them within textual citation relationships. Assuming concepts precipitate as keywords [?], the task transforms into keyword identification. Given that ancient writing rarely followed modern academic citation norms, making it difficult to distinguish cited content from an author's original writing, we adopt Vierthaler et al.'s recommendation: if texts share reused content, they are considered to have a citation relationship [?]. This paper thus establishes painting theory citations through textual reuse and subsequently identifies genetic concepts within these cited texts.

1.1 BLAST Method and Citation Network Construction Text reuse is calculated through similarity measurement, broadly categorized into sequence alignment methods and bag-of-words methods. Sequence alignment compares two texts character-by-character to find the longest identical continuous character sequences, commonly using the Smith-Waterman algorithm [?] and BLAST (Basic Local Alignment Search Tool) [?]. The Smith-Waterman algorithm exhibits higher computational complexity with greater time and space costs [?]. Bag-of-words methods ignore word order in compared sentences and instead rely on word occurrence probabilities, typically requiring topic models to calculate text topic distributions for similarity judgment [?]. However, topic models can only compare entire texts and cannot capture locally overlapping sentences. Moreover, their similarity calculation depends on word segmentation, making them unsuitable for ancient texts where segmentation is problematic. For detecting similarity among painting theoretical texts, BLAST offers a better solution for ancient text comparison by maintaining high precision while significantly reducing runtime and performing local sequence alignment without requiring word segmentation [?]. Therefore, this study employs BLAST to construct a citation network of Chinese painting theoretical texts.

BLAST uses a heuristic algorithm that operates as follows: First, it employs the N-gram method—a common vocabulary identification language model—to calculate terms from each painting theory as gene candidates. Due to N-gram's comprehensive word formation coverage, no keywords are lost, achieving 100% recall rate and ensuring comprehensive concept coverage. The input text is segmented into fragments (where N represents the number of characters per fragment), with each fragment assigned a fragment ID, its position in the text, and a text ID to build an indexable BLAST database while constructing k potential seed words. Next, these seed words are aligned against sequences in the pre-indexed BLAST database. Then, dynamic programming extends from high-scoring continuous positions in the seed words, using the Levenshtein al-

gorithm to calculate similarity scores. Extension terminates when scores fall below a threshold, outputting high-scoring sequences. Since multiple similar sequences may exist between texts, the total score aggregates all matching sequence scores, calculated as $\sum_i \text{score}_i$, where each sequence score equals the number of matched characters multiplied by its similarity. Finally, painting theories with scores above a certain threshold are pairwise linked to construct the painting theory citation network. BLAST similarity calculation and citation network construction are implemented using Python's Levenshtein and networkx packages, with similar sentences shown in Table 1.

1.2 Kuhn Gene Concept Measurement and Improvement Genetic concept identification primarily relies on citation networks. Kuhn et al. proposed measurement metrics based on citation networks: the Sticking coefficient and Sparking coefficient [?], considered effective for identifying genetic concepts [?]. The formulas are:

$$P_m = \frac{d_m^{\rightarrow m}}{d_m^{\rightarrow m} + \text{noise}} \times \frac{d_m^{\rightarrow m} + \text{noise}}{d_m^{\rightarrow m} + \text{noise}} \times f_m$$

where P_m represents the propagation score of genetic concept m , $d_m^{\rightarrow m}$ denotes the number of documents containing concept m that cite at least one document also containing m , $d_m^{\rightarrow m}$ represents all documents citing at least one document containing m , $d_m^{\leftrightarrow m}$ indicates documents containing m that do not cite documents containing m , and $d_m^{\leftrightarrow m}$ counts all documents not citing documents containing concept m . The noise element serves two purposes: preventing division-by-zero errors that would yield infinite scores, and penalizing genetic concepts with singular propagation to avoid selecting meaningless concepts. The term $\frac{d_m^{\rightarrow m}}{d_m^{\rightarrow m} + \text{noise}}$ constitutes the Sticking coefficient of concept m , representing the proportion of concept m appearing in texts that cite other texts containing m , indicating inheritance and transmission. The term $\frac{d_m^{\rightarrow m} + \text{noise}}{d_m^{\rightarrow m} + \text{noise}}$ forms the Sparking coefficient, representing the ratio of independently containing documents to non-citing documents, indicating fragmentation during diffusion. f_m is the frequency of concept m across all documents, and d_m is the number of documents containing m , while d_{all} is the total number of documents.

