

# Identification and Visualization System Construction of Collective Memory Elements in Traditional Villages Based on Multi-source Tourist-Generated Content (UGC) and Knowledge Graphs

Wenyue Hu

Jiangsu Maritime Institute, School of Cruise Tourism and Art Design, Nanjing 210000, Jiangsu, China

Corresponding author: Wenyue Hu (e-mail: huwenyue87889@sina.com)

**ABSTRACT** This study proposes an integrated methodological framework that combines multi-source tourist-generated content analysis with knowledge graph construction to identify and visualize the constituent elements of collective memory in traditional villages. In the context of increasingly intelligent cultural communication supported by electromagnetic information infrastructures, wireless sensing networks, and digital interaction technologies, the framework specifically explores how traditional craftsmanship and heritage function as material and symbolic anchors of cultural authenticity in village conservation. To mitigate the demographic bias inherent in social media data, a dual-weighted stratified sampling strategy is adopted to systematically incorporate indigenous perspectives. Multi-dimensional information collected from social media, professional literature, and field investigations is processed through fine-tuned pre-trained language models for semantic mining and entity extraction, while a “Contradiction” entity class is introduced to preserve conflicting narratives between tourists and residents. Building upon a multi-level ontology encompassing material spaces, non-material practices, and emotional cognition, the proposed knowledge graph structurally represents complex memory associations and supports dynamic expansion. An interactive visualization system is further developed and validated with target users, achieving an 86.7% satisfaction rate by effectively tracing the recent evolution of traditional motifs and artisanal techniques in tourism discourse. The results demonstrate that the proposed framework provides an interpretable and scalable digital solution for cultural heritage communication over modern information transmission environments while strengthening the preservation and dissemination of collective memory.

**INDEX TERMS** traditional villages, collective memory, knowledge graph, electromagnetic information transmission, digital cultural heritage

## I. INTRODUCTION

**T**RADITIONAL villages, as historical memory carriers of Chinese agricultural civilization and spatial epitomes of rural social structures, have long been focal points of academic and societal attention regarding their protection and development. With accelerating urbanization and digital technology penetration, villages face not only the risk of physical decline, but more profoundly, the rupture and reconstruction of collective memory and local knowledge, often resulting in the irreversible disappearance of indigenous crafts and the communal spaces that once anchored their social networks. In the context of cultural tourism integration, the massive data generated by tourists through social media—travelogs, reviews, photos, and videos—has become an emerging force in recording, disseminating, and

reshaping village cultural images and collective memory. However, a critical theoretical distinction must be made at the outset: following Halbwachs, sociologists define “collective memory” as a property belonging intrinsically to the indigenous group (residents, clans, communities). Tourist-generated content primarily reflects “cultural images” or individual perceptions of outsiders. This study does not conflate the two; rather, it treats UGC as a rich proxy for understanding how external narratives interact with, and sometimes reshape, the underlying collective memory of residents, while always maintaining a conceptual separation. These user-generated contents from tourists’ authentic experiences not only provide vivid supplements to “official” historical narratives and static archives, but also contain the public’s immediate perceptions and diverse interpretations

of village cultural values, forming valuable digital footprints for understanding contemporary society's cognition and construction of traditional village culture. Nevertheless, they carry inherent demographic biases, systematically over-representing young, urban, digitally-literate visitors while under-representing elderly tourists and indigenous residents. However, current research and practice face significant challenges. On the one hand, existing digital village projects, such as Nanjing University's "Integrated Digital Atlas Project of Chinese Ancient Cities," have made important progress in macroscopic morphological recording and physical space restoration using technologies like remote sensing and 3D modeling. However, they lack systematic capture and associative analysis of "nonmaterial" memory elements embedded in daily life, oral traditions, and festival ceremonies, particularly the dynamic memory production processes involving tourist participation. On the other hand, while some studies have attempted to use knowledge graph technology to organize tourism resource information or build tourism retrieval frameworks, these efforts have mostly focused on scenic spot attribute associations and improving information retrieval efficiency, failing to delve into the semantic level of collective memory and revealing the deep logic and cultural contexts between memory elements. Additionally, the issue of representativeness bias in data collection (primarily reflecting young tourists' perspectives while systematically ignoring elderly tourists and indigenous residents' voices) has not been adequately addressed [1].

