

ORIGINAL ARTICLE

Constructing a five-dimensional framework for traditional village building protection in China: A BERTopic bibliometric analysis linked to rural tourism

Wenjun Ge^{1,2}, Mohd Shahrudin Abd Manan^{1*}, and Syazwani binti Sahrir³¹Department of Architecture, Faculty of Design and Architecture, University of Putra Malaysia, Serdang, Selangor, Malaysia²Department of Art, Faculty of Art, Danchi College, Kunming, Yunnan, China³Department of Environmental Management, Faculty of Forestry and Environment, University of Putra Malaysia, Serdang, Selangor, Malaysia***Corresponding author:**Mohd Shahrudin Abd Manan
(am_shahrudin@upm.edu.my)

Citation: Ge, W., Manan, M. S. A. & Sahrir, S. (2026). Constructing a five-dimensional framework for traditional village building protection in China: A BERTopic bibliometric analysis linked to rural tourism. *Journal of Chinese Architecture and Urbanism*, 8(3):025480092. <https://doi.org/10.36922/JCAU025480092>

Received: November 28, 2025**Revised:** January 22, 2026**Accepted:** January 29, 2026**Published online:** April 17, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract

Protecting traditional village buildings is a global challenge, as it relates not only to cultural identity but also to the sustainable development of rural tourism. Establishing a unified protection system has become an important objective in the field of cultural heritage. This study applies a bibliometric approach using the BERTopic model to analyze 602 academic papers published between 1999 and 2024 in the China National Knowledge Infrastructure and Web of Science databases. The analysis provides quantitative evidence for the integrated development of traditional building protection and rural tourism. Topic evolution analysis based on the BERTopic model reveals that research on traditional village buildings has progressed from building culture to material technology, then to function adaptation, community participation, and finally to digital transformation. This trajectory reflects a shift from static preservation to dynamic regeneration and from a focus on physical space to social relationships. Using topic similarity analysis and semantic clustering, nine original topics were consolidated into five main dimensions—building features, materials and construction, spatial and functional adaptability, building culture, and social culture—forming a five-dimensional framework for evaluating building protection. This framework has theoretical depth and potential for policy use. This study advances beyond traditional qualitative approaches by proposing a quantitative model of semantic modeling, topic evolution, and framework construction, offering structured and visualized data to support theoretical understanding and policy development in cultural heritage protection and rural tourism.

Keywords: Chinese traditional village architecture; Heritage conservation; BERTopic bibliometric analysis; Rural tourism; Adaptive reuse

1. Introduction

1.1. Research background

As the rural revitalization strategy advances, traditional village buildings—important carriers of local culture and historical memory—are receiving increasing attention at the national level. Supported by policies such as the Guiding Opinions on the Protection and Development of Traditional Villages, the List of Chinese Traditional Villages, and the 14th Five-Year Plan for Culture and Tourism Development, the protection and utilization of traditional buildings have become key measures for promoting rural cultural renewal and tourism growth. Compared with buildings officially listed as cultural heritage, many traditional houses and public buildings with historical value remain unrecognized as protected sites, yet they still possess significant cultural and economic potential (Long *et al.*, 2022; Y. Wang *et al.*, 2021). In practice, traditional buildings face challenges such as over-commercialization, homogenized appearances, and distorted renovations. In some areas, building structures and spatial layouts have been damaged due to the absence of systematic protection plans (Gao & Wu, 2017; S. Li *et al.*, 2017; Ji & Abd Manan MS, 2024; Xie *et al.*, 2024). While tourism development brings economic benefits, it often undermines the buildings' original cultural functions. When traditional buildings are converted into commercial spaces, their cultural significance becomes difficult to preserve (Cattaneo *et al.*, 2018). In addition, current protection strategies show limited community participation and often fail to align with local residents' needs, hindering the long-term sustainability of building conservation (Qianda *et al.*, 2021). Under continued national policy support, it is therefore critical to explore frameworks that balance historical value, modern functionality, and ecological harmony. This study uses BERTopic to identify the main themes in traditional village building protection, drawing on a large body of literature. It aims to propose a theoretical framework (Cano *et al.*, 2013; Su *et al.*, 2018) that provides academic support and practical insights for achieving a balanced development of cultural protection and rural tourism.

At present, the research and practice of traditional village building protection face several challenges. First, the implementation of protection policies varies widely across regions, with significant differences in resource allocation, funding, and management capacity. A unified and standardized system has yet to be established (Xie *et al.*, 2024). Second, existing protection efforts mainly focus on physical restoration and appearance repair, with limited attention to the social and cultural roles of buildings or their adaptation to modern living needs. Consequently, some

restored villages gradually lose practical use and become empty or purely decorative spaces (Gao & Wu, 2017; Manan & Yuan, 2024). Most existing studies rely on case studies or qualitative analyses. While these approaches provide in-depth insights into the protection experiences of certain areas, they are limited in terms of broader application and theoretical generalization. In addition, there is currently no standardized method for systematically analyzing the key factors affecting the protection of traditional buildings (Ji & Abd Manan MS, 2023; Qianda *et al.*, 2021). To address these gaps, this study uses BERTopic to review and analyze 602 related papers published from 1999 to 2024. It aims to construct a traditional building protection framework that includes dimensions such as historical value, functional use, and ecological harmony. By identifying key topics from the literature and integrating factors such as cultural authenticity, material restoration, and community participation, this study proposes protection strategies with both methodological and practical value (Blei *et al.*, 2003). Through a scientific protection framework, it is possible to maximize the dual goals of cultural heritage conservation and modern utilization (Olgun & Karatosun, 2019).

1.2. Literature review

This section reviews existing studies on rural heritage protection using diverse research methods. Table 1 summarizes key characteristics from these studies, including research subjects, objectives, research methods, presence of protection frameworks, policy relevance, and connections to rural tourism development.

1.2.1. Innovations

Current research on rural heritage has predominantly approached the topic from one-dimensional perspectives, such as historical authenticity or technical restoration, often lacking a systematic framework that integrates culture, function, and community. The main contribution of this study lies in its conceptual shift from a “static heritage” paradigm to a “living heritage” perspective. Unlike traditional approaches that prioritize physical preservation, this study emphasizes the interactive relationship between heritage and local inhabitants and identifies “continuous use” and “dynamic conservation” as core principles. Furthermore, this research addresses the methodological limitations of previous qualitative reviews by employing BERTopic modeling. Through a systematic analysis of 602 papers published from 1999 to 2024, the study identifies key factors affecting heritage protection and constructs a five-dimensional cooperative structure. This theoretical model provides a more integrated and methodologically rigorous foundation for future research in this field.

Table 1. Overview and classification of research on rural architecture protection frameworks

Author (year)	Research subject	Objectives	Research method	Research type	Includes heritage protection framework	Policy applicability	Integrated with rural tourism development
Du & Shi (2019)	Revitalization of rural heritage	Explore conservation strategies	Comparative analysis	Rural architecture	Yes	Yes	No
Zhao & Greenop (2019)	Semi-traditional rural architecture in China	Identify protection models	Photo comparison, interview	Rural architecture	Yes	No	No
Olukoya (2021)	Rural heritage value framework	Construct a land value model	Theoretical modeling	Rural architecture	Yes	No	Yes
Miao & Chiou (2013)	Architectural identity and tourism imagery	Understand identity transmission	Image analysis, semantic coding	Historic village	No	Yes	Yes
Olğün & Karatosun (2019)	Earthen heritage in Turkey	Sustain earthen architecture	Case study field observation	Architecture	Yes	No	No
Aktürk (2025)	Protection in rural Serbia	Summarize conservation practices	Case study, field investigation	Rural architecture	Yes	Yes	No
Tao, Chen & Aoki (2025)	Relocation of rural buildings in China	Analyze preservation ideologies	Literature review, theoretical analysis	Rural architecture	No	Yes	No
Z. Wu <i>et al.</i> (2025)	Rural tourism landscape evaluation	Propose a landscape assessment method	Spatial perception, assessment, interview	Rural architecture	No	Yes	Yes
Lindholm & Ekblom (2019)	Cultural landscape governance in China and the United Kingdom	Compare cultural landscape frameworks	Comparative analysis	Cultural heritage research	Yes	Yes	No
Djabarouti & Ren (2024)	Governance and policy mechanisms in multi-ethnic heritage	Explore institutional and community interactions	Policy analysis	Governance research	Yes	Yes	No
Saputra (2024)	GIS applications in heritage protection	Integrate spatial and sustainability data	GIS analysis	Technical research	Yes	Yes	No
B. Liu <i>et al.</i> (2024)	Topic structure identification in architectural heritage	Extract semantic topic structures	BERTopic modeling	Data-driven research	Yes	Yes	No

(cont'd...)

Table 1. (Continued)

Author (year)	Research subject	Objectives	Research method	Research type	Includes heritage protection framework	Policy applicability	Integrated with rural tourism development
T. Nguyen <i>et al.</i> (2025)	Cultural heritage value identification in Northern England	Reassess heritage values using topic modeling and grounded theory	Mixed methods	Cultural heritage research	Yes	Yes	No
Tenzer & Schofield (2024)	Relocation of rural architecture in China	Explore adaptive reuse concepts	Theoretical analysis	Case study	Yes	Yes	No
Tao <i>et al.</i> , 2025	Rural heritage and tourism integration in Turkey	Build a sustainable management model	Case study	Empirical research	Yes	Yes	No
Arias Tapiero <i>et al.</i> , 2025	Digital feature analysis of traditional architectural heritage	Explore information and communication technology-based methods in heritage conservation	ICT-based modeling	Technical research	Yes	Yes	No
This study	Rural heritage topics & protection	Identify and build a framework	BERTopic	Rural architecture	Yes	Yes	Yes

Abbreviations: GIS: Geographic information system; ICT: Information and communications technology

(a) Framework development for living heritage

Research on rural living heritage has traditionally focused on single-dimensional analyses such as historical authenticity, technical restoration, or cultural reuse, which lack a systematic theoretical framework that integrates culture, function, and community. Unlike static heritage, living heritage emphasizes the symbiotic relationship between heritage and the community, with its core characteristics defined by continuous use and dynamic conservation. Building on previous studies, this research proposes a five-dimensional cooperative structure to construct a theoretical model for the conservation of rural living heritage buildings, thereby laying the foundation for subsequent systematic research.

(b) Method development and advantages of BERTopic technology

BERTopic is a neural topic modeling method proposed by Maarten Grootendorst in 2022. The model integrates transformer-based semantic embeddings with a class-based

Term Frequency–Inverse Document Frequency (c-TF-IDF) mechanism to generate topic representations with enhanced semantic consistency and interpretability. The model first converts documents into vector representations using a pre-trained BERT or Sentence-BERT model. Dimensionality reduction is then performed using Uniform Manifold Approximation and Projection algorithm (UMAP), followed by density-based clustering methods such as Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN). Finally, the c-TF-IDF method is used to extract representative keywords for each topic cluster (Grootendorst, 2022). Comparative studies have shown that BERTopic outperforms traditional topic modeling methods, such as Latent Dirichlet allocation and Non-negative Matrix Factorization models, in terms of topic consistency, interpretability, and semantic clarity. These advantages are particularly evident in datasets characterized by strong semantic connections, uneven word frequencies, or complex contextual relations (Egger & Yu, 2022; Hellwig *et al.*, 2024). Specifically, applying

BERTopic to more than 13,000 abstracts in the field of natural language processing enabled researchers to identify topic structure, analyze their temporal evolution, and examine patterns of topic emergence and decline (Hellwig *et al.*, 2024). In addition, BERTopic has demonstrated strong performance in short text datasets or data with limited language resources, such as social media or low-resource language texts, where traditional topic models often struggle. With appropriate parameter settings, the model can still generate meaningful and interpretable topics (Medvecki *et al.*, 2024).

In interdisciplinary research domains with complex semantics structures and strong cross-disciplinary features—such as studies on architectural protection and cultural heritage—BERTopic is particularly effective in extracting higher-level thematic patterns from medium or large datasets. These patterns may include topics such as material deterioration mechanisms, community participation pathways, policy implementation mechanisms, and digital restoration technologies. In addition, BERTopic enables the construction of topic evolution trajectories over time. Its strength lies in the ability of transformer-based embeddings to capture the semantic variation across contexts, while c-TF-IDF highlights topic-specific keywords, thereby enhancing both analytical depth and interpretability (Egger & Yu, 2022).