Introducing f_m effectively controls the influence of low-frequency genetic concepts. M_m represents the Kuhn value, indicating the importance of selected genetic concepts.

However, Kuhn's algorithm only considers hierarchical relationships and citation frequencies in citation networks, neglecting temporal dimensions. In reality, earlier concepts exert greater formative influence on later culture, making the discovery of early concepts crucial. Therefore, this study improves Kuhn's measurement by maintaining P_m and f_m while incorporating temporal sequence through reward factors. Using the function $\text{contain}(m \rightarrow t)$ to determine whether genetic concept m exists in time period t , concepts appearing across

more time periods are more important, reflecting continuity (denoted as cour1). Concepts appearing earlier are also more important, reflecting priority (denoted as cour2). The formulas are:

$$\text{cour1} = \sum_{t=1}^T \text{contain}(m \rightarrow t)$$

$$\text{cour2} = \frac{\text{contain}(0)}{\text{contain}(T - \text{begin_t})}$$

where T represents all time periods, t is each specific period, and begin_t is the first period when gene m appears. M'_m is the Kuhn value calculated using the improved metric.

To better understand the improved Kuhn measurement, this paper provides a test case using a citation network with 10 nodes [n1, n2, n3, n4, n5, n6, n7, n8, n9, n10] across periods T1, T2, T3, T4, arranged chronologically. Nodes are distributed across periods with sequential order, and their citation network is shown in Figure 1 [Figure 1: see original paper]. The test genetic concepts are [M1, M2, M3, M4], with citation details in Table 2 and propagation network in Figure 2 [Figure 2: see original paper]. With noise coefficient set to 1, calculation results for both original and improved Kuhn algorithms are shown in Table 3, where rank represents original Kuhn value ranking and rank' represents improved algorithm ranking.

The original Kuhn algorithm yields ranking [M4, M2, M1, M3], while the improved algorithm produces [M1, M2, M3, M4]. Concept M4 first appears in T4 with Sticking coefficient 0.75, Sparking coefficient 0.25, propagation score $P_m = 3$, and frequency $f_m = 0.4$. Concept M1 first appears in T1 with Sticking coefficient 0.60, Sparking coefficient 1, propagation score $P_m = 0.6$, and frequency $f_m = 0.7$. The original Kuhn value for M4 is nearly triple that of M1. However, after introducing reward factors cour1 and cour2, high-frequency, early-appearing concept M1 receives a boost, elevating it from third to first place, aligning with its historical significance.

1.3 Genetic Concept Co-word Network Construction While Kuhn's gene measurement extracts important genetic concepts from literature, these concepts exist independently. Knowledge evolution typically results from multiple genetic concepts working together. Interrelated concepts mutually explain each other, reflecting domain knowledge priorities and evolutionary directions. Therefore, exploring relationships among genetic concepts becomes crucial for understanding knowledge evolution, with co-word network analysis being the standard method. Co-word analysis examines deep associations between words by using text paragraphs or sentences as window units, counting co-occurring vocabulary and linking them pairwise to establish relationships. For genetic concept co-word network construction, the process is as follows [?]:

1. Use genetic concepts identified by Kuhn' s algorithm as a dictionary
2. Segment all literature by window units (paragraphs or sentences)
3. Position the pointer at the start of a window unit
4. Extract dictionary-contained genetic concepts from the window unit using forward maximum matching
5. Link co-occurring genetic concepts pairwise in the window unit, accumulating association frequencies
6. Repeat steps 3-5 for the next window unit until all units are processed

After constructing the genetic concept co-word network, we extract a subnetwork of the top 20 nodes ranked by Kuhn value, with node size influenced by Kuhn value. Nodes in this subnetwork possess importance from Kuhn measurement while network topology reveals relational significance. Since the network is built on window units (paragraphs, sentences), it ensures semantic completeness of genetic concepts, yielding more accurate selections.

