ARTICLE

OPEN

Check for updates

How is the development of industrial digital-real integration progressing?—Evidence from China's cultural industries

Rui Huang[1] & Hongqian Wang[2 ✉]

Digital-real integration refers to embedding digital technologies across the entire industrial chain and lifecycle of the real economy, thereby forming a high-quality industrial ecosystem centered on digital technologies. Based on this concept, this study aimed to propose a methodology for scientifically assessing the progress of digital-real integration in the cultural industry. To achieve this, the study used Chinese cultural industry as an example by employing social network analysis, text classification and text similarity methods to construct a patent network, aiming to meet the research objectives by examining network characteristics, division of labor measurement and other dimensions. The findings indicated that the development of digital-real integration in China's cultural industry had surpassed the initial stage of technological integration and progressed to the intermediate stage of innovation chain and industrial chain integration. However, it had not yet reached the advanced stage of digital ecological operation. Therefore, customized and targeted policies should be implemented to address weak links, facilitating the establishment, strengthening and supplementation of industrial chains in the cultural industry.

[1] Faculty of Finance, Jilin University of Finance and Economics, Changchun 130117, China. [2] School of Public Policy and Management, University of Chinese Academy of Sciences, Beijing 100190, China. ✉email: wanghongqian24@mails.ucas.ac.cn

Introduction

Nowadays, new technologies, such as cloud computing, big data, and artificial intelligence, have reshaped the traditional culture industry. Digitization is igniting the new culture economy from online, and data traffic upgrading and field expansion have boosted the cultural industry's accelerated fission and cross-border reorganization. For instance, Steven Spielberg revealed to the foreign media Showbiz411 that the movie *Ready Player One* was expected to experiment with further creativity in meta-universe, i.e., creating a 3A-level MMO open meta-universe experience that matches the iconic IP. The Louvre's website already categorizes its collection by fields and themes, allowing visitors to "immerse themselves" in the classic collection without leaving their homes via an interactive mapping experience. Roblox has also been experimenting with "immersive advertising," a combination of digital and real that offers users a richer and more diverse experience and opens up new development opportunities for digital marketing. From this perspective, the cultural industry has been increasingly exhibiting new expressions of networked, intelligent, and interactive digital-real integration development. Therefore, scientific insight and evaluation of the progressing in the development of digital-real integration in the cultural industry nowadays played a significant role in optimizing digital transformation and development.

The so-called digital-real integration refers to integrating digital technology into the real economy with a full industrial chain and complete life cycle to form a high-quality industrial ecosystem with digital technology as the core (Meng et al. 2023). The development of digital-real integration in the cultural industry was supported by digital technology, driven by data elements, and carried on by the network. Through a complete digital transformation of all the processes, all the directions, and all the chains, such as cultural creative design and R&D, cultural product manufacturing and producing, and cultural product distribution and circulation, it realized the operational mode of improving the total factor productivity of the cultural industry (Passiante and Secundo, 2002). It would effectively connect the value chain, industrial chain, supply chain, innovation chain, and ecological chain of the cultural industry both online and offline (Xu et al. 2021; Tsolakis et al. 2023) and make the flow of information, technology, talents, capital and supplies in the cultural industry more efficient and smooth. At the same time, digital-real integration could facilitate the cultural industry's flexible production and digital cluster operation through a networked, flattened, and platform-based structure (Tracey and Clark, 2003).

Nevertheless, how should the progress of industrial digital-real integration be assessed? Urbinati et al. (2020) identified the stages and characteristics of industrial digital-real integration from the perspectives of technological aspects and integration pathways using a qualitative method (Elia et al. 2020). Wang Z, et al. (2024) broadly reflected the state of industrial digital-real integration by calculating comprehensive values using a coupling coordination degree model. Thus, the development of industrial digital-real integration is a gradual process that cannot be achieved overnight. While its ultimate goal is to achieve industrial ecological operations centered on digital technologies, more precise identification of the detailed progress and state of digital-real integration remains urgently needed. Therefore, unlike previous studies that focused on qualitative analyses or generalized quantitative approaches, this study proposes an innovative methodology for assessing the progress of digital-real integration in the cultural industry based on China's patent information through social network analysis, text classification and text similarity methods. Specifically, after an introduction, the next section reviewed the literature on the development of digital-real integration of

industries. The next section presented a theoretical analysis to define the characteristics of different stages in the development of digital-real integration. Next two sections cover the data, methodology and research findings, respectively, identifying the progress of digital-real integration in the cultural industry using China as an example. Last section offered a discussion and conclusion.

Related work

The integration of digital-real industries is a complex and gradual process that cannot be accomplished instantaneously. It is anticipated to evolve through distinct phases, including digitalization, intelligence and ecologization (Wang Y, et al., 2025). Given the unique characteristics and challenges associated with each stage, it is imperative to employ tailored methodologies to accurately evaluate the progress and depth of digital-real integration. Against this backdrop, this section provides a comprehensive review and synthesis of existing research on the evolution and measurement of digital-real integration, aiming to establish a robust foundation for further scholarly inquiry.

The evolution of the development of digital-real integration. Li (2021) pointed out that digital-real integration was a process of reconfiguring the business logic of the real economy, changing the organizational form, and innovating the value model through the integration and innovation of digital technology, application scenarios, and business models. Fu et al. (2022) found that the deep integration of the digital economy and real economy would follow the "new real enterprise-digital ecology-new real economy" route. The digital-real integration covered data elements, digital technology, the platform economy, and the deep integration of digital sharing with the real economy (Zhengzhen, 2021). At the technical level, it involved a highly connected linkage between industrial technology and digital technology. Enterprise entities would have a possibility of digital operation in the cloud through digital transformation (Borangiu et al. 2019). At the meso-level, the digital-real integration addressed the deep coupling of the industrial chain and the innovation chain. Namely, a distinctive main value chain has been formed within the enterprise cooperation network both online and offline (Tiwari, 2021). At the macroscopic level, the digital-real integration implied that the interactions and exchanges between the main entities of the industrial chain group would display significant ecological characteristics (de Koning et al. 2023). Xiao (2020) studied the main approaches for integrating the digital economy and the real economy at the enterprise, industry, and macroeconomic levels. He concluded that integrating the digital economy and real economy should be accelerated from innovation and introduction of digital technology, improvement and connection of digital infrastructure, cultivation of digital industrial ecosystem, and support and regulation of digital platform enterprises. Balaji and Roy (2017) identified that the industrial Internet, as the primary field of Internet of Things (IoT) application, shared a similar value creation model with the network architecture for developing digital-real integration. He suggested that digital platforms should be used to facilitate the digital transformation of enterprises and that digital ecosystems could be used to optimize industrial ecosystems, and this is the only way to consolidate the depth and breadth of digital-real integration continuously.

The effect measurement for the development of digital-real integration. Song, Yuxin (2020) used the coupling coordination degree model and gray correlation analysis to carefully measure the integration level of the digital economy and real economy.

The coupling coordination degree of digital-real integration development presented a trend of agglomeration followed by dispersion, while the regional multi-level differentiation feature was significant. Lee et al. (2015) measured the integration level of the digital economy and real economy based on the patent co-classification analysis method using patent application data and patent-industry matching information. Findings revealed that the digital-real integration network featured an obvious core-edge structure, and the digital-real integration still had a large space for industrial expansion. Xu et al. (2021) explored the spatiotemporal evolution characteristics and driving factors of integrating the digital economy and real economy through the entropy method, coordination degree, and spatial econometric model. The research found that the current level of digital-real integration was dominated by weak integration and the primary level of integration, and the digital-real integration was characterized by low agglomeration. Zhang and Wang (2022) examined the factual characteristics, regional differences, and dynamic evolution laws of digital-real integration using traditional and spatial Markov chains, Moran's I, and other methods. He argued that the current level of comprehensive development of the real economy lagged behind the level of development of the digital economy. However, with the popularization and development of digital technology, the polarization of the development of digital-real integration across regions would be alleviated.