(c) International trends in data-driven heritage research

Recent international scholarship exhibits a clear shift toward computational and multi-scalar quantitative methodologies in rural heritage assessment. Compared with traditional qualitative approaches, contemporary international frameworks increasingly prioritize the integration of machine learning techniques and high-resolution spatial statistics to quantify architectural vulnerability and heritage value (Arias Tapiero *et al.*, 2025). In European and Mediterranean contexts, this methodological transition is characterized by the application of multi-criteria decision analysis and predictive modeling to simulate the long-term degradation of traditional building materials under diverse environmental and socio-economic stressors (Olğün & Karatosun, 2019; Shirvani Dastgerdi & Kheyroddin, 2023).

Furthermore, the growing utilization of big data has redefined the boundaries of heritage monitoring. International studies have increasingly employed social media analytics and mobile signaling data to generate dynamic, quantitative evidence of tourism impacts on rural architectural integrity. This approach supports a transition from static preservation models toward responsive governance frameworks (Huo *et al.*, 2024; T. Nguyen *et al.*, 2025). Such data-driven practices facilitate a more rigorous

identification of the symbiotic relationship between physical structures and their functional environments, effectively bridging empirical observation and systematic theoretical modeling (X. Liu, 2024). In contrast, while domestic Chinese research has made substantial contributions through policy-oriented induction and qualitative case studies, it remains in the nascent stages of adopting large-scale natural language processing and semantic vectorization for framework construction—a gap that the present study seeks to address via the application of BERTopic methodology.

Therefore, this paper introduces topic modeling techniques, particularly BERTopic, to reconstruct a theoretical framework in a data-driven manner. This approach aims to establish a more generalizable and structured analytical model for domestic research, responding to current limitations in structural and reusable methods. At the same time, geographic information systems (GIS) have become a mainstream tool in cultural heritage protection and play an increasingly important role in linking sustainability and heritage research (Huo *et al.*, 2024).

Advanced topic modeling technologies have been increasingly applied to reshape the analytical landscape of architectural and cultural heritage. Methodologically, the hybridization of computational modeling with humanistic inquiry provides a sophisticated lens for re-evaluating heritage significance in regional contexts (Tenzer & Schofield, 2024). The empirical utility of this approach is further evidenced by its ability to synthesize large-scale user-generated content, effectively bridging the gap between visitor perceptions and management features (Huo *et al.*, 2024). Concurrently, the use of longitudinal topic analysis has proven effective in tracing thematic evolution and identifying research hotspots within specialized technical literatures over time (Harth, 2024). Collectively, these diverse applications highlight that models such as BERTopic function not merely as tools for data reduction, but are essential for constructing structured, data-driven theoretical frameworks within the complexities of heritage protection.

1.2.2. Theories related to framework construction

In research on the integration of architectural protection and rural tourism, the construction of a theoretical framework not only supports the structured development of academic inquiry but also provides an important foundation for multi-objective governance and policy translation. In international research, framework construction is gradually evolving into a comprehensive structure that combines multiple dimensions such as

cultural landscape, ecological symbiosis, and community participation. The interwoven relationship among cultural identity, local memory, and social aesthetics forms the core logic of protection strategies, as reflected in the cultural landscape governance traditions in both China and the United Kingdom (Djabarouti & Ren, 2024). Related studies further emphasize that local knowledge, ecological memory, and collective responsibility form the basis of five-dimensional frameworks for biocultural heritage protection (Lindholm & Ekblom, 2019). At the same time, institutional tension and governance complexity within cultural policy contexts also deeply influence the framework construction pathways. In multi-ethnic countries, the interaction between policy design and community mechanisms in heritage management has given rise to protection models that balance governance integration and cultural diversity (Saputra, 2024). In cultural landscapes characterized by high disaster risks and climate uncertainty, resilience-oriented perspectives have been incorporated into the framework structure to strengthen adaptability and institutional flexibility (Shirvani Dastgerdi & Kheyroddin, 2022). The theoretical logic underlying cultural tourism development has also evolved. Variables such as multi-level participation, cultural adaptation, and local carrying capacity are increasingly recognized as core components of framework construction. In China, framework development often relies on the coordination between national strategies and localized practices. However, challenges such as theoretical fragmentation and weak structural cohesion persist. In response, data-driven approaches have emerged as a new direction for framework construction. Topic modeling tools, such as BERTopic, show strong capabilities in structure extraction and pathway identification within architectural heritage research through semantic vectorization and topic distribution analysis (X. Li & Lei, 2021; T. Nguyen *et al.*, 2025). These methods extract key thematic relationships from aggregated literature structures and move beyond the traditional “induction–application” model based on predefined dimensions, thereby providing methodological support for constructing protection frameworks with systematic structure, semantic interpretability, and cross-domain adaptability (Blei *et al.*, 2003).

At the global level, theoretical conceptualizations of rural heritage protection are increasingly oriented toward dynamic, resilience-based frameworks. Contemporary international research emphasizes the quantification of living heritage through integrated structural matrices that encompass biocultural diversity and socio-technical resilience (Lindholm & Ekblom, 2019; Saputra, 2024). This paradigm shift is increasingly supported by information and communication technology-based modeling and GIS

spatial analysis, which provide the empirical granularity necessary to validate multidimensional frameworks (Arias Tapiero *et al.*, 2025; Saputra, 2024). By incorporating quantitative indicators of stakeholder engagement and functional adaptability, these international frameworks offer a more objective basis for assessing the sustainability of traditional villages within the rural tourism context (T. Nguyen *et al.*, 2025; Z. Wu *et al.*, 2025).

Guided by these global trends, this study focuses on constructing an architectural evaluation system for rural tourism contexts and adopts the following two theories as its theoretical foundation.

(a) Theory of building adaptation

This theory of building adaptation, systematically proposed by Douglas (2006), emphasizes that, while respecting the original structure, form, and context of a building, coordination between historical architecture and contemporary functional needs can be achieved through methods such as functional replacement, spatial reorganization, and technical intervention (Douglas, 2006). In the context of historic building renovation and reuse associated with rural tourism development, this theory provides strong practical guidance for balancing revitalization and preservation. In this study, BERTopic modeling is used to identify and extract key dimensions of adaptive intervention from the literature, including material strategies, structural strategies, and functional strategies. These dimensions collectively form the basis of the adaptability component within the architectural evaluation framework.

(b) Rural tourism theory

Rural tourism theory emphasizes the activating role of tourism in rural social, economic, and cultural systems, highlighting the mediating effects of local identity, community participation, environmental carrying capacity, and cultural inheritance in tourism development processes (Lane, 1994; Sharpley, 2002). As a research context, rural tourism requires that architectural evaluation extend beyond physical attributes and be understood within community, cultural, and development-oriented systems. In this study, rural tourism theory informs the application of BERTopic for topic modeling and the extraction of tourism-driven factors from existing literature. These factors are then used to construct architectural evaluation dimensions that reflect social sustainability and cultural adaptability.

The theory of building adaptation provides the basis for constructing the architectural attribute dimensions, while rural tourism theory offers the logical foundation for embedding social and cultural contexts and developing the

sustainability-oriented dimensions. Together, these two theories form the theoretical underpinning of the study, enabling BERTopic modeling results to be systematically translated into structural evaluation dimensions that support the framework's scientific validity, interpretability, and applicability.

1.2.3. The BERTopic method

In terms of research methodology, existing studies predominantly rely on case studies, policy analysis, and field investigations. While these methods enable in-depth protection of individual villages, their findings often lack general applicability and are limited in their ability to synthesize the diverse experiences of architectural conservation. Some studies have attempted to introduce text mining tools, such as co-word analysis and high-frequency keyword extraction; however, these methods are limited in their ability to capture large-scale textual semantics and to reveal latent topic structures and evolution trends across studies.

BERTopic is a topic modeling approach designed to identify latent thematic structures by leveraging semantic embeddings and clustering techniques. Unlike traditional methods that rely primarily on word co-occurrence patterns, BERTopic uses pre-trained Transformer-based models (such as BERT) to capture deep semantic relationships between documents and then applies clustering algorithms (such as HDBSCAN) to extract coherent topics. Compared with traditional qualitative methods, BERTopic offers several key advantages: (i) it can efficiently handle large-scale textual datasets; (ii) it has strong topic clustering ability; and (iii) it supports the analysis of topic evolution over time (Blei *et al.*, 2003). In high-dimensional literature spaces, the model can distinguish nuanced semantic relationships among topics and trace their developmental trajectories, thereby revealing the internal logic and development trends of a research field. Empirical applications demonstrate the effectiveness of BERTopic in uncovering complex thematic structures. By modeling about 16,000 papers, L. Nguyen *et al.* (2024) built a hierarchical structure of research hotspots, revealing the dynamic connections among topics over time. X. Li & Lei (2021) applied a semantic nested structure in their bibliometric analysis, advancing the theoretical systematization of architectural heritage research by tracing the evolution of topics across temporal scales.

1.3. Research gap

Although scholarly attention to rural living heritage and building conservation has increased in recent years, several key limitations remain.

First, most existing studies continue to rely on qualitative methods, such as field surveys and policy analysis, with limited integration of systematic theoretical modeling and quantitative data support.

Second, domestic research has largely focused on individual case analyses and policy-oriented interpretations, with relatively few studies employing semantic or topic evolution methods to demonstrate structural relationships across cases and research domains.

Third, there is a notable lack of automated and objective approaches for generating structured frameworks. Many previous studies struggle to bridge the gap between raw textual data and theoretical dimensions, relying heavily on fragmented qualitative induction. Currently, there is no widely adopted methodology capable of maintaining high semantic fidelity while simultaneously identifying the hierarchical relationships among disparate research topics—an essential requirement for constructing a robust, multidimensional conservation framework.

In response to these gaps, we need a data-based framework to reveal the internal connections among culture, function, and community in living heritage conservation. This study applies the BERTopic model to extract latent topic structures from a large corpus of academic papers and constructs a multidimensional framework for rural living heritage conservation, thereby addressing these methodological and theoretical limitations.

1.4. Research contributions

To address the challenges of insufficient theoretical framework, outdated methodological innovation, and limited empirical approaches, this study proposes the following innovations and contributions:

1.4.1. Theoretical contribution

This study constructs a protection framework for traditional village architecture from a five-dimensional collaborative perspective encompassing cultural authenticity, functional adaptability, and community participation. By revealing the interactive logic and internal mechanisms among these three themes, the study expands the theoretical scope of architectural protection research.

1.4.2. Methodological innovation

By introducing the BERTopic model, this study conducts large-scale text mining and topic evolution analysis based on 602 academic papers. It proposes a quantitative research pathway linking topic evolution, framework construction, and practical guidance, offering a technical alternative to traditional experience-based research methods.

1.4.3. Solid empirical foundation

The study draws upon diverse data sources, including the domestic China National Knowledge Infrastructure (CNKI) database, the Web of Science (WoS) database, and relevant local policy and regulatory texts. This multi-source dataset ensures the representativeness and scientific validity of the proposed analytical framework.

This chapter systematically reviews the research background, relevant theories, prevailing methods, and existing research gaps, thereby clarifying the theoretical foundation and practical orientation of this study and laying a solid basis for subsequent framework construction and empirical analysis.

2. Materials and methodology

This study adopts a combination of qualitative and quantitative methods to analyze the relationship between traditional village architectural conservation and rural tourism development in China. The methodology encompasses data collection, analysis procedures, and the construction of an evaluation index system.

2.1. Materials

The study draws on multiple data sources, including academic literature, policy documents, and local standards, to provide a comprehensive foundation for analysis.

2.1.1. Data sources

The data sources (Figure 1) used in this study are classified into three categories:

(a) China National Knowledge Infrastructure

CNKI provides a wide range of high-quality Chinese-language journal papers related to traditional village protection, architectural heritage management, and rural tourism development. The collected literature covers topics such as regional policy interpretation, building renovation cases, and architectural culture analysis.

(b) Web of Science core collection

The WoS Core Collection serves as the primary source of English-language studies. The search employed keywords such as “traditional villages,” “architectural conservation,” and “rural tourism,” ensuring coverage of global research and enabling comparative analysis with international perspectives. A total of 67 relevant English records were retrieved. To maintain semantic consistency with the Chinese dataset, the titles and abstracts of these records were translated into Chinese and manually verified by two researchers for technical terminology accuracy.