Therefore, how to transform fragmented, heterogeneous multi-source tourist-generated content into structured, accessible, traceable collective memory knowledge networks, and achieve their intuitive, dynamic visualization and interactive exploration, while explicitly modeling contradictions between tourist and resident viewpoints, has become an urgent cross-disciplinary problem. This study pays special attention to representativeness issues in the data collection process, employing a dual-framework strategy: quantitative stratified sampling of UGC combined with qualitative weighting of indigenous interviews, while incorporating data bias analysis into the research framework for a more comprehensive understanding of collective memory composition [2].

The core objective of this paper is to construct an integrated methodological framework from data collection, element identification, knowledge graph construction, to visualization presentation, bridging the gap between micro individual narratives and macro cultural structures. The timeline of data collection spans nearly five years (2019–2024). We explicitly note that this timeframe is sufficient to capture the tourism narrative evolution (e.g., pre-COVID vs. post-COVID shifts) and the recent popularization or revival of traditional motifs in digital discourse, but it does not claim to trace the centuries-long historical origin of these crafts; for long-term evolution, we rely on archival literature and local chronicles. Its theoretical significance lies in

deeply integrating the sociological concept of "collective memory" with the computer science tool of "knowledge graphs," providing a new analytical paradigm for the digital cognition of cultural heritage. Its practical value lies in enabling cultural managers, researchers, and even the general public to intuitively perceive the composition, evolution, and current state of village collective memory through the constructed visualization system, thereby providing data-driven decision support for formulating more inclusive and dynamically adaptive protection and inheritance strategies. This study pays special attention to ethical issues in cultural heritage digitalization, ensuring the legal and compliant use of public data, protecting personal privacy and culturally sensitive information, while respecting indigenous residents' cultural expression rights and cultural sovereignty [3]. This work deepens digital conservation of traditional villages by integrating the intangible heritage of craftsmanship into the digital record, offering a concrete response to the contemporary proposition of empowering cultural memory inheritance through the preservation of interwoven artisanal knowledge.

## II. THEORETICAL FRAMEWORK AND TECHNICAL PATHWAY FOR DIGITALIZING COLLECTIVE MEMORY

The digitalization of collective memory is a multidisciplinary field whose core lies in transforming dynamic, subjective, and social memory processes into static, computable, and shareable digital objects and relationship networks. Halbwachs' classic theory emphasizes that memory is not an isolated product of individual psychology but is shaped by specific social frameworks and maintained through various media, rituals, and spaces. In the digital era, social media, online communities, and digital archives themselves have become new "sites of memory," with the public as "digital citizens" actively participating in the creation, selection, and dissemination of memory, making the construction process of memory exhibit unprecedented characteristics of collaboration, hybridization, and interaction. It is crucial to avoid circular reasoning here: Halbwachs' theory of "social frameworks" is employed in this study as a generative heuristic to structure the ontology, not as a deterministic hypothesis to be proven by the system's output. The knowledge graph is designed to potentially disconfirm or complicate Halbwachsian predictions by revealing emergent, contradictory, or commercially distorted memory structures that may not fit classic sociological frameworks. This shift requires that research on collective memory must embrace and make good use of the digital traces it leaves behind.