Previous research has primarily relied on methodologies such as the coupling coordination degree to evaluate the integrated development level of digital-real integration from a holistic perspective. However, such approaches often neglect a nuanced examination and deconstruction of the distinctive characteristics associated with different stages of industrial digital-real integration, consequently limiting their ability to accurately assess the true progress of this phenomenon. In response to this limitation, the current study focuses on the cultural industry in China as a case study. By utilizing patent data, social network analysis, text classification and text similarity analysis, this research seeks to systematically analyze the developmental trajectory of digital-real integration within the cultural industry, with a particular emphasis on technological advancements, industrial chain dynamics and ecological chain division of labor. This study represents an innovative and pioneering contribution to the field, offering fresh insights into the evolving landscape of digital-real integration.
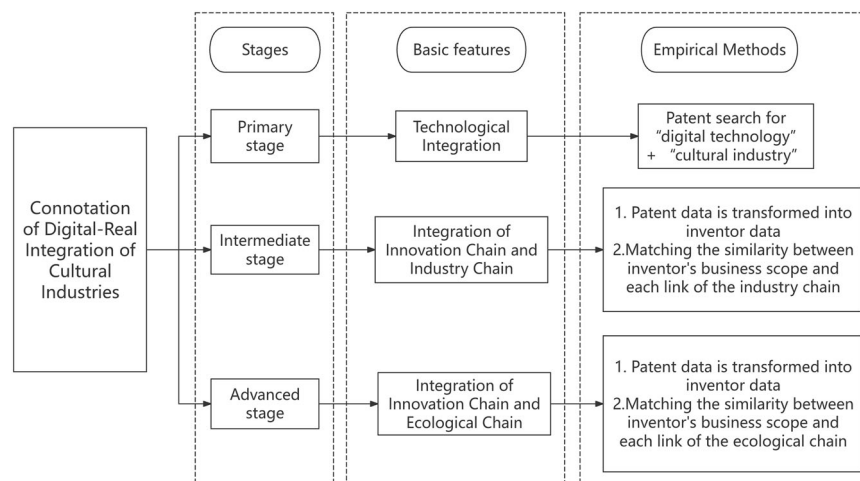
Theoretical framework

Theoretical foundation. Only by clarifying the characteristics of digital-real integration in the cultural industry at different stages can the progress of a specific country's cultural industry be accurately assessed. To address the limitations of previous studies that used the coupling coordination degree method for overly generalized observations of digital-real integration, this study begins with the concept of digital-real integration. The development is divided into primary, intermediate, and advanced stages. Using China as a case study, the analysis examined the progress and characteristics of digital-real integration in the cultural industry.

(1) *Analysis of characteristics at the primary stage of digital-real integration*: Previous studies focused predominantly on technological integration at the primary stage of digital-real integration in the cultural industry (Rosenberg N, 1963; Bygstad, Lanestedt, 2009). Specifically, researchers have highlighted the deep integration and application of digital technologies such as the Internet, big data, artificial intelligence, cloud computing, and the Internet of Things in the cultural industry. For instance, digital asset management and blockchain technologies have been widely applied to extract high-precision mural elements, reconstruct caves in 3D, and enable virtual tours of the Dunhuang caves. The creation of "Digital Dunhuang" ensured the proper preservation and long-term retention of invaluable cultural heritage in digital form. Consequently, the integration and transformation of traditional cultural technologies through digital technologies were identified as the most prominent features of the primary stage of digital-real integration (Breidbach and Maglio, 2015). The emergence and application of cultural technology patents characterized by both digitalization and platformization have been considered critical drivers of digital-real integration in the cultural industry (Alshammari et al., 2024).

(2) *Analysis of characteristics at the intermediate stage of digital-real integration*: At the intermediate stage of digital-real integration, more characteristics predominantly focused on restructuring cultural industry organizations by integrating innovation and industrial chains (Wang and Huang, 2024). At this stage, digital-real integration moved beyond simple technological integration. Digital technologies drove the digital transformation of the entire cultural industry chain. i.e., digital technologies not only accelerated the horizontal and vertical integration of production, distribution, and consumption in the cultural industry through highly integrated network platforms but also brought numerous participants into virtual spaces, expanding the boundaries of cultural industry clusters (Djamfa, 2024). For example, the National Public Culture Cloud was identified as a representative case. In the data storage and processing stage, cloud computing and big data technologies enabled the platform to classify and store vast cultural resources while implementing intelligent tagging for management. In the data exchange and management stage, blockchain and automated content review technologies ensured copyright protection and compliance for the platform's content. In the content production stage, platform technologies steered the vertical circulation of cultural industry production models. In the promotion and preservation stages, digital technologies expanded the boundaries of cultural dissemination and heritage through tools such as precise recommendation algorithms and virtual reality. Therefore, the integration of cultural industry chains and innovation chains was regarded as an inevitable trend (Firican, 2024).

(3) *Analysis of characteristics at the advanced stage of digital-real integration*: At the advanced stage of digital-real integration, more characteristics focused on reshaping organizational relationships in the cultural industry by integrating innovation and ecological chains (Lerro et al. 2022). At this stage, digital-real integration transitioned to an ecosystem integration model characterized by industrial interconnectivity, and its cultural industry organizations exhibited distinct ecological contexts (Bustamante, 2004; Calisi and Botta, 2022). The organizational relationships in digital-real integration included service groups (e.g., scenario creation, ecosystem interconnectivity, financial and tax services, and platform governance), key groups (e.g., cultural creativity, content production, and product marketing), innovation groups (e.g., 5G, AI, big data, and the metaverse), and resource groups (e.g., data collection, processing, trading, distribution, and presentation). For instance, the Palace Museum in Beijing launched the " The Digital Hall of *Da Gao-Xuan Dian* (大高玄殿数字馆, *Lit. The Digital Hall of the Great Shrine of Mystery Highness*)," the 3.0 version of its digital exhibitions, which was regarded

Fig. 1 Theoretical framework.

as a model of the cultural metaverse (Pais, 2019). This platform integrated online curation and offline exhibitions, enabling ultra-high-definition displays of digital artifacts, immersive interactive experiences, and comprehensive digital ecosystem planning for promoting cultural exhibitions of the Palace Museum. Thus, the advanced stage of digital-real integration was marked by ecosystem-driven operations in organizational relationships underpinned by digital technologies. At this stage, all aspects of the cultural industry ecosystem were fully transformed by digital technologies. The integration of the cultural industry ecosystem and innovation chains fostered the most professional and detailed modular division of labor within the industry (He and Wen, 2025). This integration was identified as the ultimate goal of digital-real integration in the cultural industry.

Analytical framework. In summary, to scientifically assess the progress of digital-real integration in the cultural industry, this study begins with the conceptual framework of digital-real integration and conducts empirical tests based on theoretical analyses of its characteristics at the primary, intermediate and advanced stages, as shown in Fig. 1. Specifically, at the primary stage, the study will use the keywords "digital technologies" plus "cultural industry" to search for relevant patent information, representing the technological integration of the cultural industry. At the intermediate stage, the study will extract information about patent inventors from the aforementioned patent data. It will then match the business scope of patent inventors with the various segments of the cultural industry chain based on similarity, showcasing the integration of innovation and industrial chains in the cultural industry. At the advanced stage, the study will similarly match the business scope of patent inventors with the various segments of the cultural industry ecosystem chain based on similarity, demonstrating the integration of innovation and ecosystem chains in the cultural industry.