(c) Local laws and regulations:

Local laws and regulations from 11 provinces in China were collected, including building control codes, function-

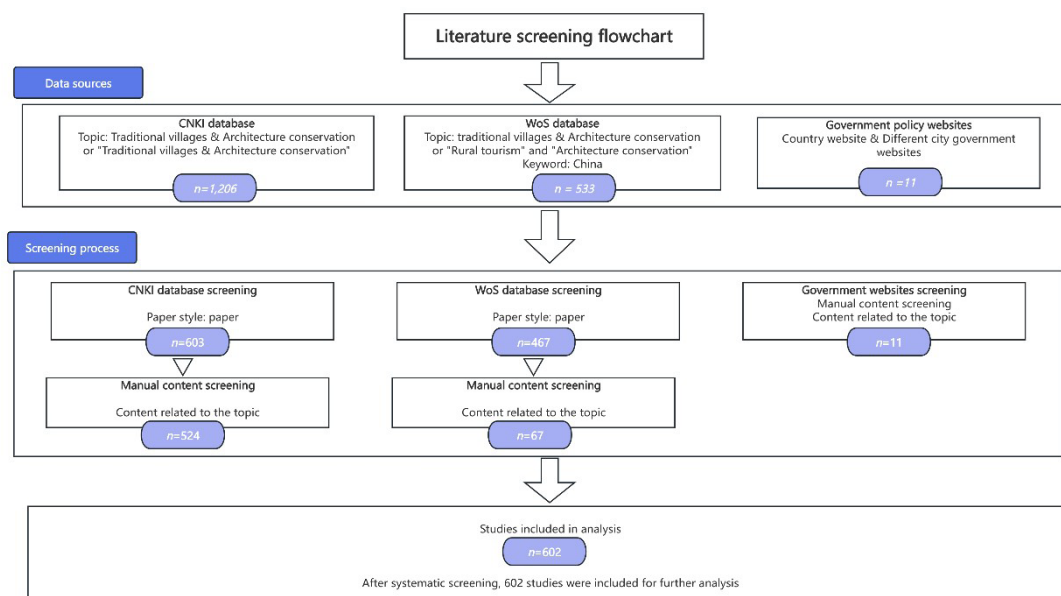


Figure 1. Data sources and data collection
Source: Drawing by the authors

al use standards, and landscape protection policies. These documents provide practical references for analyzing policy pathways and regional differences in the implementation of traditional village architectural protection in China.

2.1.2. Data collection

In the data collection stage (Figure 1), literature was systematically screened and retrieved from both Chinese and English databases using tailored strategies:

For the Chinese sources, keywords such as “traditional village,” “rural tourism,” and “architectural protection” were combined using Boolean operators (AND/OR). The initial search returned 1,206 papers. Limiting the results to journal articles reduced this to 603 records. Abstracts and keywords were then manually reviewed to remove duplicates and irrelevant entries, resulting in 524 valid papers.

For the English sources, the WoS core collection was searched using the keywords including “traditional village(s),” “tourism village,” “building protection,” and “China.” A total of 533 papers were identified, and 67 core papers were retained after screening for relevance and completeness.

Policy documents related to traditional village architectural protection were also collected from the official websites of provincial housing and urban-rural development departments and public policy databases. These include building control codes, functional use standards, and landscape protection regulations.

The final data set comprises 602 documents, including 591 academic papers (524 in Chinese and 67 in English) and 11 policy documents, covering theoretical research, policy analysis, and case-based practices. This multi-source dataset provides a solid foundation for subsequent modeling and analysis.

2.1.3. Data preprocessing

To ensure modeling quality, the collected text underwent systematic cleaning and standard preprocessing, which involved removing HTML tags, deleting numerical values, filtering out stop words, eliminating punctuation, and performing lemmatization. Specifically, for Chinese text, word segmentation was performed using the Jieba library, while for English text, lemmatization was conducted before translation to ensure semantic alignment and consistency.

Formally, the original text collection $D = \{d_1, d_2, \dots, d_N\}$ was transformed into a cleaned collection D' through the preprocessing operator P , expressed as:

$$D' = \{P(di) : i = 1, 2, \dots, N\} \quad (1)$$

This step aims to improve the accuracy of semantic text modeling and remove non-semantic noise from the data.

2.2. Methodology

2.2.1. BERTopic modeling methods

This study employs the BERTopic model to perform semantic modeling on the cleaned text dataset. BERTopic is a hybrid topic modeling algorithm that integrates BERT-based embeddings, dimensionality reduction, density-based clustering, and keyword extraction. Following the approach proposed by Grootendorst (Grootendorst, 2022), text embeddings were generated using Sentence-BERT (Reimers & Gurevych, 2019), dimensionality reduction was applied via UMAP (McInnes *et al.*, 2020), and keywords were extracted using the c-TF-IDF method (Grootendorst, 2022). In addition, a hierarchical structure of topics was constructed using the average linkage hierarchical clustering algorithm, as summarized by Murtagh and Contreras (2011).

To maintain semantic consistency and lexical uniformity across the corpus, the titles and abstracts of the 67 English papers from WoS were translated into Chinese. The translation was performed using Google Translate and subsequently verified by two researchers to ensure technical terminology accuracy. This preprocessing allowed unified Chinese-based c-TF-IDF keyword extraction, effectively preventing the fragmentation of semantic features caused by bilingual terminology.

(a) Text vectorization (BERT embedding)

Each preprocessed text d'_i was input into a pre-trained BERT model (Sentence-BERT) to obtain its semantic vector representation, denoted $\mathbf{e}_i \in \mathbb{R}^d$:

$$\mathbf{e}_i = \text{BERTEmbed}(d'_i)_n \quad (2)$$

This vector preserves the contextual semantic information, ensuring that semantically similar texts are closer in the vector space.

To ensure lexical uniformity, 67 English abstracts from WoS were translated into Chinese and verified by researchers. For high-quality semantic alignment, the paraphrase-multilingual-MiniLM-L12-v2 model was selected as the encoder to ensure a unified Chinese corpus.

(b) Uniform Manifold Approximation and Projection

High-dimensional embedding vectors generated by BERT are not suitable for clustering operations; therefore, they

were mapped into low-dimensional vectors using the UMAP $z_i \in R^k$:

$$z_i = f_{UMAP}(e_i) \quad (3)$$

UMAP preserves the local structure of the high-dimensional space. In this study, the hyperparameters were set as $n_{neighbors} = 15$ and $n_{components} = 5$ to balance the preservation of local and global semantic structures.

(c) Clustering (Hierarchical Density-Based Spatial Clustering of Applications with Noise)

In the low-dimensional space, the HDBSCAN algorithm was applied for clustering. The algorithm forms clusters based on local density and automatically identifies outliers (noise). The clustering labels can be expressed as follows:

$$C(i) = \text{cluster_id or noise} \quad (4)$$

HDBSCAN is more suitable than k-means for text clustering with irregularly shaped topic distributions. Based on the corpus scale, the core hyperparameters were set as $min_cluster_size = 10$ and $min_samples = 5$. Through this density-based clustering, nine initial topic clusters were automatically identified, providing a data-driven foundation for subsequent framework construction.

(d) Keyword extraction (class-based Term Frequency-Inverse Document Frequency)

To capture the semantic features of each cluster, CountVectorizer was first applied with $ngram_range = (1, 2)$ to capture both unigrams and bigrams, and a $min_df = 5$ to filter out low-frequency noise words. c-TF-IDF was then used to calculate the keywords within each cluster. The calculation formula is as follows:

$$cTFIDF(t, c) = \frac{tf(t, c)}{\sum_t tf(t', c)} \times \log \frac{N}{df(t)} \quad (5)$$

Here, $tf(t, c)$ represents the frequency of the term t in cluster c , $df(t)$ represents the number of clusters in which the term t appears, and N is the total number of clusters. In the end, each topic cluster can be represented by its keyword semantics.

To enhance theoretical interpretability and reduce subjectivity, the nine initial clusters were synthesized into five core evaluation dimensions. The merging followed three criteria: (i) Statistical similarity: Hierarchical clustering was performed, and clusters with a cosine similarity > 0.85 were considered for merging; (ii) Semantic redundancy: Clusters with overlapping c-TF-IDF keywords were integrated; (iii) Theoretical alignment: Clusters were mapped onto established architectural conservation pillars

to ensure the framework's systematicity.

(e) Topic merging and evolution analysis

If initial clusters were overly detailed or semantically repetitive, BERTopic's topic merging mechanism was used, which reclusters c-TF-IDF representation vectors between clusters to merge similar topics. In addition, using document publication dates, the rise and decline of topics were also tracked, enabling a preliminary topic evolution analysis.

2.2.2. Hierarchical clustering analysis

To explore the internal relationships among topics from a structural perspective, this study conducted hierarchical clustering analysis on the topic vectors or TF-IDF representations generated by BERTopic to construct a topic hierarchy map. A bottom-up agglomerative clustering method was adopted, using the average linkage approach to calculate the distance between clusters.

$$D(r, s) = \frac{1}{|r| \cdot |s|} \sum_{i \in r} \sum_{j \in s} d(i, j) \quad (6)$$

Here, $d(i, j)$ represents the distance between two topic vectors, which can be calculated using either cosine distance or Euclidean distance. This method effectively reveals the aggregation structure and classification basis among topics, providing a foundation for constructing the subsequent evaluation dimension framework.

The hierarchical clustering results (Dendrogram) were analyzed to evaluate the semantic affinity among the seven initial clusters. By applying a cosine similarity threshold of 0.85, we observed that certain topics exhibited high thematic overlap. Consequently, these seven clusters were synthesized into five primary evaluation dimensions. This step bridges the gap between raw data-driven clustering and structured theoretical induction, ensuring the framework's conciseness and interpretability.

2.2.3. Visualization and framework construction

Based on the BERTopic clustering results and the hierarchical clustering structure, representative topic categories and their keyword semantic labels were extracted. Visualization was carried out through word clouds, topic distribution maps, evolution trend charts, and dendrograms. These visualizations facilitate the summarization of structured architectural evaluation dimensions, specifically building style, building material, building function, building culture, and social culture. Following further semantic categorization and structural integration, these topic dimensions are consolidated into an architectural evaluation framework grounded in

empirical text modeling.

To verify the robustness of the framework, a sensitivity analysis was conducted by varying the `min_cluster_size` parameter from 5 to 20. The consistency of the five core dimensions across different parameter settings confirms the robustness and reliability of the proposed evaluation framework.

3. Results

Based on semantic mining and topic modeling of 602 Chinese- and English-language publications, this section systematically analyzes the evolution trends, research hotspots, and theoretical framework construction in studies of traditional village architectural protection. The analysis focuses on the distribution characteristics of the literature, topic number optimization, coherence and perplexity evaluation, the research hotspot evolution, and the construction of the final five-dimensional protection framework.

3.1. Distribution characteristics of literature

To illustrate the developmental trajectory of research on traditional Chinese villages, the temporal distribution of the 602 publications (Figure 2) was divided into three distinct stages:

(i) Initial development and policy launch stage (1999–

2013),

(ii) Policy-driven growth and rapid expansion stage (2014–2019), and

(iii) Diversified integration and deepened development stage (2020–2024).

(a) Stage 1: Initial development and policy launch (1999–2013)

This stage represents the early accumulation of research on traditional Chinese villages. During this period, publication outputs were relatively limited, with studies primarily focused on the identification of cultural spaces and the documentation of vernacular architecture. The formal inception of national protection efforts was marked by the Notice on Carrying out the Investigation of Traditional Villages (MOHURD *et al.*, 2012) and the subsequent Guiding Opinions on the Protection and Development of Traditional Villages (MOHURD, 2013). These policy milestones provided critical policy momentum, which is empirically reflected in the clear inflection point in Figure 2, where publication volume began to rise markedly after 2013.

(b) Stage 2: Policy-driven growth and rapid expansion (2014–2019)

Following 2014, a rapid increase in scholarly output was observed, closely aligned with the implementation of na-

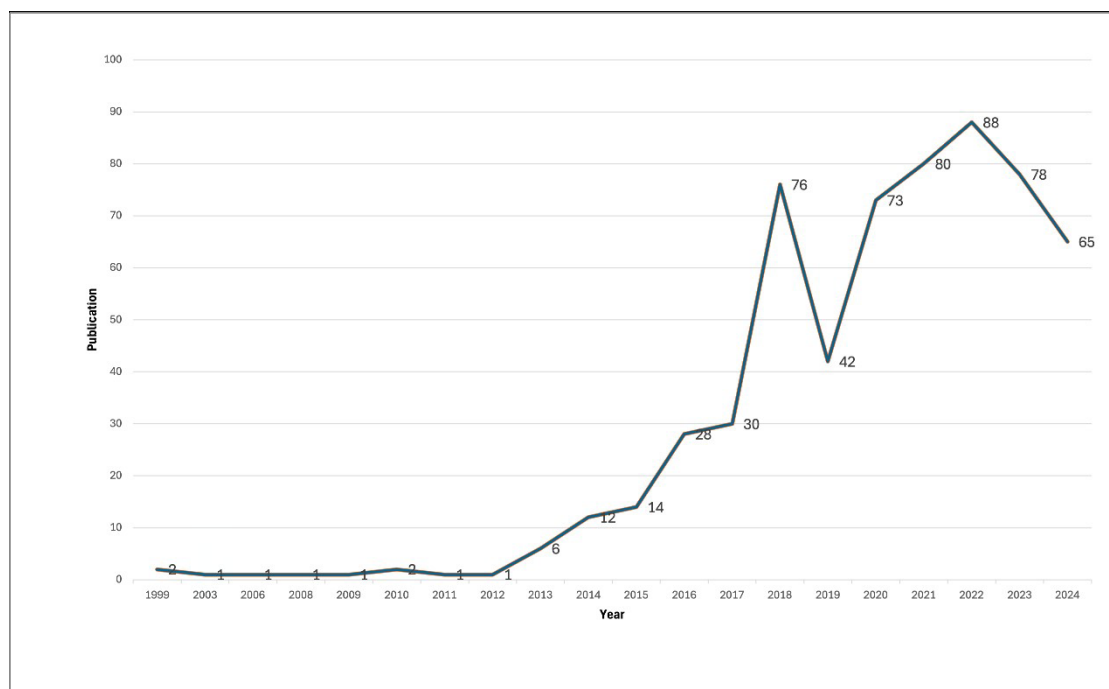


Figure 2. Number of publications of different years
Source: Graphs by the authors

tional initiatives such as the Rural Revitalization Strategy. During this stage, research themes expanded beyond architectural protection to include community participation, cultural reconstruction, and tourism-driven development.