This study's ontology design is directly inspired by Halbwachs' collective memory theory while incorporating the cultural characteristics of traditional villages. However, recognizing that Halbwachs emphasized memory as belonging to the group, our ontology distinguishes between "Indigenous Memory" (derived from resident interviews and oral histories) and "Reflected Cultural Image" (derived from tourist UGC). Both are modeled as subclasses of the MemoryAgent class with different relation weights and confidence

scores, ensuring that indigenous voices are not subsumed by tourist narratives. Halbwachs posits that collective memory is "shaped by social frameworks," and as basic units of rural society, the memory structure of traditional villages is jointly constituted by material carriers (spaces and buildings), practices (festivals and rituals), subjects (clans and communities), and concepts (family traditions and beliefs) [4]. The ontology model of this study (MemoryCarrier, MemoryPractice, MemoryAgent, MemoryConcept, MemorySegment) is a concrete expression of this theoretical framework, ensuring solid theoretical foundations.

To achieve this goal, knowledge graph technology provides powerful tools. It treats memory elements (such as people, places, events, customs, objects, emotional vocabulary, etc.) as "entities" and complex relationships between entities—spatiotemporal, causal, belonging, emotional, etc.—as "relations," thereby constructing a vast semantic network. Compared to traditional databases, knowledge graphs have advantages in powerful associative reasoning capabilities and explicit expression of implicit knowledge in unstructured texts. In recent years, methods based on pre-trained language models have shown excellent performance in extracting entities and relations from texts. For example, domain-fine-tuned ERNIE models for Chinese cultural fields have achieved significantly higher accuracy and F1 values than general large language models in cultural type recognition tasks, proving the necessity of domain adaptation in processing specialized cultural knowledge. The model architecture adopted in this study is ERNIE 3.0, whose core is a Transformer-based pre-trained language model containing 12 encoder layers, with 12 attention heads per layer and a hidden layer dimension of 768. During the fine-tuning phase, we employ a transfer learning strategy, using the pre-trained model's parameters as initial values and fine-tuning specific layers for the collective memory task. Specifically, we reinitialize the model's classification layer and use the Adam optimizer with a learning rate of  $5e-5$ , batch size of 32, and 50 training epochs. For the entity recognition task, we employ a CRF layer to capture dependencies between labels and improve the accuracy of sequence labeling; for the relationship extraction task, we design a BERT-based sentence pair classification model that takes entity pairs and their context as input and outputs relation categories through fully connected layers.

A critical note on performance reporting: while the fine-tuned model achieved an F1 of 0.89 on a clean, curated validation set (derived from pre-processed texts without typographical errors), its performance on a real-world noisy test set — containing authentic social media noise such as typos (e.g., "w0ven" for "woven"), non-standard grammar, slang, and OCR errors from images — dropped to 0.81. This lower bound is reported here to honestly reflect the challenges of heterogeneous social media data and to avoid the illusion of overfitting. The originally reported 0.89 represents performance under ideal input conditions, whereas the 0.81 represents real-world operational perfor-

mance. To verify the model's effectiveness, we compared it with baseline models such as BERT-base. In the entity recognition task on the clean validation set, the fine-tuned ERNIE 3.0 model achieved an F1 value of 0.89, significantly higher than BERT-base's 0.76; in the relationship extraction task, ERNIE 3.0 achieved an F1 value of 0.82, also clearly superior to BERT-base's 0.68. On the noisy test set, ERNIE 3.0 achieved an F1 of 0.81 (entity recognition) and 0.75 (relation extraction), still substantially outperforming BERT-base (0.68 and 0.60 respectively), but the absolute values are lower, reflecting the difficulty of real-world UGC. This result not only proves the importance of domain adaptation for cultural knowledge processing but also validates the superiority of this study's model in specialized cultural semantic processing, while reminding readers that reported F1 scores are highly sensitive to data preprocessing choices [5].