Data and methods
Data source
Patent information. Since the development of digital-real integration in the cultural industry was based on digital technology, this paper selected relevant patent information to interpret, which fitted very well, and the specific data came from the patent database of China National Intellectual Property Administration. First, by referring to Guan and Liu (2016), following the approach

of Benson, Magee (2013) and Montecchi et al. (2013), we used specific keywords to search for relevant patents, this paper selected 18 keywords related to "digital-real integration" from documents and reports such as "Guiding Opinions on Promoting the Deep Integration of Culture with Science and Technology (S&T)" and "Top 10 Frontier Application Trends of Culture and S&T in 2022", and then constructed a search formula for "digital-real integration". Second, the keywords related to the cultural industry were selected from the Classification of Cultural and Other Related Industries in China (2018). Finally, link the keywords "digital-real integration" and "cultural industry" with the "AND" operator to construct the "cultural industry digital-real integration" keyword search formula.

For example, TS = (cultural industry keywords) AND (digital-real integration keywords) (see Table 1 for details). The use of these keywords helped us obtain technological patent data related to the integration of the digital and physical aspects of the cultural industry, and further identify relevant enterprises through patent citation relationships. At the same time, this paper took 2021, the release year of China's "14th Five-Year Plan" for the development of the cultural industry, as the base year, and as of June 2023, cross-searches were conducted by applying the above search formula. Finally, 12,672 patents and their citation relationships related to "digital-real integration of cultural industry" were obtained. By further summing up the patents to the individual level, a total of 2237 individuals (including enterprises, research institutions, universities, and so on) were covered, and a total of 3146 pairs of inter-individual citation relationships were obtained. Therefore, the nodes in the network of this paper represent various entities in the cultural industry, rather than individual patents. These entities include enterprises, research institutions, universities, etc., and each plays a different role in the patent network. The 2237 entities (nodes) represent various innovators in the field of cultural industry digital-physical integration. The connections between the nodes are determined by patent citation relationships. When a node's patent cites another node's patent, a directed edge is formed, indicating the technological transfer path from the citing patent to the cited patent. Specifically, if Entity A's patent cites Entity B's patent, a directed edge is established between A and B, and the number of citations determines the strength of the connection between A and B. Therefore, the connection is based on the citation relationship between patents. In this network, the connections between nodes not only reflect technological relationships but also reveal the interactions and collaborations of different entities in the technological innovation process (Karvonen and Kässi, 2013; Choi and Yoon, 2022).

Table 1 Related patent search equation.

	Keywords	Source
Digital-substantive integration	Generative AI, AIGC, Digital Twins, Virtual Productions, Meta Human, 3D Digital Modeling, Ultra HD Video, Intelligent Advertising, Game Engines, Real-time Audio, NFT, Meta-Universe, AR, VR, Extended Reality, Mixed Reality, Virtual Simulation, Blockchain	China's Ministry of Science and Technology, the Central Propaganda Department, the Central Internet Information Office, the Ministry of Finance, the Ministry of Culture and Tourism, and the General Administration of Radio and Television jointly researched and formulated the ≪Guiding Opinions on Promoting the Deep Integration of Culture and Science and Technology≫ ≪Top 10 Frontier Application Trends of Culture and Technology in 2022≫Published by Tencent Research Institute
Cultural industry	News, Newspaper, Radio, Television, Internet Information Services, Publishing, Movie, Creative Performances, Games, Libraries, Cultural Heritage, Cultural Relic, Museums, Crafts, Artistic Ceramics, Creative Design, Advertising, Artworks, Collectibles, Playgrounds, Parks, Scenic Spots, Gardens, Conventions and Exhibitions, Cultural Equipments, Copyrights, Data Assets	≪Classification of Cultural and Other Related Industries in China≫Published by the National Bureau of Statistics of China in 2018, this classification reflects the production activities of China's cultural and related industries in a more comprehensive manner and provides a standard basis for classification.

Text information on business scope. The "business scope" refers to the range of business activities and fields disclosed in an enterprise or institution's business registration information. This information typically includes the main products or services the enterprise engages in, reflecting its market positioning and technological capabilities. Additionally, as part of the business registration information, the business scope has a high degree of standardization and credibility, providing a foundation for large-scale analysis. In the context of the cultural industry, the business scope may include various aspects such as cultural creativity, cultural content development, and cultural consumption. Therefore, the business scope not only provides detailed information about an enterprise's involvement in the industrial and ecological chains but also helps clarify the role and function of each enterprise in the integration process at different levels. In the research process, we use the business scope data to deepen our understanding of each enterprise and institution in the industrial and ecological chains. The construction of the industrial and ecological chain corpora is based on the classification and stratification of these business scope texts. Specifically, by analyzing the business scope of each enterprise, we extract keywords and themes related to the integration of digital technologies and the cultural industry, mapping them to different links in the industrial and ecological chains.

Through the Enterprise Search platform QiChaCha.com, this paper batch searched the individual names involved in the above patents to get the business scope information of 2231 of them and used them as the text to analyze.

(1) *Industrial chain corpus*: The cultural industry chain includes three major links: cultural creativity supply, cultural content R&D, and cultural product marketing (Hao and Huang, 2021). When constructing the industrial chain corpus, we set keywords for each industrial chain link (such as "cultural creativity", "cultural development", "cultural consumption", etc.) and collect the business scope information of enterprises involved in these industrial links. Based on this information, we construct the industrial chain corpus. Miric et al. (2023) used text classification methods to categorize AI patents. This paper follows their approach by classifying enterprises based on their business scope and assigning each entity to a specific industrial chain link. This classification method is based on the specific content of the business scope text and employs a Naive Bayes classification method. Therefore, in the process of supplying cultural creativity, this paper searched for "cultural creativity" as the keyword in the

QiChaCha platform. In the process of cultural content R&D, this paper used "cultural R&D" and "cultural science and technology" as keywords to search; in the process of cultural product marketing, this paper used "cultural marketing", "cultural consumption", "cultural experience", and "cultural derivatives" as keywords to search. Finally, a total of 600 individual names and business scope information from the three industrial chain links were obtained. This paper took this as the basis for analyzing the division of labor in the industry chain for the development of digital-real integration in the cultural industry.

(2) *Ecological chain corpus*: In constructing the ecological chain corpus, the business scope text information helps us distinguish the position of enterprises in the broader ecosystem of digital-physical integration in the cultural industry. We set keywords for each ecological chain module (such as "scene construction", "ecological interconnection", "VR/AR", etc.), collect business scope information from enterprises involved in these ecological modules, and use this information to build the ecological chain corpus. Hoberg and Phillips (2016) used text analysis techniques to examine product descriptions in company annual reports, proposing the concept and measurement of product similarity. This paper adopts their methodology, and using the cosine similarity algorithm, analyzes the business scope of the entities in the study, categorizing them based on the module they occupy within the ecological chain.This paper collected the keywords of each ecological group from official documents and authoritative reports and obtained the individual names and business scope information related to the ecological chain for the development of digital-real integration in the cultural industry with the help of QiChaCha.com Platform, thus taking it as the basis for the classification of the ecological chain, as shown in Table 2.