(c) Stage 3: Diversified integration and deepened development (2020–2024)

Since 2020, driven by policy initiatives such as the *Rural Construction Action Plan* and *Digital Countryside Strategy*, research has entered a phase of integration and diversification. The integration of digital technologies such as GIS, building information modeling, and virtual reality (VR) has become increasingly prominent in heritage conservation, signaling a methodological transition toward precision documentation and digital preservation (B. Liu *et al.*, 2024).

3.2. Topic modeling optimization process and results

3.2.1. Model optimization and parameter validation

To ensure the transparency and reliability of topic extraction, a systematic model optimization and parameter validation process was conducted (Table 2).

First, the optimal number of topics was determined using the topic coherence score (C_v). A range of topic numbers ($k = 5\text{--}15$) was tested by adjusting the `nr_topics` parameter in BERTopic. As shown in Table 2, the coherence score reached its maximum at $k = 9$ ($C_v = 0.61$), indicating that this topic granularity yields the highest semantic consistency. When k exceeded 12, the coherence score declined, reflecting increased topic fragmentation and redundancy.

Second, to address the stochastic nature of the UMAP dimensionality reduction algorithm, a stability test was performed by running the model with five different random seeds. The results showed that, although minor variations were observed in the spatial positioning of topics within the Intertopic Distance Map, the core keyword clusters and overall thematic structure remained highly stable. Finally, a sensitivity analysis was conducted on the `min_topic_size` parameter, with values ranging from 10 to 50. The results confirmed that the five-dimensional framework synthesized in Section 3.3 is robust and insensitive to minor parameter variations, further supporting the reliability of the modeling outcomes.

3.2.2. Topic extraction and distribution overview

Using the BERTopic model, nine principal topics were identified from the corpus of 602 publications addressing traditional village architecture and rural tourism. These topics encompass diverse research perspectives, including architectural protection, cultural inheritance, rural spatial transformation, and technological intervention. The distribution of topics and their corresponding keyword weights are summarized in Table 3.

Topic 0 is centered on the keywords village, tradition, protection, and architecture. This topic reflects research emphasizing the preservation and spatial revitalization of traditional settlements within a broader rural development framework. Existing studies demonstrate that architectural conservation is closely tied to cultural recognition and serves as a foundation for revitalizing rural spatial identity (Feng & Xiao, 2020; Yao, 2016).

Topic 1 focuses on architecture, culture, village, and

Table 2. Performance comparison of different topic numbers (k)

Number of topics (k)	Coherence score (C_v)	Average similarity	Description
5	0.49	0.35	Insufficient granularity; loss of technical detail.
7	0.53	0.39	Merges social and cultural themes.
9 (Optimal)	0.61	0.42	Best semantic distinctness and interpretability.
12	0.58	0.52	Topic fragmentation starts to occur.
15	0.54	0.65	Significant semantic overlap between topics.

tradition, emphasizing architecture as a medium for expressing cultural values. The literature shows that architectural form serves as a carrier through which cultural memory and regional identity are expressed, particularly in the renovation of Hui-style historical villages that integrate traditional aesthetics with contemporary functional requirements (H. Hu *et al.*, 2021; L. Wang, 2021).

Topic 2, represented by heritage, building, architectural, and village, reflects a growing interest in the conceptualization of heritage architecture. Research within this topic examines interrelationships between heritage conservation, architectural typology, and local identity, which collectively shape the spatial logic of rural settlements (Lindholm & Ekblom, 2019; Qianda *et al.*, 2021).

Topic 3 addresses the interplay among protection, culture, and tradition, underscoring the social construction of heritage through community engagement and cultural continuity. This body of literature emphasizes how protection practices intersect with local social structures and daily life.

Topic 4, characterized by keywords such as Drum tower, Dali, and Xili, explores ethnic and regional architecture as cultural symbols within tourism-oriented heritage landscapes. These distinctive buildings act as visual markers of ethnic identity and as spatial representations of collective memory (González Martínez, 2019; J. Zhou, 2021).

Topic 5 includes relics, Yaodong, and tradition, focusing on technical conservation and adaptive reuse practices. Research highlights the integration of traditional craftsmanship with contemporary construction techniques, particularly in regions such as the Loess Plateau, where vernacular forms like Yaodong dwellings are preserved and reinterpreted for modern use (Y. Wang *et al.*, 2021; X. Zhang, 2015).

Topic 6 emphasizes technology, digitalization, and models, representing the increasing adoption of digital tools in heritage research. The application of GIS, BIM, and three-dimensional (3D) modeling technologies has enhanced the precision of architectural documentation and facilitated new modes of visualization, simulation, and interpretation (B. Liu *et al.*, 2024).

Topic 7 is dominated by heritage, history, and an ancient village, focusing on the historical evolution of heritage protection frameworks and the transformation of institutional approaches to conservation. This topic reflects a continuous refinement of multidimensional protection systems (Miao & Chiou, 2013; Tenzer & Schofield, 2024).

Topic 8 combines culture, protection, and tradition, partially overlapping with Topics 0 and 3. It emphasizes the symbolic reconstruction of cultural meaning and local identity through lived practices and community narratives, particularly in minority and remote mountain villages where cultural landscapes remain central to heritage continuity (Peng *et al.*, 2021; P. Tang *et al.*, 2019).

3.2.3. Topic trend and weight analysis

The topic trend chart (Figure 3) generated by BERTopic illustrates temporal changes in the prevalence of each topic across the study period (Grootendorst, 2022). The horizontal axis represents publication year, while the vertical axis indicates the proportion of documents assigned to each topic. Each colored area corresponds to a topic. The width of each colored area reflects the relative prominence of a topic in a given year. This visualization helps illustrate the temporal evolution of research focus within the corpus.

Topic 0, characterized by village–tradition–conservation–architecture, shows a steady increase in frequency since 2011 and remains the dominant theme in the field. Topic 1, defined by architecture, tradition, culture, and village, follows a similar growth trajectory. Both themes maintain a strong association with architectural heritage conservation and the spatial continuity of traditional villages (Gao & Wu, 2017; C. Tang *et al.*, 2023; J. Zhou, 2021).

After 2011, several clustered topics show accelerated growth that aligns with national policy milestones, particularly the release of the Traditional Villages List in 2012 and the Guidelines for the Protection and Development of Traditional Villages in 2013.

From 2015 onward, new themes related to culture, village, and heritage expanded rapidly, indicating increasing thematic diversification.

Topics associated with digital technologies display the most pronounced upward trend. The increased prevalence of themes such as architecture, village, technology, and digitalization demonstrates the expanding use of GIS, BIM, and 3D modeling in rural heritage research (Harth, 2024; Medveckí *et al.*, 2024; Zhu *et al.*, 2022).

The topic weight analysis (Figure 4) indicates that Topic 0 and Topic 1 together account for more than half of the dataset, confirming their dominant influence on the field's overall structure. Although the topics such as theme heritage, building, architectural, and village occupy a smaller proportion, they serve as an important link between architectural conservation and cultural heritage studies (Cattaneo *et al.*, 2018; Medveckí *et al.*, 2024).

While topics related to history, conservation, and memory (Topic 5 and Topic 7) appear less frequently, they remain essential for understanding the socio-cultural dimensions of architectural protection (Barber *et al.*, 2006; Douglas, 2006; Lane, 1994). Overall, the results demonstrate that traditional architectural preservation remains the core research focus, while emerging topics show increasing diversification and methodological expansion.

The topic weight bar chart (Figure 4) visualizes the relative importance of each topic within the corpus. The horizontal axis represents the topics, and the vertical axis shows their weights, calculated based on the number of documents assigned to each topic. Higher bars indicate dominant or frequently discussed themes, while lower bars represent less prominent or emerging topics.

The topic weight bar chart (Figure 4) indicates that Topic 0 and Topic 1 together account for more than half of the total dataset, confirming their dominant influence on the structural composition of the field (Harth, 2024). The theme heritage–building–architectural–village occupies a comparatively smaller proportion, but functions as an important conceptual bridge between architectural conservation and cultural studies (Cattaneo *et al.*, 2018; Medvecki *et al.*, 2024).

Although topics related to history–conservation–memory (Topic 5 and Topic 7) appear less frequently, they remain highly relevant for exploring the socio-cultural dimensions of architectural protection (Barber *et al.*, 2006; Douglas, 2006; Lane, 1994). Overall, the dataset demonstrates that traditional architectural preservation continues to dominate the research landscape, while emerging topics reflect increasing thematic diversification and methodological expansion.

3.3. Topic structure and similarity analysis

The BERTopic model initially extracted nine distinct topics, whose internal logical relationships and structural hierarchy were subsequently explored through a similarity heatmap (Figure 5) and a hierarchical clustering tree (Figure 6). These visualizations provided a quantitative foundation for synthesizing the data-driven clusters into a structured evaluation framework.

The transformation from empirical topic clusters to normative dimensions was governed by a multi-stage validation process that integrated quantitative metrics with qualitative researcher review. Candidate clusters were first identified based on dendrogram proximity and cosine similarity scores, typically exceeding 0.75. To ensure theoretical rigor, the research team then conducted a semantic audit of the top keywords and their corresponding weights, as summarized in Table 4. This involved cross-referencing automated clusters with established architectural conservation standards to ensure that the final categories remained academically relevant. For instance, topics with high semantic overlap regarding physical components (such as “Yaodong” and “Material” in Topics 5 and 8) or technological tools (such as “Digitization” in Topic 6) were synthesized into their respective normative dimensions. Ultimately, this iterative refinement led to the establishment of the five core evaluation dimensions: Building style, Building material, Building function, Building culture, and Social culture.