2. Empirical Analysis

2.1 Research Framework Based on the above theories and methods, this paper establishes the overall research framework for empirical analysis of Chinese painting theory genetic concepts, as shown in Figure 3 [Figure 3: see original paper]:

1. Construct a Chinese painting theory citation network using the BLAST method
2. Apply the improved Kuhn genetic identification algorithm to examine temporally-sensitive Chinese painting theory genetic concepts and select candidate genes
3. Use forward matching to traverse candidate genetic concepts in Chinese painting theory, establishing connections among co-occurring concepts in the same paragraph to build a co-word network, detecting semantic relationships and establishing Chinese painting theory' s conceptual system

2.2 Data Sources and Preprocessing This study' s data primarily derives from Yu Jianhua' s *Comprehensive Survey of Chinese Painting Theory Through the Ages*, which collects 453 representative painting theories from pre-Qin to Qing dynasty. This corpus is selected for several reasons [?]: (1) Comprehensive coverage, integrating painting principles, techniques, formulas, poetry, criticism, catalogs, and treatises, including historically significant colophons with profound technique discussions; (2) Chronological organization by dynasty, providing foundational material for genetic concept excavation; (3) Authenticity and purity, with forged works identified and painting historical texts excluded to maintain theoretical integrity.

First, the paper texts were vectorized, manually proofread, and saved as text files titled "author-dynasty-painting theory name." Then, BLAST was applied to calculate text reuse among Chinese painting theories, constructing the cita-

tion network shown in Figure 4 [Figure 4: see original paper], where node size represents total score and edge width indicates citation frequency.

The highest degree centrality score belongs to *Qing-Yun Ge-Nantian Tiba* with total score 11,032.0, followed by *Qing-Shen Zongqian-Jiezhou Xuehua Bian* (9,238), *Qing-Tang Yifen-Huaquan Xilan* (5,739.067), *Qing-Fang Xun-Shanjingju Hualun* (5,322.0), and *Qing-Da Chongguang-Huaquan* (5,237.067).

2.3 Measurement of Genetic Concepts in Chinese Painting Theory

Chinese painting theory genetic concepts precipitate in keywords. To comprehensively capture them, segmenting painting theoretical texts by dynasty is essential. Based on genetic concept characteristics, we use 2-gram (two-character words) and 4-gram (four-character words) for segmentation, then calculate genetic concepts using Kuhn's algorithm.

Painting theory quantities vary dramatically across dynasties, with Ming and Qing works approximately triple those of earlier periods. According to literature growth theory, publications in any field increase over time with accelerating growth until saturation [?]. Due to explosive growth during Ming-Qing periods and correspondingly more citations, keywords from these eras—such as “Jing Guan Dong Ju” (artists' names), “qinglv shanshui” (blue-green landscape), “yongbi yongmo” (brush and ink usage), and “ziran shengdong” (natural vitality)—rank higher than earlier dynasties' keywords (see Table 4 and Table 5).

Chinese painting theory genetic concepts require consideration of both priority and continuity—concepts appearing earlier and spanning more dynasties are more likely to be important genetic concepts. The improved Kuhn algorithm incorporates temporal dimensions, with t representing the dynasty when concept m first appears. The Tang dynasty serves as a pivotal boundary because painting entered a conscious period during Tang. Liang Qichao noted that “Chinese painting, broadly speaking, pre-mid-Tang represents one era, while post-Kaiyuan/Tianbao marks a new era, with the watershed at the early Kaiyuan period” [?]. Teng Gu considered Tang “the golden age of Chinese painting history, significant for landscape painting's independent development into China's mainstream and Buddhist painting's detachment from foreign influence to form Chinese style” [?]. Su Shi stated in *Postscript on Wu Daozi's Painting* that “the wise create, the capable transmit. Painting did not develop through one person alone. For scholars in learning and craftsmen in technique, from the Three Dynasties through Han to Tang, all was complete. Thus poetry reached its peak with Du Fu, literature with Han Yu, calligraphy with Yan Zhenqing, and painting with Wu Daozi. All changes from ancient to modern, all capabilities under heaven, are fulfilled” [?]. Following these perspectives, we divide painting theory periods into pre-Tang, Tang, Five Dynasties, Song, Yuan, Ming, and Qing. Improved Kuhn algorithm results are shown in Table 6 and Table 7.