Research methods
Social network analysis. To better meticulously present the progress of the integration of digital-real aspects in the cultural industry, the social network analysis method was used to quantitatively examine the relationships between network entities to understand the characteristics of the network structure (Wasserman, 1994; Scott, 2012). Therefore, this article referenced the approach of Hu et al. (2024) and conducted a social network analysis to characterize the network features of the development

Table 2 Related search keywords for links in the ecological chain.

Ecological llevels	Ecological groups	Keywords	Source
Industrial level	Scene construction	New Scenes of Digital Cultural Consumption, Immersive Opera, High Definition Image Database, Digital Exhibition, Virtual Reality Dance Performance, Scene on the Cloud, Online Exhibition Viewing, Live-action Performance, Intelligent Tourism, Cultural Creative Parks, Cultural Activity Venues, Cultural Innovation Incubators, Digital Media Workshops, Cultural Heritage Preservation Centres, Cultural Industrial Parks	China's Ministry of Culture and Tourism released «Digital Innovation to Promote High-Quality Development of Culture and Tourism»[a]
	Ecological interconnection	Internet Audiovisual Convergence, Interactive Entertainment, Animation Media, Digital Music, Online Video Interaction, Live Webcasting, Cultural Technology R&D Chain, Internet Advertising Service, Cultural Industry Cross-Industry Integration	«New Forms of Cultural Industry in the Context of Interconnection and Interoperability» (Hua, 2015)
	Financial and tax services	Deposit and Loan Business, Securities Issuance, Insurance, Financial Consultancy, Financial Services, Business Guarantee, Popular Financl, Financing Guarantee, Financing and Credit Enhancement, Tax Services, Investment Management, Asset Management	«Notice on Strengthening the Implementation of Financial and Tax Support Policies Promoting the High-Quality Development of Small and Medium-Sized Enterprises»[b]
	Platform control	E-commerce Platform, Industry Internet, Public Culture Cloud Construction, Supply Chain Management, Internet Platform, Artificial Intelligence Platform, Double Innovation Service Platform, Cultural Data Service Platform, Cultural Digital Infrastructure Design	China Development and Reform Commission General Administration of Market Supervision Central Internet Information Office Ministry of Industry and Information Technology Ministry of Human Resources and Social Security Ministry of Agriculture and Rural Development Ministry of Commerce Ministry of the People's Bank of China General Administration of Taxation Jointly released «Opinions on Promoting the Standardized, Healthy and Sustainable Development of the Platform Economy»[c]
Technology level	5G	5G	See the theoretical section above
	AI	AI Artificial Intelligence	
	Big Data	Big Data	
	VR, AR	VR AR Virtual Reality Augmented Reality	
	Meta-Universe	Meta-Universe	
	Blockchain	Blockchain	
	Cloud Computing	Cloud Computing	
Foundation level	Cultural Data Collection	Cultural Data Collection	
	Cultural Data Processing	Cultural Data Processing	
	Cultural Data Trading	Cultural Data Trading	
	Cultural Data Distribution	Cultural Data Distribution	
	Cultural Data Presentation	Cultural Data Presentation	

The division of the links in the cultural industry ecosystem chain refers to the paper by Huang and Yu (2024).
[a] The Ministry of Culture and Tourism of China announced a number of "Top 10 Cases" and "Excellent Cases" of digital innovation demonstration in culture and tourism, focusing on the demonstration and promotion of the latest achievements in digital innovation application scenarios in the field of culture and tourism, https://zwgk.mct.gov.cn/zfxxgkml/zcfg/zcjd/202310/t20231011_949074.html.
[b] In accordance with the decision-making and deployment of China's Party Central Committee and State Council, a series of fiscal and taxation policies have been introduced to support small and medium-sized enterprises (SMEs).
[c] The platform economy is a new type of economic form with the internet platform as the main carrier, data as the key production factor, new-generation information technology as the core driving force, and network information infrastructure as the important support. The opinion clearly points out that it is necessary to adapt to the law of development of the platform economy, establish and improve the rules and systems, optimize the environment for the development of the platform economy, and further promote the standardized, healthy and sustainable development of the platform economy, https://www.gov.cn/zhengce/zhengceku/2022-01/20/content_5669431.htm.

of digital-real integration in cultural industries. Specific network topology indicators selected were as follows.

(1) Degree had been defined as the number of relationships between nodes, which characterized the active degree of a node's connection with other nodes. The mean value of the degrees of all the nodes in the network was the average degree of the network, and the formula was

$$D = d(x) \tag{1}$$

(2) Network density was used as a measure of the overall closeness of the network and was defined as the ratio of the total number of links actually presented in the network to the theoretical maximum total number of links. Using the

formula below:

$$D = \frac{\sum_t^N k_i}{N(N-1)} \qquad (2)$$

where k_i was the degree centrality of the node, and N was the number of nodes.

(3) The average length of the path was the average value of the shortest length of the path among all pairs of nodes in the network. The smaller the value, the smoother and more convenient the connectivity between the nodes, and the formula was:

$$L = \frac{\sum_j^N d_{ij}}{N(N-1)} \qquad (3)$$

where d_{ij} was the shortest length of the path between node i and node j, and N was the number of nodes in the network.

(4) The clustering coefficient reflected the clustering of nodes in the network. It was defined as the ratio of the number of edges between a node and all its neighbors to the theoretically possible maximum number of connected edges. The formula was:

$$C = \frac{2E_i}{Nk_i(k_i-1)} \qquad (4)$$

where E_i was the actual number of connected edges between node i and all neighboring nodes, N is the number of nodes in the network, and k_i is the degree of node i.

(5) The modularity was mainly used to measure the strength of the network divided into clusters. The network with a higher degree of modularity has a strong connection in each module, while different modules are more sparsely connected to each other. It is defined as the number of edges within a module minus the expected value of the number of edges when randomly distributed using the formula.

$$Q = \sum_i e_{ii} - E_r \qquad (5)$$

Text classification method—Naive Bayes classifier. Text classification was a method that calculated the category information to which the document data belonged by calculating it based on the content of the document data in a classification system that had already been trained and learned (McCallum and Nigam, 1998; Kim et al. 2006). The specific steps of the text classification method included five steps: text pre-processing, text representation, classification model training, performance evaluation, and classification determination (Kang et al. 2020).

(1) *Data pre-processing*: In this paper, the *Jieba Chinese Text Segmentation* in the Python environment was selected to conduct segmentation processing on the business scope text of the above industrial chain and ecological chain enterprises, and the business scope texts were cut into separate words. Further, the deactivated stopwords from the Harbin Institute of Technology (HIT) were used to delete the useless contents in the text, such as punctuation marks, adverbs, and so on. Because the enterprise business scope text used in this paper also contained useless information such as state-restricted operation, projects required to be approved according to the law, general scope, business license, and so on, which also affected the accuracy of the text classification, so they were also deleted in the text pre-processing process.

(2) *Text representation*: Text representation was to use the relevant module to denote the text into a machine-readable pattern, i.e., the text was transformed into the form of word vector, which was easy to process later. This paper selected a bag-of-words model for the textual representation of the above textual information, which is implemented based on the CountVectorizer tool in Scikit-learn. The bag-of-words model assumed that the text consists of a collection of several words combined, and it calculated the frequency of the words in the text appearing in the full lexicon, which constitutes the word vectors of that text (Salton and Buckley, 1988). Even though the bag-of-words method ignored the sentence structure of phrases and sentences and lost a large amount of information contained in the word sequence, given that the text analyzed in this paper was the information on the business scope of enterprises, the content was more scattered and fragmented and did not have the problem of ordering. Hence, the bag-of-words method was still applicable to a certain extent.