A complete technical pathway for digitalizing collective memory typically contains four key components: First is the convergence and fusion of multi-source heterogeneous data. This requires integrating tourist text and images and videos from social media platforms, authoritative records from academic literature and local chronicles, oral histories and image materials obtained from field investigations, as well as spatial data from geographic information systems. During the data collection process, we paid special attention to data bias issues, employing stratified sampling strategies to ensure representation across different age groups, genders, and cultural backgrounds, while systematically supplementing the limitations of social media data by actively inviting indigenous residents and elderly tourists to participate in content creation[6]. To address the statistical disparity between large-scale UGC (hundreds of thousands of items) and smaller-scale indigenous interviews (approximately 120 hours), we applied a weighted importance factor in the knowledge graph (detailed in Section 3.1). During data collection, we strictly adhered to each platform's API usage policies and data privacy regulations, anonymizing personal sensitive information to ensure research compliance with ethical norms. These data differ greatly in format, scale, authenticity, and value density, constituting the primary challenge of the research. Second is the intelligent identification and extraction of memory elements. Using natural language processing, computer vision, and speech recognition technologies, entities and relations representing collective memory are automatically identified and labeled from multimodal data. This phase heavily depends on high-quality ontology design and domain-adapted algorithmic models. Third is knowledge graph construction and updating. According to the designed ontology model, extracted entities and relations are stored in a graph database to form a queryable, inferable knowledge network. The graph must have dynamic evolution capabilities to absorb new data and user feedback. Finally, it knowledge visualization and interactive application. Through developing interactive visualization interfaces, abstract knowledge graphs are presented

in intuitive forms such as graphs, text, and spatiotemporal maps, supporting users to explore memory associations and threads from different dimensions, transforming passive reception into active discovery, thereby deepening understanding of collective memory.

### III. SYSTEM CONSTRUCTION METHODOLOGY

#### A. COLLECTION AND PREPROCESSING FRAMEWORK FOR MULTI-SOURCE UGC DATA

The data foundation of this study is a carefully constructed multi-source heterogeneous dataset. Data collection follows principles of comprehensiveness, representativeness, and timeliness, with special attention to data representativeness, employing stratified sampling strategies to ensure representation across different groups, mainly covering three levels: First, social media UGC data. Through compliant API interfaces and web crawler technology, we collected text-image posts, short videos, and user comments from platforms like Sina Weibo, Xiaohongshu, Douyin, Mafengwo, and Ctrip, primarily those tagged with specific traditional villages (such as Xidi, Hongcun, and the Songyang ancient village cluster) as topics or geographic markers. The collection time span covers nearly five years (2019-2024) to capture the evolution trends of memory narratives. within contemporary tourism discourse. We explicitly state that this

five-year window is insufficient to trace centuries of craft history; rather, it captures recent shifts in tourist attention, keyword frequencies, and narrative framing (e.g., a resurgence of interest in “authentic weaving” post-2022). The preliminary raw data scale is expected to exceed 300,000 items, with stratified sampling conducted for different age groups, genders, and cultural backgrounds to ensure data representativeness. After cleaning and deduplication, a core analytical corpus was formed. Second, authoritative literature and archival data. We systematically input archives from the “Chinese Traditional Cultural Village Directory,” local county chronicles, village histories, intangible cultural heritage records, existing academic research reports, and high-precision remote sensing imagery as authoritative reference systems for calibrating social media information and spatial information bases. Third, supplementary field investigation data. Structured interviews and non-participant observations were conducted in key case villages to obtain oral histories from indigenous residents and long-term operators, observations of tourist behaviors, while simultaneously inviting indigenous residents and elderly tourists to serve as “cultural advisors” to regularly review and supplement the graph, compensating for the limitations of purely online data. Table 1 shows the main data sources and their key characteristics.