Two-character concepts from pre-Tang include “yongbi” (brush usage), “shanshui” (landscape), “ziran” (nature), and “qiyun” (spirit resonance), all occupying

crucial positions. These concepts persist in later dynasties with both continuity coefficient *cour1* and temporal coefficient *cour2* reaching or approaching 1, demonstrating their genetic attributes.

Four-character pre-Tang concepts include “*qiyun shengdong*” (spirit resonance and vitality), “*gufa yongbi*” (bone method brush usage), and “*jingying weizhi*” (composition arrangement). Tang concepts include “*yuanqi linli*” (pervading vitality), “*fan hua shanshui*” (all landscape painting), “*Gu Lu Zhang Wu*” (artist names), and “*yuanren wumu*” (distant figures have no eyes). Song concepts like “*pingdan tianzhen*” (plain and natural) and “*shanshui renwu*” (landscape and figures) permeate ancient Chinese painting theory. These genetic concepts, to some extent, determine Chinese painting’s direction. Comparing original and improved Kuhn algorithms for four-character concepts reveals that the improved version captures more genetic concepts, except for the Qing portion, which represents the final dynasty in our study where subsequent transmission cannot be demonstrated. Comparison results appear in Figure 5 [Figure 5: see original paper].

2.4 Construction of Genetic Concept Co-word Network After extracting core genetic concepts using the improved Kuhn algorithm, we reveal relationships among them through co-word network analysis. Using painting theory paragraphs as window units and two-character and four-character genetic concepts as dictionaries, we traverse all paragraphs via maximum matching, linking co-occurring concepts pairwise to build co-word networks. We extract subnetworks displaying the top 20 concepts ranked by Kuhn value, with node size determined by Kuhn ranking (see Figures 6 [Figure 6: see original paper] and 7 [Figure 7: see original paper]).

In the two-character network, “*shanshui*” (landscape), “*bimo*” (brush and ink), “*renwu*” (figures), and “*guren*” (ancient people) occupy central positions, indicating these concepts rank highly in Kuhn’s algorithm and frequently connect with other keywords in network topology, demonstrating semantic importance. For instance, “*yongbi*” (brush usage) ranks first in Kuhn’s algorithm but lower in the co-word network, while “*shanshui*” (landscape) ranks second in Kuhn’s algorithm but becomes prominent in the network, playing a pivotal role in theoretical development. Chinese culture can be summarized through six concepts: *xushi* (void and solid), *yinyang*, *tiandi* (heaven and earth), *qili* (qi and principle), *xinyi* (mind and intention), and *shen* (spirit). The first three belong to cosmology, while the latter three signify subtle connections between human spirit and the cosmos [?]. Landscape painting makes these connections visible and subject to contemplation, successfully shifting from external representation of nature’s grandeur to internal resonance with nature [?]. Rooted in Chinese cultural soil, the landscape concept remains constant as a painting subject even as styles evolve. Regarding landscape transformation, Zhang Yanyuan noted “landscape change began with Wu (Daozi) and was completed by the two Lis (Sixun and Zhao dao)” [?]. Guo Ruoxu stated: “In landscape painting, only Yingqiu Li

Cheng, Chang' an Guan Tong, and Huayuan Fan Kuan achieved divine wisdom and superior talent, establishing standards for a hundred generations. Though works by Wang Wei, Li Sixun, and Jing Hao survive, how can they compare?" [?]. Wang Shizhen observed: "Landscape painting changed with the two Lis, changed again with Jing, Guan, Dong, and Ju, again with Li Cheng and Fan Kuan, again with Liu, Li, Ma, and Xia, and again with Dachi and Huanghe" [?]. These three observations on "shanshui" come from painting historians of the 9th, 11th, and 16th centuries, respectively—sustained interest rarely shown for other genres like figures or flowers-and-birds. Shou-chien Shih notes that despite these historians' differing perspectives and concerns, "their narratives on landscape history all express a high degree of interest" [?], reflecting landscape concept's sustained vitality and exploratory enthusiasm, which explains its status as a Chinese painting genetic concept.