(3) *Classification model training*: Classification model training was used to select and train a suitable classifier to get the classification model. In this paper, the approach of Kaushik et al. (2022) was taken as a reference to select a Naive Bayesian classifier for training and accomplishing the final classification goal. The idea was to know the prior probability, calculate the conditional probability of a given text to be classified into each category, compute the posterior probability based on the prior probability and the conditional probability, and finally select the category with the highest calculated posterior probability as the category to which the text belongs to, which was calculated by the formula:

$$p(c_j|d_I; \theta) = \frac{p(c_j|\theta)p(d_i|c_j; \theta)}{p(d_i|\theta)} \qquad (6)$$

where θ denoted the unknown parameter set in the probabilistic model; $p(c_j|d_I; \theta)$ represented the posterior probability that text d_i belonged to category c_j, under the condition that the text d_i and parameter θ were known; $p(c_j|\theta)$ stood for the prior probability of category c_j. $P(d_i|\theta)$ implied the probability of text d_i if the parameter set θ is presupposed.

(4) *Performance evaluation of classifiers*: This part evaluated the classification model's performance using the existing pre-labeled test dataset for classification performance, which was commonly evaluated by the accuracy rate, checking precision rate, recall rate, and F1 value metrics (Meque et al. 2023). Among them, the Accuracy rate was used to evaluate the overall recognition ability of the model, i.e., the number of correctly classified texts as a percentage of the total number of texts, which was calculated by the formula:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \qquad (7)$$

The Precision rate is the ratio of the number of texts correctly determined in a category to the total number of texts determined to be in that category, with the formula:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad (8)$$

The Recall rate was the ratio of the number of texts correctly identified for a category to the number of texts actually belonging to that category, and the formula was:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad (9)$$

The Precision and Recall rate reflect the performance of the classification algorithm from different perspectives, while the F1 value better balances the Precision rate and recall rate by combining the two for a comprehensive evaluation,

which was given by the formula:

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (10)$$

Among them, TP indicated the number of texts correctly determined as belonging to a category; FP indicated the number of texts incorrectly determined as belonging to a category by the classifier; FN indicated the number of texts actually belonging to a category but determined as belonging to other categories by the classifier, and TN indicated the number of texts incorrectly determined by the classifier as not belonging to the category.

(5) *Classification evaluation*: Use the trained classifier to classify the unlabeled to-be-classified text to accomplish the final text classification goal.

Text matching method—corpus-based cosine similarity algorithm. The text-matching method was identifying and locating specific text in the original data, mainly done by measuring the similarity between words, sentences, paragraphs, and documents (Islam and Inkpen, 2008; Alqahtani et al. 2021). The steps of text matching included data pre-processing, text representation, and similarity calculation. Among them, data pre-processing and text representation were done as above.

In terms of similarity calculation, this paper selected the cosine similarity algorithm (Huang, 2008; Bewoor and Patil, 2018), which was widely used today, to calculate the similarity between the texts of cultural enterprise operation and the ecological chain corpus, using the following formula:

$$S(D1, D2) = \cos\theta = \frac{\sum_{i=1}^{n} D1_i D2_i}{\sqrt{\sum_{i=1}^{n} D1_i^2} \times \sqrt{\sum_{i=1}^{n} D2_i^2}} \qquad (11)$$
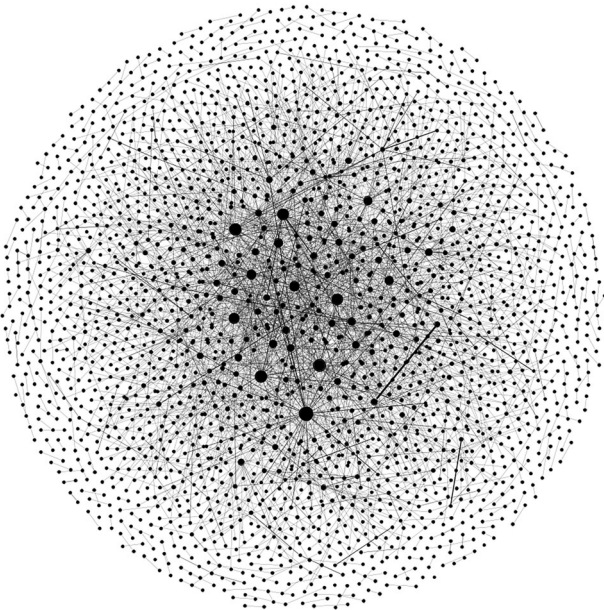
Fig. 2 Organizational form for the development of digital-real integration in cultural industries.

where $D1_i$, $D2_i$ were the word vectors of different texts, respectively. The calculated cosine similarity is between 0 and 1. The closer the value is to 1, the higher the degree of similarity between the texts. In this paper, the threshold of cosine similarity was set to 0.7 (Okazaki and Tsujii, 2010), which is used to determine the existence of ecological operation for the development of digital-real integration in the cultural industry.

Results

The characteristics of organizational form. In this paper, 3146 pairs of individual patent citations information were imported into Gephi software, and the number of citations among their patents was used as the weight of the edges. Based on the Fruchterman Reingold layout, the values of zone, gravity, and speed parameters are set to 22,370, 10, and 1, respectively. The organizational form for the development of digital-real integration of cultural industry was visualized and presented, and the results were shown in Fig. 2. Among them, the points represented different individuals in the digital-real integration network; the edges represented the referential relationship between individuals; the thickness of the edges represented the size of reference frequency between individuals. Further, this paper utilized Gephi software to present its network structure in detail.

As shown in Table 3, the average degree and network density of the organizational network for the development of digital-real integration in the cultural industry were 1.343 and 0.001, respectively. Based on the measure of Hu et al. (2024), the overall activity of its network was low, and the nodes were more loosely connected to each other. Meanwhile, the average path length and clustering coefficient of the digital-real integration chain cluster network of the cultural industry were 6.41 and 0.01, respectively, which indicated that in the network's internal nodes, any two points needed to be mediated more than six times on average before a connection can be generated. Namely, the internal accessibility was low, and the flow of information was not well disseminated in the network. In addition, the relatively low clustering coefficient of the chain-cluster network also indicated that the network did not form an obvious agglomeration pattern, i.e., it was not characterized as a small-world network (Eslami et al. 2013). However, it was worth noting that the network's modularity reached 0.715, indicating that the network had typical modularity characteristics, i.e., tighter cliques within the network (Lacasa and Shubbak, 2018).
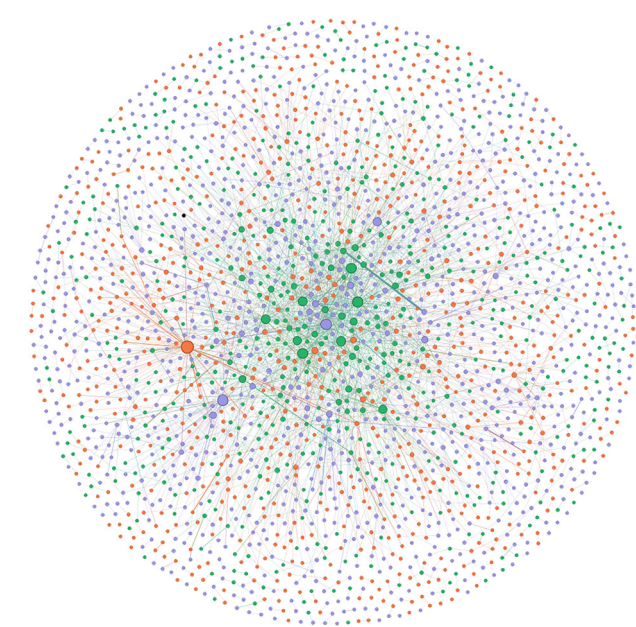
Division of labor in the industrial chain. First, this paper used the train_test_split command in the scikit-learn library under the Python environment to split the industrial chain corpus into a training and test set. In reference to Kaushik et al. (2022), 80% was used as the training dataset and 20% as the test dataset, and this was used as the basis for training the classifier using the Multinomial NB command. The results showed that the overall accuracy of the model was 95.59%, and the accuracy, precision, recall, and F1 values for each category were shown in Table 4, which were all above 80%. It can be determined that the model has good classification performance.