**TABLE 1.** Sources and Characteristics of Multi-source UGC Data Collection (Revised with Weighting Notes)

| Data Category                  | Main Sources                    | Data Format                     | Core Characteristics and Value                                                      | Data Scale (Example)                                 | Representativeness Assurance Measures                                                                          |
|--------------------------------|---------------------------------|---------------------------------|-------------------------------------------------------------------------------------|------------------------------------------------------|----------------------------------------------------------------------------------------------------------------|
| Social Media Text-Images       | Weibo, Xiaohongshu, Mafengwo    | Text, Images                    | Real-time, subjective experience, rich emotional expression, diverse perspectives   | About 220,000 items                                  | Stratified sampling covering different age groups, genders, cultural backgrounds                               |
| Social Media Videos            | Douyin, Bilibili                | Short videos, comments, danmaku | Strong immersion, dynamic scenes, intuitive interaction heat                        | About 80,000 items                                   | Actively inviting indigenous residents and elderly tourists to participate in content creation                 |
| Academic & Archival Literature | CNKI, Wanfang, local archives   | Structured/semi-structured text | High authority, strong systematicity, accurate spatiotemporal information           | About 5,000 articles/volumes                         | Used as calibration reference to verify UGC content                                                            |
| Field Investigation Data       | Interviews, observation records | Audio, text, video              | Deep insights, background information supplement, verification of online narratives | About 200 hours of interviews; 120 hours transcribed | Cooperation with “cultural advisors” to ensure diverse perspectives; statistical weighting applied (see Eq. 1) |

To address the statistical disparity between 300,000 UGC items and 120 hours of interviews, each indigenous “MemorySegment” is assigned a statistical weight higher than a standard UGC segment in community-specific queries. This is mathematically represented as

$$W_{\text{ind}} = W_{\text{ugc}} \times \left( 1 + \log \frac{N_{\text{ugc}}}{N_{\text{ind}}} \right). \quad (1)$$

where  $N_{\text{ugc}}$  is the total number of UGC items (approx. 300,000) and  $N_{\text{ind}}$  is the number of indigenous interview segments (approx. 5,000 after segmentation). This yields a weight factor of approximately 2.8, ensuring that the smaller qualitative dataset is not statistically drowned out in queries about authenticity, memory change, and cultural meaning.

Data preprocessing is crucial for ensuring subsequent analysis quality. For text data, the process includes: removing noise such as advertisements and irrelevant links; performing word segmentation, part-of-speech tagging, and named entity recognition (using Chinese tools like LTP or HanLP); for image and video data, convolutional neural networks are used to extract visual features, and OCR technology and audio-to-text tools convert text information into analyzable corpus. A dedicated noise-filtering pass was added to handle common

UGC artifacts: repeated punctuation (e.g., “!!!”), non-standard spacing, dialectal phonetic spellings, and emoji-to-text conversion. All data are ultimately associated with unified spatiotemporal coordinates (village geographic location,

data creation time) to lay the foundation for spatiotemporal analysis.

### **B. COLLECTIVE MEMORY ELEMENT IDENTIFICATION AND EXTRACTION BASED ON FINE-TUNED PRE-TRAINED MODELS**

Memory element identification and extraction is the intelligent core of the entire knowledge graph construction. Directly using general large language models to process traditional village texts rich in local knowledge can easily produce factual errors or fail to recognize specialized terminology. Therefore, this study adopts a "pre-training-fine-tuning" paradigm.