3.3.1. Topic similarity heatmap analysis

The topic similarity heatmap (Figure 5) illustrates the semantic relationships among BERTopic-derived topics. The heatmap was constructed using cosine similarity

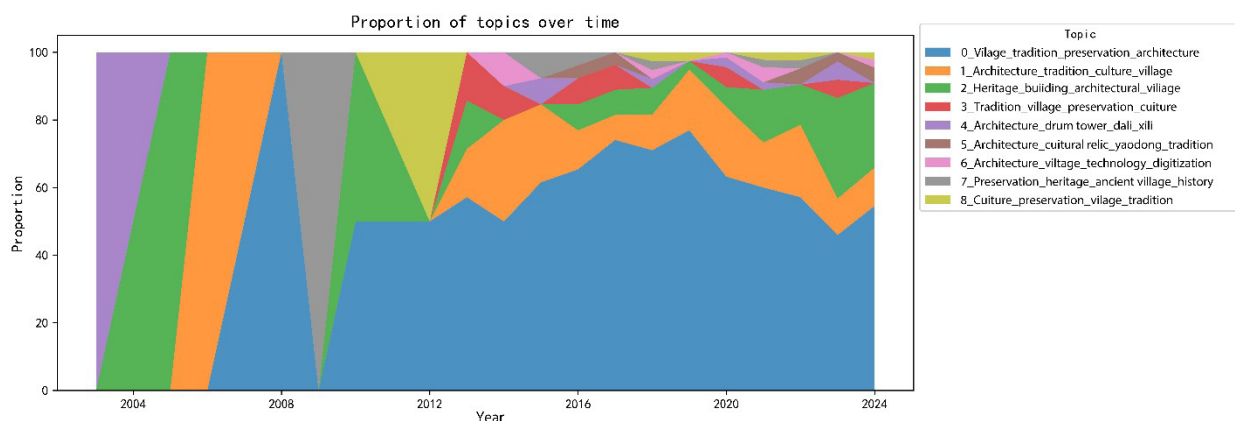


Figure 3. Topic trend chart
Source: Chart by the authors

Table 3. Topic word distribution

Topic0	T0-weight	Topic1	T1-weight	Topic2	T2-weight	Topic3	T3-weight	Topic 4	T4-weight	Topic5-word	T5-weight	Topic6-word	T6-weight	Topic7-word	T7-weight	Topic8-word	T8-weight
Village	0.062	Architecture	0.047	Heritage	0.05	Tradition	0.071	Architecture	0.050	Architecture	0.071	Architecture	0.107	Preservation	0.090	Culture	0.128
Tradition	0.062	Tradition	0.044	Building	0.04	Village	0.067	Drum tower	0.040	Cultural relic	0.054	Village	0.082	Heritage	0.084	Preservation	0.089
Preservation	0.051	Culture	0.044	Architectural	0.03	Preservation	0.051	Dali	0.037	Yaodong	0.053	Technology	0.080	Ancient village	0.079	Village	0.075
Architecture	0.047	Village	0.039	Village	0.03	Culture	0.040	Xili	0.036	Tradition	0.052	Digitization	0.071	History	0.061	Tradition	0.068
Culture	0.038	Preservation	0.033	Cultural	0.03	Symbiosis	0.027	Dajiacang	0.034	Preservation	0.045	Tradition	0.062	Value	0.059	History	0.066
History	0.026	Space	0.031	Traditional	0.03	Revitalization	0.026	Preservation	0.032	Loess	0.042	Preservation	0.060	Culture	0.045	Inheritance	0.051
Residential housing	0.021	Landscape	0.028	Area	0.03	Architecture	0.026	Villager	0.030	Model	0.034	Model	0.060	Theory	0.043	Cultural heritage	0.043
Space	0.021	History	0.027	Value	0.03	Southern Fujian	0.025	Village	0.027	Implementation	0.032	Data	0.055	Village	0.043	Planning	0.039
Implementation	0.020	Environment	0.021	Architecture	0.03	Space	0.024	Guailu	0.025	Village	0.031	Xian village	0.050	Tradition	0.043	Space	0.029
Village	0.020	Implementation	0.021	China	0.03	Planning	0.023	Implementation	0.024	Style	0.026	Space	0.045	Society	0.031	Material	0.027



Figure 4. Topic weight bar chart
Source: Chart by the authors

between topic embeddings, with lighter colors indicating higher similarity values. This visualization facilitates the identification of conceptually related topic clusters and provides empirical support for subsequent grouping or dimensional analysis.

The topic similarity heatmap (Figure 5) was generated using cosine similarity between the embedding vectors of each topic, where higher similarity scores indicate closer semantic relationships. The results show several distinct clusters with strong internal connections. Topics 2, 4, and 5 show similarity values above 0.62, forming a distinct cluster. Keywords such as “architecture,” “drum tower,” and “Dali” suggest a shared focus on building form, style, and visual representation. Another cluster, which includes Topics 0, 7, and 8, also demonstrates strong semantic alignment. In this cluster, Topic 0 and Topic 7 reach a similarity score of 0.66, and Topic 8 remains close to both (≈ 0.58). These topics share keywords such as “culture,” “heritage,” “conservation,” and “history,” indicating strong connections in cultural and heritage-related aspects. Topics 1, 3, and 8 also show above-average similarity values, though weaker than the first two clusters. Their keywords—“villagers,” “activities,” “tourism,” and

“space”—suggest overlapping interests in social activities and spatial use. In contrast, Topic 6 demonstrates very low similarity with all other topics (<0.35 with Topics 0 and 1). Its keywords, “construction,” “material,” “green building,” and “renewal,” reflect a strong focus on technical processes and material systems.

3.3.2. The topic hierarchical clustering

The hierarchical clustering diagram (Figure 6) visualizes the semantic relationships among topics by grouping them according to the similarity of their embeddings. Topics that merge at lower levels in the dendrogram share higher semantic similarity, whereas those merging later are more distinct. This visualization aids in identifying clusters of related topics and supports the structural interpretation of thematic dimensions.

The hierarchical clustering diagram (Figure 6) illustrates the semantic proximity among topics based on their embedding distances. Topics that merge earlier in the dendrogram share higher semantic similarity, whereas those merging later are more distinct.

The results show that Topics 2, 4, and 5 form the lowest-level cluster, indicating a strong semantic relationship.

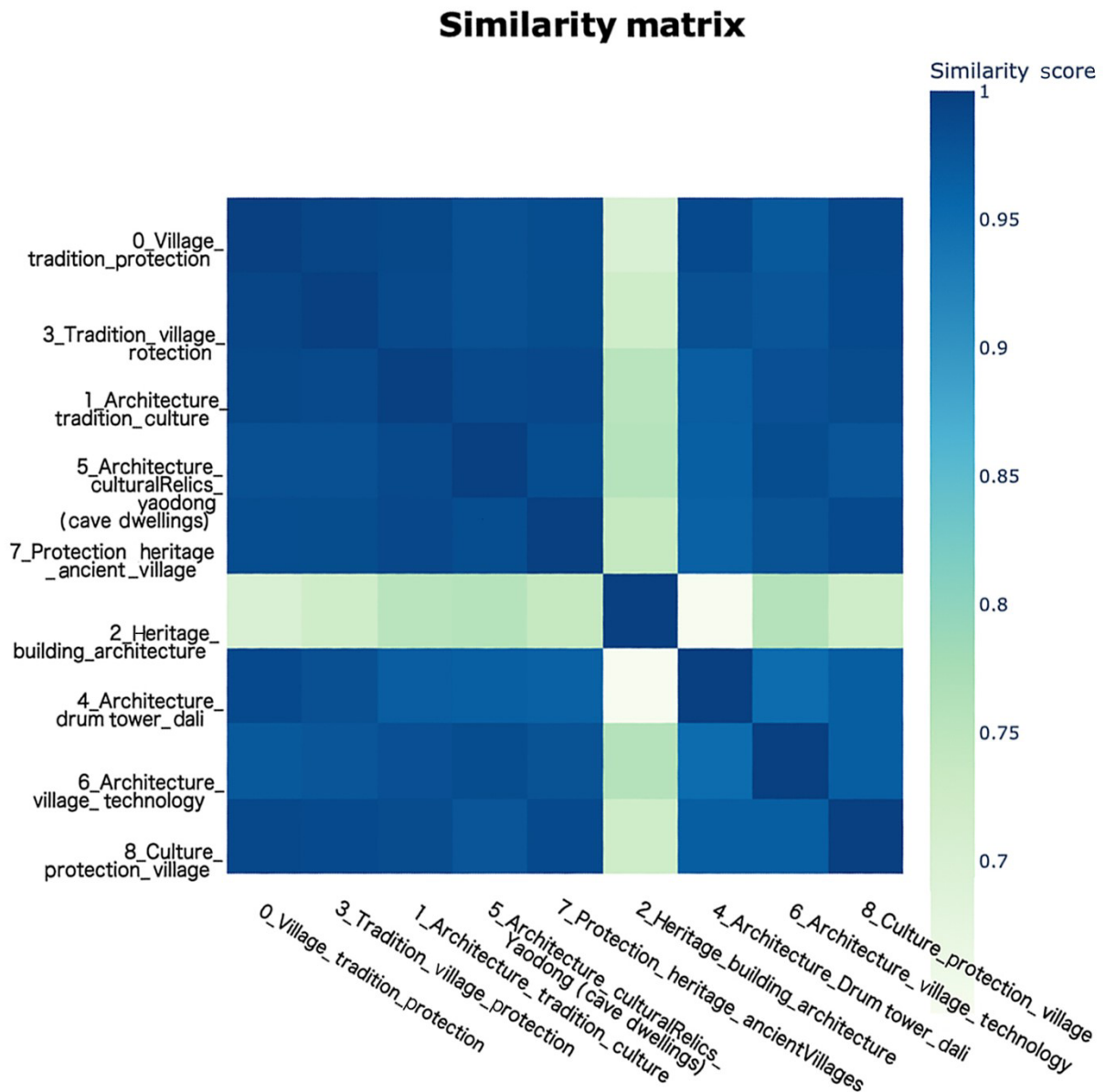


Figure 5. Topic similarity heatmap
Source: Heatmap by the authors

Topics 1, 3, and 8 occupy the same branch but with moderate distances, suggesting partial similarity and interrelation. Topics 0, 7, and 8 constitute another major cluster, with particularly close semantic connections between Topics 0 and 7. In contrast, Topic 6 remains isolated until the final merging stage, reflecting its high degree of independence from the other clusters.

3.3.3. Topic merging and grouping

Based on a threefold verification analysis using the semantic similarity heatmap (Figure 5), hierarchical clustering

diagram (Figure 6), and the spatial distribution results in the Intertopic Distance Map (Figure 7), the nine original topics extracted by BERTopic display a clear structural coupling rather than an even distribution. The Intertopic Distance Map (Figure 7) visualizes the spatial relationships among topics based on their semantic distances in a two-dimensional plane. Each bubble represents a topic, and the distance between bubbles reflects semantic similarity. Larger bubbles indicate topics with higher prevalence or greater weight within the corpus. This visualization aids in identifying overlapping themes and distinguishing

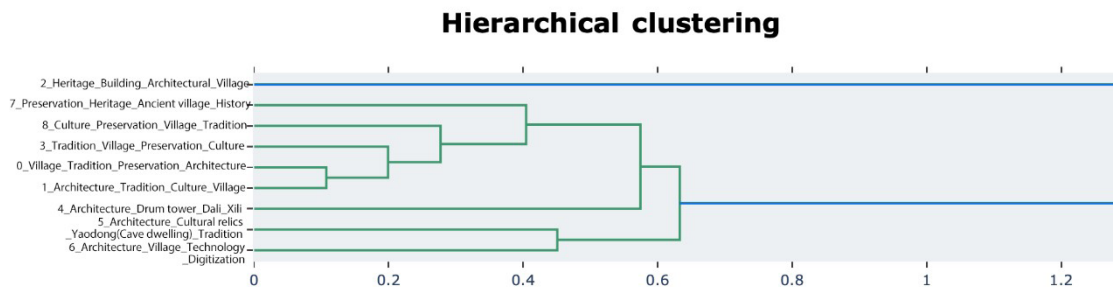


Figure 6. Hierarchical clustering diagram
Source: Diagram by the authors

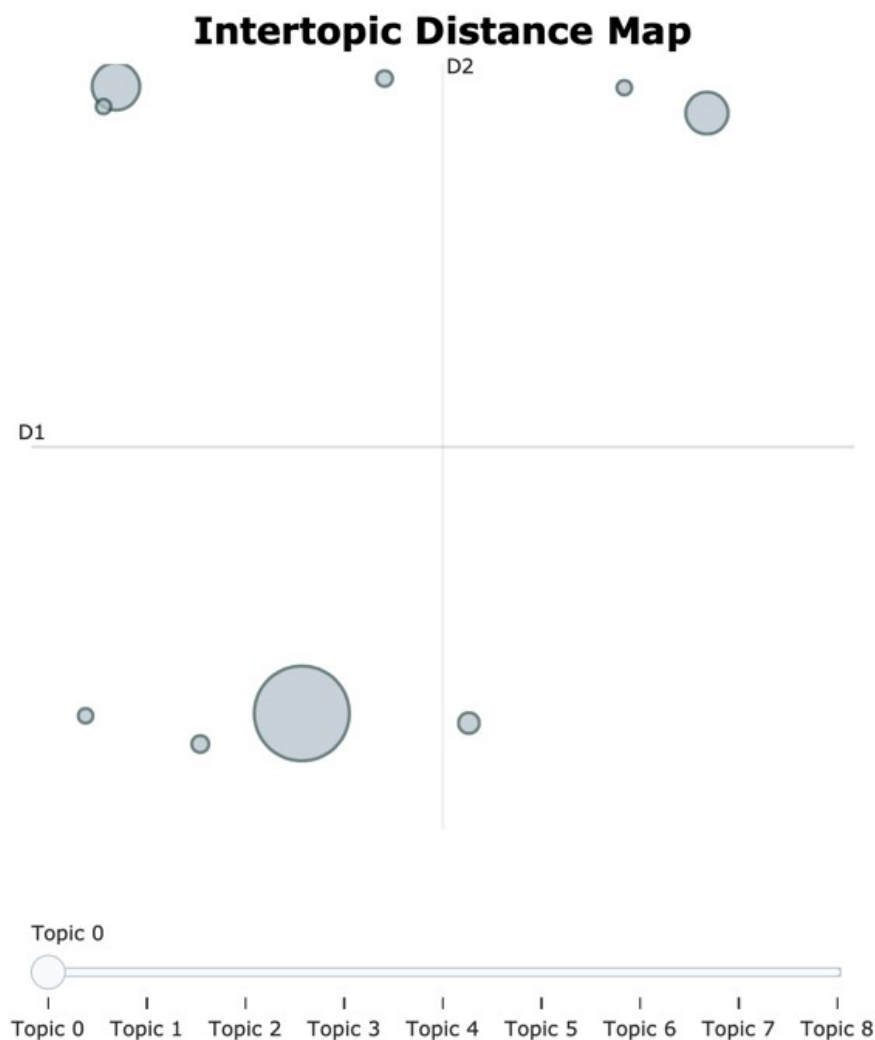


Figure 7. Intertopic Distance Map
Source: Map by the authors

independent topic clusters.