The classifier trained above was used to explore the division of labor in the industrial chain for the development of digital-real integration in the cultural industry, as shown in Fig. 3. The nodes

Table 3 Network topology.					
Indicator name	Average degree	Network density	Average path length	Clustering coefficient	Modularity
Numerical value	1.343	0.001	6.41	0.01	0.715

Table 4 Classification results.

Industrial chain links	Precision	Recall	F1-score
Cultural creativity supply	0.86	0.87	0.86
Cultural content R&D	0.94	0.80	0.86
Cultural product marketing	0.81	0.93	0.87

Fig. 3 Diagram of the division of labor in the industry chain for the development of digital-real integration in cultural industry.

of different colors in the figure represented individuals of different segments in the cultural industry chain. Among them, the purple point represented the category of cultural creativity supply, which occupied 44.32% of the overall industry chain; the green point indicated the category of cultural content R&D, which accounted for 32.36% of the overall industry chain; and the orange point represented the category of cultural product marketing, which accounted for 23.32% of the overall industry chain. It could be seen that even though the level of network agglomeration in the development of digital-real integration in the cultural industry was limited, a distinctive division of labor in the industry chain had been formed.

Division of labor in the ecological chain. The ecological chain similarity calculation was carried out using the 'cosine_similarity' command in Python software, and the results were shown in Table 5. At the industry level, the number of enterprises with a similarity degree above 0.7 in the finance and tax service and platform management group reached 52 and 46, respectively, among which the representative enterprises of Tianjin Jincheng Bank Co., Ltd. and Datang Convergence IOT Technology Wuxi Co., Ltd. both had similarity degrees above 0.9. In the ecological interconnection group, there were 12 enterprises with a similarity degree of 0.7 or above. In contrast, the similarity degree of the scene construction group was generally lower than 0.7, which indicated that the scene construction group was the weak group in the ecological chain in the digital-real integration network for the cultural industry. At the basic level, the similarity degree of the cultural data presenting group had reached 45 enterprises with a similarity degree of 0.7 or above, among which the

similarity degree of Weichuang Group Company Limited was 0.97. The similarity degrees of the cultural data processing and the cultural data trading groups were 14 enterprises each, and each enterprise's similarity degree was 0.7 or above.

In contrast, the number of enterprises with a similarity of 0.7 or more in cultural data collection and cultural data distribution was 3 and 0, respectively, which meant an urgent need to build and supplement chains. At the technology level, the number of enterprises with a similarity of 0.7 or more in the blockchain chain reached as high as 101, and the similarity degree of the representative enterprise, Shandong Guodun Information Science and Technology Company Limited, was 0.836. The numbers of enterprises with a similarity of 0.7 or more in 5G, big data, and cloud computing technologies were 33, 20, and 49, respectively. However, the applications of AI and VR/AR technologies were fewer, and the number of enterprises with a similarity degree above 0.7 was not more than 10. At the same time, none of the meta-universe technology group had a similarity degree of 0.7. This indicated that the development of digital-real integration in China's cultural industry had not yet reached the advanced stage of ecological operation and that the technology gap still needed to be further guided and reinforced.

Discussion and conclusion
Given that previous research on the integration of digital-real industries has predominantly relied on qualitative analysis and employed overly generalized methods for assessing the level of integration, this study takes the cultural industry in China as a case study. By leveraging social network analysis, text classification and text similarity analysis, this paper systematically compares the characteristic differences across various stages of digital-real integration in the cultural industry over time. Furthermore, it proposes a scientifically robust methodological framework for evaluating the development of digital-real integration, offering a significant advancement in the field.

Conclusion. Based on China's patent information, this paper found that the cultural industry has gradually moved towards the development of digital-real integration, and its cooperation network has presented a topological structure and possessed a complete industrial chain. However, the degree of agglomeration of the cooperative network for the development of digital-real integration in the cultural industry still remains to be improved. It has not yet realized a fully ecological operation, especially in the ecological chain links such as cultural data collection, cultural data distribution, meta-universe technology, VR/AR technology, and so on. Increasing the investment to promote the cultural industry better and build, strengthen, and supplement the chain is still necessary. To summarize, this paper argues that the road to the development of digital-real integration in the cultural industry is still a long way to go.

Theoretical implications. The study provides at least two theoretical implications. First, it suggests that science should be used to examine the digital-real integration industrial development trend. Previous studies focused on the development of digital-real integration of industries have mainly focused on the level of qualitative logical deduction, and few studies have quantitatively portrayed the development trend of digital-real integration of industries. At the same time, there is a lack of detailed insight into its organizational characteristics and division of labor. Therefore, this paper measures the development trend of digital-real integration in the cultural industry from the network characteristics and division of labor levels in the industrial chain and ecological

Table 5 Results of the division of labor in the ecological chain and representative enterprises.

Level	Module	Number of enterprises with a similarity ratio of 0.7 or more	Representative individuals (similarity)
Industrial level	Scene building	0	Guangzhou Prosperous Star Hi-Entertainment Information Technology Co. (0.643)
	Ecological interconnection	12	Shenzhen Zhugegua Technology Co. (0.874)
	Property tax services	52	Tianjin Jincheng Bank Co. (0.919)
	Platform governance	46	Datang Fusion IOT Technology Wuxi Co. (0.983)
Foundation level	Cultural data collection	3	Speed Technology Co. (0.824)
	Cultural data processing	14	Nanchang Black Shark Technology Co. (0.928)
	Cultural data exchange	14	Hangzhou Yulin Technology Co. (0.867)
	Cultural data distribution	0	Hangzhou Yulin Technology Co. (0.696)
	Cultural data presentation	45	Weichuang Group Co. (0.97)
Technology level	5G	33	Hangzhou Tianshi Information Technology Co. (0.907)
	AI	8	Xi'an Zhongyijian Technology Group Co. (0.801)
	Big Data	20	Shaanxi Yanyi Yimou Industrial Intelligence Technology Co. (0.791)
	VR, AR	5	Shenzhen Simaiyun Technology Co. (0.941)
	Universe	0	Hangzhou Ant Kool-Aid Technology Co. (0.677)
	Blockchain	101	Shandong Guodun Network Information Technology Co. (0.836)
	Cloud Computing	49	Zhongcheng Hualong Computer Technology Co. (0.986)

chain, expanding the research field and enhancing the research tools.