In the evolution from single to multi-character terms, four-character expressions provide increasingly specific guidance and system construction for painting. The top 20 four-character genetic concepts in Figure 7 [Figure 7: see original paper] show "gufa yongbi," "qiyun shengdong," "jingying weizhi," and "suilei fucai" frequently appearing and occupying central positions—these being key components of Xie He's Six Principles. Xie He stated: "Painting criticism assesses painting quality. All paintings clarify admonitions and record rise and decline; through a thousand years of solitude, opening the scroll provides reflection. Though painting has six principles, few can master them all. Since ancient times to present, each excels in one aspect. What are the six principles? First, qiyun shengdong; second, gufa yongbi; third, yingwu xiangxing; fourth, suilei fucai; fifth, jingying weizhi; sixth, chuanyi moxie" [?]. In the co-word network, "yingwu xiangxing" and "chuanyi moxie" do not occupy prominent positions because Chinese painting shifted from "formal likeness" to "spiritual resonance," with painting's subject and advocates transitioning from craftsmen to literati, making literati painting increasingly dominant. Dong Qichang's Northern-Southern School division resulted from literati painting development. For literati painters, as Su Shi's teacher Ouyang Xiu stated, strategies for handling space, height, and distance belonged to "craftsmen's art." Literati painters were not concerned with landscape's physical reality [?]. Therefore, "yingwu xiangxing" and "chuanyi moxie," expressing realistic landscapes, became less important. As Ni Zan said: "My bamboo merely writes the leisurely spirit in my heart; why would I compare its likeness, leaf density, or branch angle? After long application, others may see it as hemp or reeds, but I cannot forcefully argue it's bamboo. What can be done with such viewers? I only wonder what they see in it" [?]. Meanwhile, "gufa yongbi," "qiyun shengdong," "jingying weizhi," and "suilei fucai" remained as universal requirements for both literati and academic painting, thus preserved as Chinese painting principles. "Yuanshui wubo" (distant water has no waves) and "yuanren wumu" (distant figures have no eyes) originate from Wang Wei's *Landscape Treatise*: "In all landscape painting, intention precedes brush. A zhang mountain, chi tree, cun horse, fen figure—this is the method. Distant figures have no eyes, distant trees have no branches, distant mountains have no

rocks appearing like eyebrows, distant water has no waves reaching cloud height—this is the formula”[?]. Later, Jing Hao’s *Landscape Fu* also mentions: “Distant figures have no eyes, distant trees have no branches, distant mountains have no texture, reaching cloud height, distant water has no waves, faintly seeming to exist—this is the pattern”[?]. Wang Wei and Jing Hao specified landscape painting techniques, constituting Chinese painting technical instruction. “Yu ertong lin” (neighboring children), “si jian yu er” (seemingly seeing children), and “xingsi jian yu” (formal likeness seen by children), though lacking independent meaning as four-character phrases, preserve their context from Su Shi’s *Two Poems on Yanling Wang’s Painted Branches*: “Judging painting by formal likeness is neighboring children. Writing poetry must be this poem, certainly not a poet. Poetry and painting share the same principle, with nature and freshness. Bian Luan’s sparrows are lifelike, Zhao Chang’s flowers convey spirit. How can these two scrolls be so sparse yet refined? Who says one red dot cannot convey boundless spring? Thin bamboo like a recluse, secluded flowers like a maiden. Sparrows on branches, rain shaking among flowers. Double plumes about to rise, numerous leaves lifting themselves. Piteable flower-gathering bees store clear honey. If one possesses heaven’s skill, spring colors enter bold paper. Knowing you can write poetry, I send sound seeking wonderful words”[?]. Su Shi’s poems discuss painting’s formal versus spiritual likeness, emphasizing spirit-resonant works and responding to both literati and academic painting directions, after which literati painting dominated the mainstream. The four-character co-word network thus reflects Chinese painting principles, literati painting direction, and landscape techniques—validating these as genetic concepts.