The map shows that several topics exhibit strong semantic and spatial convergence. For instance, Topics 2, 4, and 5 are tightly aggregated in both semantic and spatial distributions, emphasizing their shared focus on architectural typology, stylistic forms, and visual recognition.

These topics are thus classified under the Building Style dimension. Topic 6 demonstrates strong independence across all analytical layers, with a focus on construction processes, material selection, and sustainable renovation practices, and is designated as a separate Building Material dimension. Topics 1, 3, and 8 involve diverse aspects of spatial use, tourism functions, and community interaction. Despite partial dispersion in the embedding space, their evident clustering trends collectively form the Building Function dimension. Topics 0, 7, and 8 form a coherent cluster centered on cultural values, historical memory, and symbolic meaning, defining the Building Culture dimension. Finally, Topics 1 and 3 exhibit overlap in both social and cultural attributes, highlighting community participation, identity construction, and social communication. Accordingly, a fifth dimension, Social Culture, is proposed to supplement the cultural structure with a social-interactional perspective.

3.3.4. Topic evolution and changing trends

The topic evolution analysis illustrates how research themes in rural architecture have shifted in terms of activity, growth rate, and structural development over time. Using the `topics_over_time` module in BERTopic (Figure 8), temporal patterns were analyzed to identify the progression of research focuses. Based on document timestamps, the overall development can be divided into

three stages: (i) the initial and policy launch stage (1999–2013), (ii) the policy-driven and rapid growth stage (2014–2019), and (iii) the diversified integration and deepening development stage (2020–2024). This division aligns with major milestones in China's rural construction policies and reflects the evolving discourse on rural planning (H. Wang *et al.*, 2022).

In the initial and policy launch stage, Topic 0 (Culture and Heritage) and Topic 2 (Traditional Architectural Style) dominate. Research during this period primarily addressed cultural preservation, documentation of physical forms, and authenticity of architectural styles. This reflects a static protection model aligned with the institutionalization of the List of Traditional Villages and early policy frameworks.

During the policy-driven and rapid growth stage, Topic 3 (Spatial Function Adaptation), Topic 5 (Architectural Form), and Topic 6 (Green Construction Technology) rise sharply, with Topic 6 emerging for the first time. The research focus shifts from descriptive cultural documentation to spatial utilization, ecological design, and construction process optimization. This period marks the integration of rural tourism and architectural protection, as traditional buildings began to serve productive, social, and touristic functions (X. Zhang, 2015).

In the diversified integration and deepening development stage, Topic 7 (Digitalization and Technological Translation) grows rapidly, becoming the most dynamic theme. It centers on digital modeling, cultural scene reconstruction, and VR-based participatory design, signifying the increasing role of digital technology in architectural studies. Simultaneously, Topic 1 (Community Participation and Tourism Function) and Topic 3 maintain growth, highlighting the enduring relevance of social embeddedness and everyday spatial use

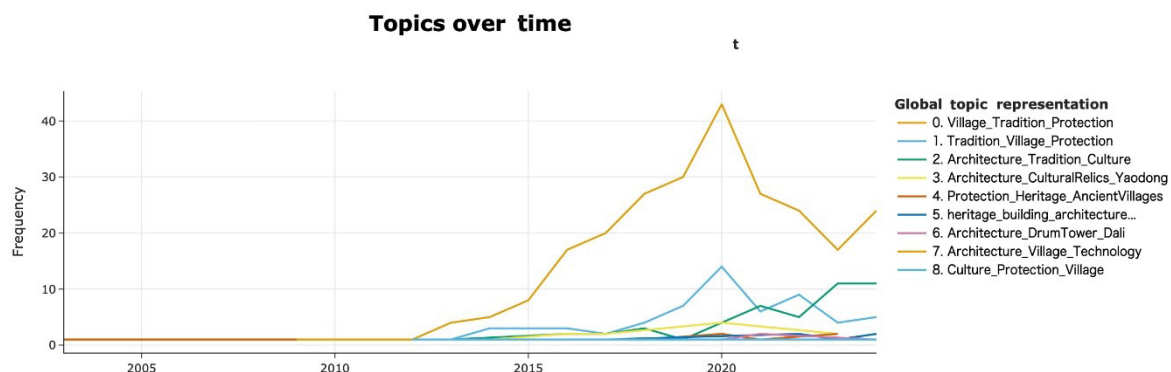


Figure 8. Topics over time
Source: Graph by the authors

(Huang *et al.*, 2024).

Comparative analysis across stages reveals distinct thematic trajectories. Topics 0 and 2 decline after initial prominence, Topics 3 and 5 show stable expansion, and Topics 6 and 7 emerge as fast-growing research directions. These patterns suggest a clear evolution from traditional conservation toward digital, participatory, and socially engaged paradigms (Yang *et al.*, 2021).

The Sankey diagram (Figure 9) further illustrates cross-dimensional linkages. Topics 2 and 4 merge into Topic 5 during later integration stages, forming the unified Architectural Style dimension. Meanwhile, portions of Topic 0 transition into Topic 8 (Cultural Communication and Symbolism), reflecting a shift from static heritage representation toward dynamic cultural interaction.

3.4. Composition of the protection framework for traditional village buildings

To address challenges in the conservation of traditional village buildings, this study uses BERTopic topic evolution analysis covering 1999–2024. Drawing on building adaptability theory and rural tourism theory, a five-dimensional protection framework was developed.

The five-dimensional framework developed in this study integrates architectural characteristics, materials and craftsmanship, spatial and functional adaptation,

technological support, environmental landscape, community and cultural interaction, and architectural cultural value. These dimensions are derived from high-frequency, evolving topics in the literature and are grounded in architectural adaptation theory (Douglas, 2006; Yung *et al.*, 2017), which emphasizes hierarchical renewal and sustainable adaptation. They also draw on rural tourism theory (Su *et al.*, 2018; M. Wu *et al.*, 2022), which highlights a dual focus on cultural authenticity and functional innovation.

By combining topic evolution analysis, theoretical integration, and practical framework construction, this study bridges the gap between large-scale text mining and conceptual development, resulting in a protection framework with both academic depth and applied relevance.

4. Discussion

The overall discussion integrates the analytical results of topic evolution, clustering, and framework construction to interpret the structural logic, methodological transformation, and theoretical implications of traditional village architectural protection research. Through BERTopic-based modeling, this study identifies a multidimensional knowledge framework that reflects shifts in research orientation from preservation to revitalization, and from material to cultural-technological integration.

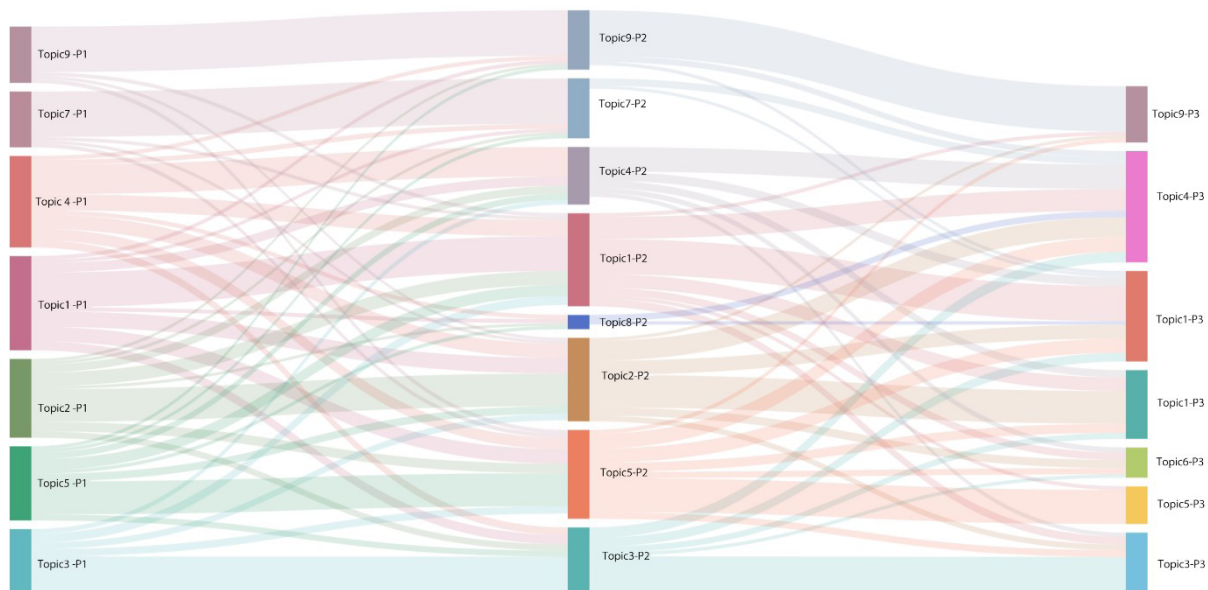


Figure 9. Sankey diagram
Source: Diagram by the authors

Table 4. Protection Framework for Traditional Village Buildings

Dimension	Sub-dimension	Periods	Rationality explanation
Building features	Construction time, integrity of appearance, style diversity	1999–2013; 2014–2019	Captures the historical and cultural value of buildings and provides a baseline criterion for prioritizing conservation (C. Hu & Gong, 2017)
Materials and construction	Traditional materials, green material updates, local craft preservation	1999–2013; 2014–2019	Shows how traditional craft continues and how builders take care of the environment. It also fits green and local construction needs (Feng & Xiao, 2020)
Space and functional adaptability	Space use, mixed functions, modern facility compatibility	2020–2024	Shows how buildings can meet new needs. Helps protection and tourism work together (L. Zhou <i>et al.</i> , 2021)
Building culture	Cultural storytelling, symbolic meaning, traditional identity	2020–2024	Highlights cultural meaning and local identity. This is an important part of continuing village history (Z. Zhang <i>et al.</i> , 2023)
Social culture	Community participation, governance methods, local identity building	2014–2019; 2020–2024	Focuses on the strength within the community. This helps make protection work last longer (Cao <i>et al.</i> , 2024)

4.1. From static preservation to dynamic regeneration

The evolution of research themes indicates a gradual transition from the static preservation of built heritage to dynamic regeneration processes. Early studies centered on authenticity and the conservation of tangible forms, aligning with values of cultural integrity and stylistic restoration (Gao & Wu, 2017). With the expansion of the “architecture, culture, protection” triad, attention broadened to include adaptive reuse and local identity construction. This shift reflects an increasing awareness of architecture as a living cultural space rather than a fixed artifact, emphasizing the role of tradition in ongoing spatial transformation (Feng & Xiao, 2020).

During the intermediate phase, topics such as spatial function adaptation and architectural form emerged, highlighting a shift towards design intervention and ecological integration. Concepts of “heritage architecture” and “locality” became analytical foundations for understanding rural settlement evolution (Qianda *et al.*, 2021). These developments suggest that architectural preservation began to align with sustainability and rural revitalization agendas, bridging the gap between conservation and productive use (X. Zhang, 2015).

4.2. Integration of digital technology and heritage research

The digital transformation of heritage studies extends beyond methodological innovation and has brought substantial changes to conservation standards and management practices. VR and 3D visualization technologies enable immersive reconstruction and simulation, improving conservation professional training, supporting public education, and enhancing cultural communication (Oumoumen, 2024). The application of GIS in heritage research supports spatial distribution analysis, real-time condition monitoring, and environmental risk assessment. In particular, GIS-based models have shown significant value in cultural tourism management and data-driven decision support (Orozco Carpio & Rolim, 2024). Historic Building Information Modelling (HBIM) integrates structural, material, and historical information into a unified digital environment, offering systematic support for restoration planning, structural diagnosis, condition assessment, and archival documentation (López-González & García-Valdecabres, 2023; Siewczyński & Jan Szot, 2025).