Second, it is better to focus on the far-reaching impact of the development of digital-real integration on the change of industrial organization. The digital-real integration not only reshapes the online and offline cooperation relationship in the cultural industry but also changes the division of labor in the cultural industry chain and facilitates the formation of new business models and industrial ecology. This paper has revealed that digital-real integration is boosting the evolution of industrial organizations in a more flattened, networked, and collaborative direction through studying patent information, which will provide theoretical support for transforming and upgrading cultural enterprises in the context of digital-real integration.

Practical implications. The study provides at least two practical implications. First, customized policy support for weak links. Different from the previous general suggestions for optimizing the development of digital-real integration in the cultural industry, this paper points out the weak links in the industrial chain and ecological chain in the development process of digital-real integration in the cultural industry through empirical measurements, which is beneficial for the cultural industry itself and the relevant authorities can provide support and supplementation in a more targeted way. In other words, only by clarifying more clearly what the cultural industry needs to improve in building, strengthening, and supplementing the chain. It is possible to find the root causes of the problem to help the cultural industry develop digital-real integration better.

Second, multi-party cooperation is needed to promote innovative applications. Given that the level of cooperation network agglomeration in the development of digital-real integration in the cultural industry is still low, cultural enterprises, scientific research institutions, and government sectors must work together to formulate the direction of improvement based on the weak links to foster close collaboration and joint efforts among all parties, thus pushing forward the optimization and upgrading of the cultural industry chain and the ecological chain. The efficiency and value of each link in the cultural industry chain can only be truly enhanced by actively encouraging more cultural enterprises to explore and apply new technologies in digital-real integration.

Research limitations and future research directions. In terms of research limitations, this paper chose to utilize the patent cooperation network to approximate the morphological characteristics of the development of the digital-real integration in the cultural industry as a substitute, and based on this, it measured the division of labor in its industrial chain and ecological chain. This approach may produce some errors in the research results due to the limitations in the scope of patent search, years of experience, and degree of matching. However, since the Chinese government released the "14th Five-Year Plan for the Development of Cultural Industry" in 2021, which will vigorously promote the digitalization of the cultural industry, the data bias will not have a hard impact on the research conclusions. In the future, as the advanced stage of the development of digital-real integration for the cultural industry, the ecological operation of the cooperation network and the value co-creating mechanism among the entities in the chain group will be the possible direction for the digital-real integration of the cultural industry in the future to develop further. How to more comprehensively capture the ecological interaction behavior of the digital-real integration network of cultural industry and scientifically express the value co-creation process of cooperative entities under the ecological operation will be the focus of future research that needs to be explored urgently.

Data availability

The datasets generated and/or analyzed during the current study are publicly available and are hand-collected but are available from the corresponding author upon reasonable request.

References

Alqahtani A, Alhakami H, Alsubait T, Baz A (2021) A survey of text matching techniques. Eng, Technol Appl Sci Res 11(1):6656–6661. https://doi.org/10.48084/etasr.3968

Alshammari KH, Alshallaqi M, Al-Mamary YH (2024) Digital transformation dilemma in the era of changing dynamics: how organizational culture influence the success of digital transformation. Hum Syst Manag 43(4). https://doi.org/10.3233/HSM-230163

Balaji MS, Roy SK (2017) Value co-creation with Internet of things technology in the retail industry. J Mark Manag 33(1-2):7–31. https://doi.org/10.1080/0267257X.2016.1217914

Bewoor MS, Patil SH (2018) Empirical analysis of single and multi document summarization using clustering algorithms. Eng Technol Appl Sci Res 8(1):2562–2567. https://doi.org/10.48084/etasr.1775

Benson CL, Magee CL (2013) A hybrid keyword and patent class methodology for selecting relevant sets of patents for a technological field. Scientometrics 96:69–82. https://doi.org/10.1007/s11192-012-0930-3

Borangiu T, Trentesaux D, Thomas A, Leitão P, Barata J (2019) Digital transformation of manufacturing through cloud services and resource virtualization. Comput Ind 108:150–162. https://doi.org/10.1016/j.compind.2019.01.006

Breidbach CF, Maglio PP (2015) A service science perspective on the role of ICT in service innovation. ECIS 2015 Research-in-Progress Papers, Paper 33. https://aisel.aisnet.org/ecis2015_rip/33

Bustamante E (2004) Cultural industries in the digital age: some provisional conclusions. Media Cult Soc 26(6):803–820. https://doi.org/10.1177/0163443704047027

Bygstad B, Lanestedt G (2009) ICT based service innovation—a challenge for project management. Int J Proj Manag 27(3):234–242. https://doi.org/10.1016/j.ijproman.2007.12.002

Calisi D, Botta S (2022) Virtual reality and captured reality for cultural landscape communication. Int Arch Photogramm Remote Sens Spat Inf Sci 46:113–120. https://doi.org/10.5194/isprs-archives-XLVI-2-W1-2022-113-2022

Choi J, Yoon J (2022) Measuring knowledge exploration distance at the patent level: application of network embedding and citation analysis. J Informetr 16(2):101286. https://doi.org/10.1016/j.joi.2022.101286

Classification of Cultural and Other Related Industries in China (2018) https://www.stats.gov.cn/sj/pcsj/jjpc/4jp

de Koning K, Broekhuijsen J, Kühn I, Ovaskainen O, Taubert F, Endresen D, Schigel D, Grimm V (2023) Digital twins: dynamic model-data fusion for ecology. Trends Ecol Evol 38(10):916–926. https://doi.org/10.1016/j.tree.2023.04.010

Djamfa PN (2024) Technological innovation and development of the publishing industry: the challenging digital transformation of publishing houses in Cameroon. Open J Appl Sci 14(7):1863–1885

Elia G, Margherita A, Passiante G (2020) Digital entrepreneurship ecosystem: How digital technologies and collective intelligence are reshaping the entrepreneurial process. Technol Forecast Soc Change 150:119791. https://doi.org/10.1016/j.techfore.2019.119791

Eslami H, Ebadi A, Schiffauerova A (2013) Effect of collaboration network structure on knowledge creation and technological performance: the case of biotechnology in Canada. Scientometrics 97:99–119. https://doi.org/10.1007/s11192-013-1069-6

Firican DA (2024) Creating a digital culture for digital transformation: a literature review of practical steps. Proc Int Conf Bus Excell 18(1):1018–1028. https://doi.org/10.2478/picbe-2024-0089

Fu Y, Yu H, Da J, Kan X (2022) Discussion on the influence of digital economy development on real economy. Front Bus Econ Manag 6(3):176–179

Guan J, Liu N (2016) Exploitative and exploratory innovations in knowledge network and collaboration network: a patent analysis in the technological field of nano-energy. Res Policy 45(1):97–112. https://doi.org/10.1016/j.respol.2015.08.002

Hao TL, Huang YL (2021) On the modernization of the digital cultural industry chain under the new development pattern of dual circulation. Jianghan Tribune (4), 127–133

He Z, Wen C (2025) Construction of digital creation development model of intangible cultural heritage crafts in China. Humanit Soc Sci Commun 11(1):1–14. https://doi.org/10.1057/s41599-024-04331-4

Hoberg G, Phillips G (2016) Text-based network industries and endogenous product differentiation[J]. J Political Econ 124(5):1423–1465. https://doi.org/10.1086/688176

Hua J (2015) New forms of cultural industry under the background of interconnection. J Beijing Union Univ (Humanit Soc Sci) 13(2):24–30. https://doi.org/10.16255/j.cnki.11-5117c.2015.02.005

Huang A (2008) Similarity Measures for Text Document Clustering. In: New Zealand Computer Science Research Student Conference (NZCSRSC), Christchurch