This study applies BLAST, Kuhn gene measurement, and co-word network analysis to Chinese painting theory genetic concepts, representing digital humanities practice in Chinese painting theory. Findings reveal that terms like “yongbi,” “shanshui,” “ziran,” “qiyun shengdong,” “gufa yongbi,” and “jingying weizhi” permeate Chinese painting theory as core concepts. Research conclusions include: (1) The improved Kuhn algorithm effectively captures temporally-sensitive genetic concepts, greatly aiding identification of historically significant cultural genes. (2) Co-word network analysis reveals semantic associations among genetic concepts, providing effective connections for discrete Kuhn-ranked genes and facilitating holistic understanding of painting. (3) Two-character genetic concepts like “shanshui,” “ziran,” and “qiyun” effectively connect Chinese cultural concepts of void-solid, yin-yang, heaven-earth, qi-principle, mind-intention, and spirit, expressing subtle connections between cosmic objects and human subjects and achieving visual expression of heaven-earth-humanity. (4) Four-character genetic concepts like “gufa yongbi,” “qiyun shengdong,” “jingying weizhi,” “suilei fucai” (Xie He), “yu ertong lin,” “si jian yu er,” “xingsi jian yu” (Su Shi), and “yuanshui wubo,” “yuanren wumu” (Wang Wei) reflect Chinese painting principles, literati painting direction, and landscape techniques. (5) The improved Kuhn algorithm has methodological generalizability for identifying genetic concepts across long historical periods. For any cultural corpus that is complete, temporally extensive, and contains

citation relationships, cultural gene identification is possible, applicable to philosophical concept evolution, Chinese calligraphy concept evolution, and Confucian-Buddhist-Daoist concept evolution across long historical documents.

This study has limitations: (1) Since ancient literature lacks modern citation norms, our citation network relies on textual sequence similarity, lacking direct citations among painting theories. (2) The vast disparity in painting theory quantities across periods creates data sparsity challenges; future research could weight different historical periods by document quantity to enhance reliability.

References

- [1] Dawkins R. *The selfish gene* [M]. New York: Oxford University Press, 1976: 360.
- [2] Wen Tingxiao, Luo Xianchun, Liu Xiaoying, et al. Review of knowledge unit research [J]. *Journal of Library Science in China*, 2011, 37(5): 75-86.
- [3] Liang Zhentao, Mao Jin, Cao Yujie, et al. Analysis of domain knowledge diffusion patterns based on knowledge gene cascade networks [J]. *Information Studies: Theory & Application*, 2020, 43(4): 40-46.
- [4] Sun Xiaoling, Ding Kun. Research on the relationship between science and technology based on knowledge gene discovery [J]. *Information Studies: Theory & Application*, 2017, 40(6): 23-26, 17.
- [5] Jin Guantao, Liu Qingfeng. The historical world hidden in keywords [J]. *Journal of the History of Ideas in East Asia*, 2011(1): 55-83.
- [6] Vierthaler P, Gelein M. A BLAST-based, language-agnostic text reuse algorithm with a MARKUS implementation and sequence alignment optimized for large Chinese corpora [J]. *Journal of Cultural Analytics*, March 18, 2019: 1-19.
- [7] Funk K, Mullen A. The spine of American law: Digital text analysis and U.S. legal practice [J]. *The American Historical Review*, 2018, 123(1): 132-164.
- [8] Stephen F Altschul, et al. Basic local alignment search tool [J]. *Journal of Molecular Biology*, 1990, 215(3): 403-410.
- [9] Linder F, Desmarais B, Burgess M, et al. Text as policy: Measuring policy similarity through Bill Text Reuse [J]. *Policy Studies Journal*, 2018(2): 1-29.
- [10] Wang Zhenzhen, He Ming, Du Yongping. Text similarity calculation based on LDA topic model [J]. *Computer Science*, 2013, 40(12): 260-263.
- [11] Tobias Kuhn, Matjaž Perc, Dirk Helbing. Inheritance patterns in citation networks reveal scientific memes [J]. *Physical Review X*, 2014, 4(4): 041036.
- [12] Lu Wanhui, Tan Zongying. Research on domain theme evolution based on knowledge gene drift and recombination [J]. *Information Studies: Theory & Application*, 2019, 42(2): 101-107.
- [13] Niu Liang. The conceptual structure of the *Analects*: A digital humanities perspective [J]. *Library Tribune*, 2021, 41(2): 67-76.
- [14] Yu Jianhua. *Comprehensive survey of Chinese painting theory through the ages: Volume 1, Pre-Qin to Five Dynasties painting theory* [M]. Nanjing: Jiangsu Phoenix Art Publishing House, 2015: 1-5.
- [15] Li Jiuping, Yao Leye. Research on knowledge management literature growth models [J]. *Library Theory and Practice*, 2012(5): 36-39.
- [16] Liang Qichao. *Methods for Chinese historical research (plus two other works)* [M]. Shijiazhuang: Hebei Education Press, 2000: 350-351.
- [17] Teng Gu. *A brief history of Chinese art, Tang and Song painting history* [M]. Shanghai: Shanghai