The convergence of these digital technologies promotes the standardization and refinement of heritage evaluation

systems, enhances the scientific and transparent nature of conservation decision-making, and strengthens the linkage between research, governance, and sustainable cultural tourism (Z. Liu *et al.*, 2023).

4.3. Social participation and cultural identity

Recent growth in research on community participation and social culture underscores the increasing recognition of residents' roles in the preservation process. Cultural meaning construction, once secondary, now serves as a central mechanism connecting spatial renewal with collective identity (P. Tang *et al.*, 2019). Within this context, traditional architecture functions not only as a carrier of memory but also as a platform for interaction where tourism, governance, and cultural production intersect (Chen *et al.*, 2019).

This transition parallels broader shifts in rural governance, emphasizing bottom-up agency and collaborative management. Integrating local identity and community participation into conservation practices enhances the sustainability and legitimacy of protection frameworks (Cao *et al.*, 2024). Overall, architectural heritage research increasingly reflects social dynamics and participatory values alongside material or formal considerations.

4.4. Formation of a multidimensional framework

Building on the topic analysis and clustering results, this section synthesizes findings into a structured five-dimensional framework. To enhance conceptual clarity and ensure alignment with international heritage assessment

terminology (United Nations Educational, Scientific and Cultural Organization, International Council on Monuments and Sites, and HBIM), the terminology of each dimension has been refined and clearly defined (Table 5).

This five-dimensional framework is not a simple categorical model but a conceptual synthesis derived from both topic evolution analysis and theoretical mapping. The dimensions reflect a hierarchical relationship among physical integrity, functional adaptation, and socio-cultural continuity (Chen *et al.*, 2019; Feng & Xiao, 2020; C. Hu & Gong, 2017).

Specifically, Building Features and Materials and Construction correspond to the material dimension, emphasizing the conservation of tangible heritage through the preservation of both form and substance. Spatial and Functional Adaptability bridges the material and social domains, linking physical transformation with evolving community needs under rural tourism development. Building Culture represents the symbolic and aesthetic dimension, highlighting how meaning and identity are embedded in architectural expression. Social Culture extends the framework into the social dimension, integrating governance, participation, and intergenerational knowledge transmission.

Together, these five dimensions outline a multi-scalar analytical model that captures both the physical and cultural dynamism of traditional village buildings. This structure reflects the principles of architectural adaptation theory (Douglas, 2006; Yung *et al.*, 2017), which emphasizes

Table 5. Conceptual definition of the five-dimensional framework for traditional village building protection

Dimension	Definition
Building features	The degree to which the original spatial form, structure, and layout of traditional buildings remain intact, reflecting vernacular design logic and construction coherence.
Materials and construction	The preservation and traceability of traditional materials and local construction techniques that embody regional knowledge.
Space and functional adaptability	The flexibility of traditional spaces to accommodate new uses—particularly for rural tourism, while maintaining cultural meaning.
Building culture	The symbolic, aesthetic, and vernacular characteristics that represent local identity and cultural expression.
Social culture	The persistence of traditional architectural practices through community participation and intergenerational cultural transmission.

hierarchical renewal and sustainable reuse, and aligns with rural tourism theory (Su, 2018; M. Wu *et al.*, 2022), which balances cultural authenticity with functional innovation. Hence, the framework not only represents a taxonomic categorization but also embodies an evolutionary logic connecting material conservation, cultural meaning, and social participation.

4.5. Conceptual implications

The findings suggest that contemporary rural architecture research is moving toward a balanced model that integrates cultural continuity, technological innovation, and community engagement. The interplay of these forces demonstrates that protection is no longer merely the preservation of the past but is an adaptive, socially embedded process. The multidimensional framework developed here provides both analytical depth and practical relevance, enabling future studies to address protection, utilization, and revitalization within a unified system.

5. Conclusion

5.1. Methodological contributions

This study demonstrates the innovative application of the BERTopic model in transforming unstructured academic literature into a structured protection framework. Unlike traditional content analysis, the integration of BERT-based semantic embeddings and HDBSCAN clustering enables the detection of subtle semantic transitions within the heritage conservation discipline. This approach represents a methodological shift from descriptive data analysis to conceptual modeling, strengthening the connection between quantitative text mining and theoretical synthesis.

5.2. The evaluation framework and future directions

Based on data-driven insights, this study develops a five-dimensional evaluation framework encompassing building features, materials and construction, space and functional adaptability, building culture, and social culture. This framework integrates key theoretical perspectives, including architectural adaptation theory (Douglas, 2006; Yung *et al.*, 2017), which emphasizes hierarchical renewal, and rural tourism theory (Su *et al.*, 2018; M. Wu *et al.*, 2022), which balances cultural authenticity with functional innovation. By clarifying the interconnections between physical characteristics and cultural continuity, the framework provides a robust theoretical basis for the protection of rural architectural heritage.

In the next phase, the research team plans to conduct field investigations in Sanjia Village, Yunnan Province, a representative traditional settlement with active rural tourism development. Mixed research methods, including

semi-structured interviews, Delphi expert consultation, and field observation, will be employed to verify the feasibility, validity, and consistency of the proposed five-dimensional framework within real-world governance contexts. This empirical extension aims to transform the framework from a theoretical construct into an operational and evaluative tool, enhancing its practical relevance for policy formulation and community-based heritage management.

Several limitations should be acknowledged. First, the study primarily relies on academic databases such as CNKI, Web of Science, and Scopus, potentially omitting grey literature, including local policy documents. Second, while the framework is theoretically grounded, its practical implementation in diverse rural contexts should be viewed as a potential pathway for exploration rather than a finalized protocol. Future research should focus on empirical validation through field-based case studies to test the framework's adaptability across different geographic areas. Additionally, integrating broader stakeholder perspectives, particularly those of local residents and policy-makers, remains crucial to ensure the framework's long-term sustainability and practical impact.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Wenjun Ge

Formal analysis: Wenjun Ge

Investigation: Wenjun Ge

Methodology: Wenjun Ge

Writing—original draft: Wenjun Ge

Writing—review & editing: All authors

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

The data presented in this study are available on request from the corresponding author.

References

- Aktürk, İ. (2025). Practices for the protection of rural architecture: The case of Serbian rural area. *International Journal of Research and Innovation in Social Science*, IX(I), 1283–1288.
<https://doi.org/10.47772/IJRIS.2025.9010108>
- Arias Tapiero, J. C., Graus, S., Khei, S., Silva, D., Conde, O., Ferreira, T. M., Ortega, J., Luso, E., Rodrigues, H., & Vasconcelos, G. (2025). An ICT-enhanced methodology for the characterization of vernacular built heritage at a regional scale. *International Journal of Architectural Heritage*, 19(6), 966–984.
<https://doi.org/10.1080/15583058.2024.2320862>
- Barber, D. M., Dallas, R. W. A., & Mills, J. P. (2006). Laser scanning for architectural conservation. *Journal of Architectural Conservation*, 12(1), 35–52.
<https://doi.org/10.1080/13556207.2006.10784959>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *The Journal of Machine Learning Research*, 3, 993–1022.
- Cano, M., Garzón, E., & Sánchez-Soto, P. J. (2013). Preservation and conservation of rural buildings as a subject of cultural tourism: A review concerning the application of new technologies and methodologies. *Journal of Tourism & Hospitality*, 02(02).
<https://doi.org/10.4172/2167-0269.1000115>
- Cao, K., Liu, Y., Cao, Y., Wang, J., & Tian, Y. (2024). Construction and characteristic analysis of landscape gene maps of traditional villages along ancient Qin-Shu roads, Western China. *Heritage Science*, 12(1), 37.
<https://doi.org/10.1186/s40494-024-01155-y>
- Cattaneo, T., Giorgi, E., & Ni, M. (2018). Landscape, architecture and environmental regeneration: A research by design approach for inclusive tourism in a rural village in China. *Sustainability*, 11(1), 128.
<https://doi.org/10.3390/su11010128>
- Chen, Y., Yuan, J., & Lu, Q. (2019). Study on the usability of residential buildings in traditional villages in Southern China from the perspective of human settlements. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11585 LNCS, 14–28.
https://doi.org/10.1007/978-3-030-23538-3_2
- Djabarouti, J., & Ren, Y. (2024). Cultural convergence in heritage landscape conservation: A comparative study of Chinese and English traditions. *Arts & Communication*, 2(1), 1923.
<https://doi.org/10.36922/ac.1923>
- Douglas, J. (2006). *Building adaptation* (2nd ed.). Butterworth-Heinemann.
- Du, X., & Shi, D. (2019). Rural heritage: Value, conservation and revitalisation—From the perspective of the human-land relationship. *Built Heritage*, 3(2), 1–6.
<https://doi.org/10.1186/BF03545723>
- Egger, R., & Yu, J. (2022). A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify Twitter posts. *Frontiers in Sociology*, 7, 886498.
<https://doi.org/10.3389/fsoc.2022.886498>
- Feng, H., & Xiao, J. (2020). Dynamic authenticity: Understanding and conserving Mosuo dwellings in China in transitions. *Sustainability*, 13(1), 143.
<https://doi.org/10.3390/su13010143>
- Gao, J., & Wu, B. (2017). Revitalizing traditional villages through rural tourism: A case study of Yuanjia Village, Shaanxi Province, China. *Tourism Management*, 63, 223–233.
<https://doi.org/10.1016/j.tourman.2017.04.003>
- González Martínez, P. (2019). From verifiable authenticity to verisimilar interventions: Xintiandi, Fuxing SOHO, and the alternatives to built heritage conservation in Shanghai. *International Journal of Heritage Studies*, 25(10), 1055–1072.
<https://doi.org/10.1080/13527258.2018.1557235>
- Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv Preprint*.
<https://doi.org/10.48550/ARXIV.2203.05794>
- Harth, A. (2024). X-ray fluorescence (XRF) on painted heritage objects: A review using topic modeling. *Heritage Science*, 12(1), 17.
<https://doi.org/10.1186/s40494-024-01135-2>
- Hellwig, N. C., Fehle, J., Bink, M., Schmidt, T., & Wolff, C. (2024). Exploring Twitter discourse with BERTopic: Topic modeling of tweets related to the major German parties during the 2021 German federal election. *International Journal of Speech Technology*, 27(4), 901–921.
<https://doi.org/10.1007/s10772-024-10142-4>
- Hu, C., & Gong, C. (2017). Exploring the creation of ecological historic district through comparing and analyzing four typical revitalized historic districts. *Energy Procedia*, 115, 308–320.
<https://doi.org/10.1016/j.egypro.2017.05.028>
- Hu, H., Qiao, X., Yang, Y., & Zhang, L. (2021). Developing a resilience evaluation index for cultural heritage site: Case study of Jiangwan Town in China. *Asia Pacific Journal of Tourism Research*, 26(1), 15–29.
<https://doi.org/10.1080/10941665.2020.1805476>
- Huang, H., Xie, Y., Chen, J., Liang, S., & Chen, Z. (2024). Bioclimatic design strategy of vernacular architecture