Huang R, Yu H (2024) The integration of digital and physical aspects in the cultural industry: causes, mechanisms, and ecological vision. Jianghan Tribune (11):139–144

Hu F, Qiu L, Wei S, Zhou H, Bathuure IA, Hu H (2024) The spatiotemporal evolution of global innovation networks and the changing position of China: a social network analysis based on cooperative patents. RD Manag 54(3):574–589. https://doi.org/10.1111/radm.12662

Islam A, Inkpen D (2008) Semantic text similarity using corpus-based word similarity and string similarity. ACM Trans Knowl Discov Data 2(2):1–25. https://doi.org/10.1145/1376815.1376819

Kang Y, Cai Z, Tan CW, Huang Q, Liu H (2020) Natural language processing (NLP) in management research: a literature review. J Manag Anal 7(2):139–172. https://doi.org/10.1080/23270012.2020.1756939

Karvonen M, Kässi T (2013) Patent citations as a tool for analysing the early stages of convergence. Technol Forecast Soc Change 80(6):1094–1107. https://doi.org/10.1016/j.techfore.2012.05.006

Kaushik K, Bhardwaj A, Dahiya S, Maashi MS, Al Moteri M, Aljebreen M, Bharany S (2022) Multinomial Naive Bayesian classifier framework for systematic analysis of smart IoT devices. Sensors 22(19):7318. https://doi.org/10.3390/s22197318

Kim S-B, Han K-S, Rim H-C, Myaeng SH (2006) Some effective techniques for Naive Bayes text classification. IEEE Trans Knowl Data Eng 18(11):1457–1466

Lacasa ID, Shubbak MH (2018) Drifting towards innovation: the co-evolution of patent networks, policy, and institutions in China's solar photovoltaics industry. Energy Res Soc Sci 38:87–101. https://doi.org/10.1016/j.erss.2018.01.012

Lee WS, Han EJ, Sohn SY (2015) Predicting the pattern of technology convergence using big-data technology on large-scale triadic patents. Technol Forecast Soc Change 100:317–329. https://doi.org/10.1016/j.techfore.2015.07.022

Lerro A, Schiuma G, Manfredi F (2022) Editorial: entrepreneurial development and digital transformation in creative and cultural industries: trends, opportunities and challenges. Int J Entrep Behav Res 28(8):1929–1939. https://doi.org/10.1108/IJEBR-11-2022-008

Li L (2021) In-depth integration and development of digital economy and real economy. Sci J Econ Manag Res 3(3):253–256

McCallum A, Nigam K (1998) A Comparison of Event Models for Naive Bayes Text Classification. Proceedings in Workshop on Learning for Text Categorization, AAAI'98, pp 41–48

Meng XN, Xu SC, Hao MG (2023) Can digital-real integration promote industrial green transformation: fresh evidence from China's industrial sector. J Clean Prod 426:139116. https://doi.org/10.1016/j.jclepro.2023.139116

Meque AGM, Hussain N, Sidorov G, Gelbukh A (2023) Machine learning-based guilt detection in text. Sci Rep 13(1):11441. https://doi.org/10.1038/s41598-023-38171-0

Miric M, Jia N, Huang KG (2023) Using supervised machine learning for large-scale classification in management research: the case for identifying artificial intelligence patents. Strateg Manag J 44(2):491–519. https://doi.org/10.1002/smj.3441

Montecchi T, Russo D, Liu Y (2013) Searching in Cooperative Patent Classification: comparison between keyword and concept-based search. Adv Eng Inform 27(3):335–345. https://doi.org/10.1016/j.aei.2013.02.002

Okazaki N, Tsujii J (2010) Simple and Efficient Algorithm for Approximate Dictionary Matching. In: Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010). Coling 2010 Organizing Committee, Beijing, China, pp 851–859

Pais I (2019) Digital Platforms and the transformations in the division of labor. In: Handbook of labor, human resources and population economics. pp 1–16

Passiante G, Secundo G (2002) From geographical innovation clusters towards virtual innovation clusters: the Innovation Virtual System. ERSA conference papers ersa02p270. European Regional Science Association

Rosenberg N (1963) Technological Change in the Machine Tool Industry, 1840–1910. J Econ Hist 23(4):414–443. http://www.jstor.org/stable/2116207

Salton G, Buckley C (1988) Term-weighting approaches in automatic text retrieval. Inf Process Manag 24(5):513–523. https://doi.org/10.1016/0306-4573(88)90021-0

Scott J (2012) What is social network analysis? Bloomsbury Academic

Song J, Yuxin S (2020) An empirical study on the effect of digital economy on real economy. Sci Res Manag 41(5):32

Tiwari S (2021) Supply chain integration and Industry 4.0: a systematic literature review. Benchmarking: Int J 28(3):990–1030. https://doi.org/10.1108/BIJ-08-2020-0428

Tracey P, Clark GL (2003) Alliances, networks and competitive strategy: rethinking clusters of innovation. Growth change 34(1):1–16. https://doi.org/10.1111/1468-2257.00196

Tsolakis N, Zissis D, Tjahjono B (2023) Scrutinising the interplay between governance and resilience in supply chain management: a systems thinking framework. Eur Manag J 41(1):164–180. https://doi.org/10.1016/j.emj.2021.11.001

Urbinati A, Chiaroni D, Chiesa V, Frattini F (2020) The role of digital technologies in open innovation processes: an exploratory multiple case study analysis. R&D Manag 50(1):136–160. https://doi.org/10.1111/radm.12313

Wang M, Huang Y (2024) A digital technology–cultural resource strategy to drive innovation in cultural industries: a dynamic analysis based on machine learning. Technol Soc 6:77. https://doi.org/10.1016/j.techsoc.2024.102590

Wang Y, Shi M, Liu J, Zhong M, Ran R (2025) The impact of digital–real integration on energy productivity under a multi-governance framework: the mediating role of AI and embodied technological progress. Energy Econ 142:108167. https://doi.org/10.1016/j.eneco.2024.108167

Wang Z, Lin S, Chen Y, Lyulyov O, Pimonenko T (2024) The digital economy and real economy: the dynamic interaction effect and the coupling coordination degree. Sustainability 16(13):5769. https://doi.org/10.3390/su16135769

Wasserman S (1994) Social network analysis: methods and applications. The Press Syndicate of the University of Cambridge

Xiao XY (2020) Research on the integration of digital economy and real economy to promote high-quality economic development. In: 2020 Management Science Informatization and Economic Innovation Development Conference (MSIEID). IEEE, pp 8–11

Xu G, Lu T, Liu Y (2021) Symmetric reciprocal symbiosis mode of China's digital economy and real economy based on the logistic model. Symmetry 13(7):1136. https://doi.org/10.3390/sym13071136

Zhang D, Wang Y (2022) The mechanism and empirical study of the integration of digital economy and real economy to promote highquality economic development. Int J Sci Acad Res 8(3):4280–4287

Zhengzhen Z (2021) Research on the path of high-quality integration development of digital economy and real economy of China during the 14th Five-year Plan period. West Forum Econ Manag 32(6):29–36. https://doi.org/10.12181/jjgl.2021.06.04

Author contributions

Rui Huang: Conceptualization, methodology, writing—original draft. Hongqian Wang: Methodology, data curation, validation, and writing.

Competing interests

The authors declare no competing interests.

Ethical approval

Ethical approval was not required as the study did not involve human participants.

Informed consent

Inform consent was not required as the study did not involve human participants.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-025-04743-w.

Correspondence and requests for materials should be addressed to Hongqian Wang.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.