Calligraphy and Painting Publishing House, 2016: 187. [18] Zavadskaya. *Foreign scholars on Chinese painting* [M]. Changsha: Hunan Art Publishing House, 1986: 64-80. [19] Hong Zaixin. *Selected essays on overseas Chinese painting research* [M]. Shanghai: Shanghai People's Art Publishing House, 1992: 250. [20] Yu Jianhua. *Comprehensive survey of Chinese painting theory through the ages: Volume 1, Pre-Qin to Five Dynasties painting theory* [M]. Nanjing: Jiangsu Phoenix Art Publishing House, 2015: 117-118. [21] Yu Jianhua. *Comprehensive survey of Chinese painting theory through the ages: Volume 2, Song dynasty painting theory* [M]. Nanjing: Jiangsu Phoenix Art Publishing House, 2015: 18-19. [22] Yu Jianhua. *Comprehensive survey of Chinese painting theory through the ages: Volume 4, Ming dynasty painting theory* [M]. Nanjing: Jiangsu Phoenix Art Publishing House, 2015: 89. [23] Shou-chien Shih. *A history of Chinese landscape painting and its audiences* [M]. Shanghai: Shanghai Calligraphy and Painting Publishing House, 2019: 1. [24] Yu Jianhua. *Comprehensive survey of Chinese painting theory through the ages: Volume 1, Pre-Qin to Five Dynasties painting theory* [M]. Nanjing: Jiangsu Phoenix Art Publishing House, 2015: 62. [25] James Cahill, translated by Li Yu. *Illustrated history of Chinese painting* [M]. Beijing: SDX Joint Publishing Company, 2014: 130. [26] Yu Jianhua. *Comprehensive survey of Chinese painting theory through the ages: Volume 3, Yuan dynasty painting theory* [M]. Nanjing: Jiangsu Phoenix Art Publishing House, 2015: 144-153. [27] Yu Jianhua. *Comprehensive survey of Chinese painting theory through the ages: Volume 1, Pre-Qin to Five Dynasties painting theory* [M]. Nanjing: Jiangsu Phoenix Art Publishing House, 2015: 155-156. [28] Yu Jianhua. *Comprehensive survey of Chinese painting theory through the ages: Volume 1, Pre-Qin to Five Dynasties painting theory* [M]. Nanjing: Jiangsu Phoenix Art Publishing House, 2015: 160-161. [29] Yu Jianhua. *Comprehensive survey of Chinese painting theory through the ages: Volume 2, Song dynasty painting theory* [M]. Nanjing: Jiangsu Phoenix Art Publishing House, 2015: 212-214.

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

Source: ChinaXiv – Machine translation. Verify with original.