- in the south-east of China: A case study in Fujian, China. *International Journal of Low-Carbon Technologies*, 19, 1–17.
<https://doi.org/10.1093/ijlct/ctad079>
- Huo, H., Shen, K., Han, C., & Yang, M. (2024). Measuring the relationship between museum attributes and visitors: An application of topic model on museum online reviews. *PLOS ONE*, 19(7), e0304901.
<https://doi.org/10.1371/journal.pone.0304901>
- Ji, G., & Abd Manan, M. S. (2023). An overview of «all-for-one tourism» development and possible future research directions for Ningxia's tourism using VUP (Visual Urban Perception). *e-Review of Tourism Research*, 20(1), 23–50.
- Ji, G., & Abd Manan, M. S. (2024). Iconic visions: Shaping Yinchuan's future tourism business through landmark perceptions. *International Journal of Business and Technology Management*, 6(3), 79–91.
<https://doi.org/10.55057/ijbtm.2024.6.3.9>
- Junjie Su. (2018). Conceptualising the subjective authenticity of intangible cultural heritage. *International Journal of Heritage Studies*, 24(10), 1032–1047.
<https://doi.org/10.1080/13527258.2018.1428662>
- Lane, B. (1994). What is rural tourism? *Journal of Sustainable Tourism*, 2(1–2), 7–21.
<https://doi.org/10.1080/09669589409510680>
- Li, S.-L., Li, L., Cao, M.-W., Cao, L., Jia, W., & Liu, X.-P. (2017). Rapid modeling of Chinese Huizhou traditional vernacular houses. *IEEE Access*, 5, 20668–20683.
<https://doi.org/10.1109/ACCESS.2017.2754858>
- Li, X., & Lei, L. (2021). A bibliometric analysis of topic modelling studies (2000–2017). *Journal of Information Science*, 47(2), 161–175.
<https://doi.org/10.1177/0165551519877049>
- Lindholm, K.-J., & Ekblom, A. (2019). A framework for exploring and managing biocultural heritage. *Anthropocene*, 25, 100195.
<https://doi.org/10.1016/j.ancene.2019.100195>
- Liu, B., Wu, C., Xu, W., Shen, Y., & Tang, F. (2024). Emerging trends in GIS application on cultural heritage conservation: A review. *Heritage Science*, 12(1), 139.
<https://doi.org/10.1186/s40494-024-01265-7>
- Liu, X. (2024). The effects of commercialisation on urban heritage in Tianjin: A study of citizens' livelihood in the Five Avenues (Wudadao) historical district. *Built Heritage*, 8(1), 42.
<https://doi.org/10.1186/s43238-024-00146-z>
- Liu, Z., Zhang, M., & Osmani, M. (2023). Building Information Modelling (BIM) driven sustainable cultural heritage tourism. *Buildings*, 13(8), 1925.
<https://doi.org/10.3390/buildings13081925>
- Long, C., Lu, S., Chang, J., Zhu, J., & Chen, L. (2022). Tourism environmental carrying capacity review, hotspot, issue, and prospect. *International Journal of Environmental Research and Public Health*, 19(24), 16663.
<https://doi.org/10.3390/ijerph192416663>
- López-González, C., & García-Valldecabres, J. (2023). The integration of HBIM-SIG in the development of a virtual itinerary in a historical centre. *Sustainability*, 15(18), 13931.
<https://doi.org/10.3390/su151813931>
- Manan, M. S. A., & Yuan, Y. Z. (2025). Parasitic architecture: Developing a thematic design framework for adaptive reuse as an urban regeneration strategy in Kuala Lumpur, Malaysia. *MAJ*, 7(5), 85–96.
- McInnes, L., Healy, J., & Melville, J. (2018). *UMAP: Uniform Manifold Approximation and Projection for dimension reduction*. arXiv.
<https://doi.org/10.48550/arXiv.1802.03426>
- McInnes, L., Healy, J., & Melville, J. (2020). Algorithms for hierarchical clustering: An overview. arXiv.
<https://doi.org/10.1002/widm.53>
- Medvecki, D., Bašaragin, B., Ljajić, A., & Milošević, N. (2024). Multilingual transformer and BERTopic for short text topic modeling: The case of Serbian (Vol. 872, pp. 161–173).
https://doi.org/10.1007/978-3-031-50755-7_16
- Miao, Y., & Chiou, S.-C. (2013). Study on the wind environment of the architecture communities: Traditional typical Min Nan human settlements» case. *Mathematical Problems in Engineering*, 2013, 1–11.
<https://doi.org/10.1155/2013/467076>
- Ministry of Housing and Urban-Rural Development of the People's Republic of China (MOHURD). (2012). *Guiding opinions on strengthening the protection and development of traditional villages*. Available from: https://www.mohurd.gov.cn/gongkai/zc/wjk/art/2012/art_17339_212337.html [Last accessed on 2024 Oct 24].
- Ministry of Housing and Urban-Rural Development of the People's Republic of China (MOHURD). (2013). *Notice on issuing the basic requirements for the compilation of traditional village protection and development planning (Trial)*. Available from: https://www.mohurd.gov.cn/gongkai/zc/wjk/art/2013/art_17339_215684.html [Last accessed on 2024 Oct 24].
- Murtagh, F., & Contreras, P. (2011). Algorithms for hierarchical clustering: An overview. *WIREs Data Mining and Knowledge Discovery*, 2(1), 86–97.
<https://doi.org/10.1002/widm.53>
- Nguyen, L. T., Chansanam, W., Hunsapun, N., Chaichuay, V., Kanyacome, S., Takhom, A., Jaroenruen, Y., & Li, C. (2024).

- Evaluating the performance of topic modeling techniques for bibliometric analysis research: An LDA-based approach. *HighTech and Innovation Journal*, 5(2), 312–330.
<https://doi.org/10.28991/HIJ-2024-05-02-07>
- Nguyen, T., Van, L. N., Duc, A. N., & Viet, S. D. (2025). A framework for neural topic modeling with mutual information and group regularization. *Neurocomputing*, 645, 130420.
<https://doi.org/10.1016/j.neucom.2025.130420>
- Olgun, T. N., & Karatosun, M. B. (2019). Rural architectural heritage conservation and sustainability in Turkey: The case of Karaca village of Malatya region. *International Journal of Design & Nature and Ecodynamics*, 14(3), 195–205.
<https://doi.org/10.2495/DNE-V14-N3-195-205>
- Olukoya, O. A. P. (2021). Framing the values of vernacular architecture for a value-based conservation: A conceptual framework. *Sustainability*, 13(9), 4974.
<https://doi.org/10.3390/su13094974>
- Orozco Carpio, P. R., & Rolim, R. (2024). GIS and HBIM for tourism management: A multiscale challenge. In: *Proceedings Heritage Digital Technologies and Tourism Management - HEDIT2024*. June 20-21, 2024, Valencia, Spain.
<https://doi.org/10.4995/hedit2024.2024.17751>
- Oumoumen, K. (2024). Automation of historical buildings: Historical Building Information Modeling (HBIM) based Virtual Reality (VR). *Materials Research Proceedings*, 40, 319–322.
<https://doi.org/10.21741/9781644903117-34>
- Peng, Y., Meng, M., Huang, Z., Wang, R., & Cui, G. (2021). Landscape connectivity analysis and optimization of Qianjiangyuan National Park, Zhejiang Province, China. *Sustainability*, 13(11), 5944.
<https://doi.org/10.3390/su13115944>
- Qianda, Z., Guoquan, Z., Hussein, M. K., Ariffin, N. F. M., & Yunos, M. Y. M. (2021). Identification of rural vernacular building character and conservation strategy from the perspective of rural tourism—A case study of Yayou Gou Village in Shandong Province, China. *E3S Web of Conferences*, 251, 02076.
<https://doi.org/10.1051/e3sconf/202125102076>
- Saputra, R. (2024). Governance frameworks and cultural preservation in Indonesia: Balancing policy and heritage. *Journal of Ethnic and Cultural Studies*, 11(3), 25–50.
<https://doi.org/10.29333/ejecs/2145>
- Sharpley, R. (2002). Rural tourism and the challenge of tourism diversification: The case of Cyprus. *Tourism Management*, 23(3), 233–244.
[https://doi.org/10.1016/S0261-5177\(01\)00078-4](https://doi.org/10.1016/S0261-5177(01)00078-4)
- Shirvani Dastgerdi, A., & Kheyroddin, R. (2022). Policy recommendations for integrating resilience into the management of cultural landscapes. *Sustainability*, 14(14), 8500.
<https://doi.org/10.3390/su14148500>
- Shirvani Dastgerdi, A., & Kheyroddin, R. (2023). Building resilience in cultural landscapes: Exploring the role of transdisciplinary and participatory planning in the recovery of the Shushtar Historical Hydraulic System. *Sustainability*, 15(13), 10433.
<https://doi.org/10.3390/su151310433>
- Siewczyński, B., & Szot, J. (2025). BIM goals and uses in the management, maintenance, and preservation of historic buildings: An open access perspective. Implementation characteristics of HBIM for improved documentation and lifecycle management. *npj Heritage Science*, 13(1), 103.
<https://doi.org/10.1038/s40494-025-01588-z>
- Su, M., Sun, Y., Min, Q., & Jiao, W. (2018). A community livelihood approach to agricultural heritage system conservation and tourism development: Xuanhua Grape Garden urban agricultural heritage site, Hebei Province of China. *Sustainability*, 10(2), 361.
<https://doi.org/10.3390/su10020361>
- Tang, C., Qin, S., Dai, X., & Lv, J. (2023). A review and prospect of China's rural revitalization research from the perspective of culture and tourism. *Progress in Geography*, 42(8), 1437–1452.
<https://doi.org/10.18306/dlkxjz.2023.08.001>
- Tang, P., Wang, X., & Shi, X. (2019). Generative design method of the facade of traditional architecture and settlement based on knowledge discovery and digital generation: A case study of Gunan jie Street in China. *International Journal of Architectural Heritage*, 13(5), 679–690.
<https://doi.org/10.1080/15583058.2018.1463415>
- Tao, R., Chen, P., & Aoki, N. (2025). Conceptual changes and controversies in rural historical building relocation in China under the heritage adaptive reuse discourse. *Built Heritage*, 9(1), 6.
<https://doi.org/10.1186/s43238-024-00172-x>
- Tenzer, M., & Schofield, J. (2024). Using topic modelling to reassess heritage values from a people-centred perspective: Applications from the North of England. *Cambridge Archaeological Journal*, 34(1), 147–168.
<https://doi.org/10.1017/S0959774323000203>
- Wang, H., Liu, N., Chen, J., & Guo, S. (2022). The relationship between urban renewal and the built environment: A systematic review and bibliometric analysis. *Journal of Planning Literature*, 37(2), 293–308.
<https://doi.org/10.1177/08854122211058909>

- Wang, L. (2021). Causal analysis of conflict in tourism in rural China: The peasant perspective. *Tourism Management Perspectives*, 39, 100863.
<https://doi.org/10.1016/j.tmp.2021.100863>
- Wang, Y., Huang, W., & Yao, X. (2021). Research on the evaluation of tourism destination image based on user generated content. *2021 2nd International Conference on Artificial Intelligence and Information Systems*, 1–5.
<https://doi.org/10.1145/3469213.3470694>
- Wu, M.-Y., Wu, X., Li, Q.-C., & Tong, Y. (2022). Community citizenship behavior in rural tourism destinations: Scale development and validation. *Tourism Management*, 89, 104457.
<https://doi.org/10.1016/j.tourman.2021.104457>
- Wu, Z., Mu, D., Sun, M., Xiong, Q., & Xiao, S. (2025). Evaluating rural tourism landscapes in world heritage buffer zones using PLES (Production—Living—Ecological Space) theory. *npj Heritage Science*, 13(1), 464.
<https://doi.org/10.1038/s40494-025-02033-x>
- Xie, K., Zhang, Y., & Han, W. (2024). Architectural heritage preservation for rural revitalization: Typical case of traditional village retrofitting in China. *Sustainability*, 16(2), 681.
<https://doi.org/10.3390/su16020681>
- Yao, P. (2016). Research on the space environment characteristic and the development of the traditional village in China: Taking Pengzhuang in the northern Jiangsu province as an example. *Brazilian Archives of Biology and Technology*, 59(spe).
<https://doi.org/10.1590/1678-4324-2016160543>
- Yung, E. H. K., Zhang, Q., & Chan, E. H. W. (2017). Underlying social factors for evaluating heritage conservation in urban renewal districts. *Habitat International*, 66, 135–148.
<https://doi.org/10.1016/j.habitatint.2017.06.004>
- Zhang, X. (2015). Design of recreational corridor planning of Henan vernacular landscape based on resource integration. In: *Proceedings of the 2015 International Forum on Energy, Environment Science and Materials*. September 25–26, 2015, Shenzhen, China.
<https://doi.org/10.2991/ifeesm-15.2015.262>
- Zhang, Z., Xiong, K., & Huang, D. (2023). Natural world heritage conservation and tourism: A review. *Heritage Science*, 11(1), 55.
<https://doi.org/10.1186/s40494-023-00896-6>
- Zhao, X., & Greenop, K. (2019). From ‘neo-vernacular’ to ‘semi-vernacular’: A case study of vernacular architecture representation and adaptation in rural Chinese village revitalization. *International Journal of Heritage Studies*, 25(11), 1128–1147.
<https://doi.org/10.1080/13527258.2019.1570544>
- Zhou, J. (2021). Statistical research on the development of rural tourism economy industry under the background of big data. *Mobile Information Systems*, 2021, 1–11.
<https://doi.org/10.1155/2021/9152173>
- Zhou, L., Wall, G., Zhang, D., & Cheng, X. (2021). Tourism and the (re)making of rural places: The cases of two Chinese villages. *Tourism Management Perspectives*, 40, 100910.
<https://doi.org/10.1016/j.tmp.2021.100910>
- Zhu, X.-X., Mu, Q.-R., & Liang, W.-Z. (2022). An innovative strategic choice for stakeholders in the Chinese traditional commercial street renewal using evolutionary game theory. *Journal of Innovation & Knowledge*, 7(3), 100225.
<https://doi.org/10.1016/j.jik.2022.100225>