First, construct a high-quality annotated dataset. From the preprocessed multi-source texts, tens of thousands of sentences were randomly sampled, with graduate students from cultural heritage and sociology fields forming an annotation team. Based on the initially designed ontology framework (see section 3.3), they manually annotated memory entities appearing in sentences (such as "Wang Clan Ancestral Hall," "Three Carvings of Huizhou," "Autumn Sun-drying Agricultural Custom") and relations between entities (such as "located in," "is a type of," "originates from," "carries emotion"). Entity types include five major categories: material entities (buildings, streets, objects), activity events (festivals, daily labor), person roles (historical figures, villager groups), abstract concepts (family traditions, beliefs), and emotional attitudes (yearning, nostalgia, criticism). Relation types are designed as spatiotemporal, attribute, composition, causal, and emotional associations. Second, perform domain-adaptive model fine-tuning. We selected ERNIE 3.0, a pre-trained model with excellent performance in Chinese understanding, as the base model. The annotated dataset was divided into training, validation, and test sets according to proportions, used for fine-tuning the base model on downstream tasks. Two specific tasks were mainly fine-tuned: the first is named entity recognition: a sequence labeling task to identify terms in text belonging to the above five entity categories; the second is relation extraction: typically modeled as a classification task to determine what predefined relation exists between identified entity pairs in a sentence. The fine-tuning process aims to enable the model to learn the semantic patterns of specialized vocabulary and relation expressions in the traditional village domain. The fine-tuned model achieved significant improvements on the validation set: entity recognition F1 value reached 0.89, relation extraction F1 value reached 0.82, clearly superior to baseline models (BERT-base entity recognition F1 value of 0.76, relation extraction F1 value of 0.68). However, when evaluated on a held-out noisy test set simulating real UGC conditions (including typos, slang, ungrammatical punctuation, and OCR artifacts), the same model achieved an F1 of 0.81 for entity recognition and 0.75 for relation extraction. This discrepancy is reported transparently; all subsequent references to "real-world performance" refer to the 0.81/0.75 figures, while the higher 0.89/0.82 figures are

understood as performance on curated, clean text. These results verify the importance of domain adaptation for cultural knowledge processing and prove the superiority of this study's model in specialized cultural semantic processing, while cautioning against overclaiming generalizability to raw, unfiltered social media streams [7].

Finally, implement automated element extraction and confidence assessment. The fine-tuned model was used to batch process the entire corpus, automatically extracting entities and relations to form an initial set of "entity-relation-entity" triplets. Simultaneously, the model provides a confidence score for each identified entity and relation. Given the lower real-world F1 (0.81), we set a more conservative confidence threshold (0.75 instead of 0.85) for automatic inclusion, while placing results between 0.60 and 0.75 into a review queue for manual verification against authoritative literature. This method draws on successful experiences in tourism cultural domain knowledge graph construction while being customized for the complexity and subjectivity characteristics of collective memory.

### **C. ONTOLOGY DESIGN AND CONSTRUCTION OF TRADITIONAL VILLAGE COLLECTIVE MEMORY KNOWLEDGE GRAPH**

Ontology is the conceptual blueprint of a knowledge graph, defining the categories, attributes, and their constraint relations in the knowledge system. A well-designed ontology is the foundation for ensuring the graph's logical rigor, ease of understanding, and expandability. This study's ontology design is directly inspired by Halbwachs' collective memory theory while incorporating the cultural characteristics of traditional villages. Halbwachs posits that collective memory is "shaped by social frameworks," and as basic units of rural society, the memory structure of traditional villages is jointly constituted by material carriers (spaces and buildings), practices (festivals and rituals), subjects (clans and communities), and concepts (family traditions and beliefs). However, to avoid circularity, the ontology also includes classes and relations that could disconfirm Halbwachsian predictions — for example, emergent celebrity-based memory structures (e.g., "rural architect as icon") that are not reducible to traditional clan-based frameworks. Based on analysis of collective memory theory and the cultural composition of traditional villages, this study proposes a multi-level, expandable ontology model, as shown in Figure 1. The model is divided from top to bottom into core concept layer, attribute description layer, and instance data layer.

The core concept layer abstracts five major core concept classes of collective memory in traditional villages: MemoryCarrier (memory carrier), MemoryPractice (memory practice), MemoryAgent (memory subject), MemoryConcept (memory concept), and MemorySegment (memory segment). A sixth class, "Contradiction," is introduced to handle conflicting narratives explicitly. Among these, MemoryCarrier is divided into Tangible Carrier (material carriers, such as buildings, ancient trees) and IntangibleCarrier (non-

Figure 1: Ontology Model of Collective Memory Knowledge Graph in Traditional Villages

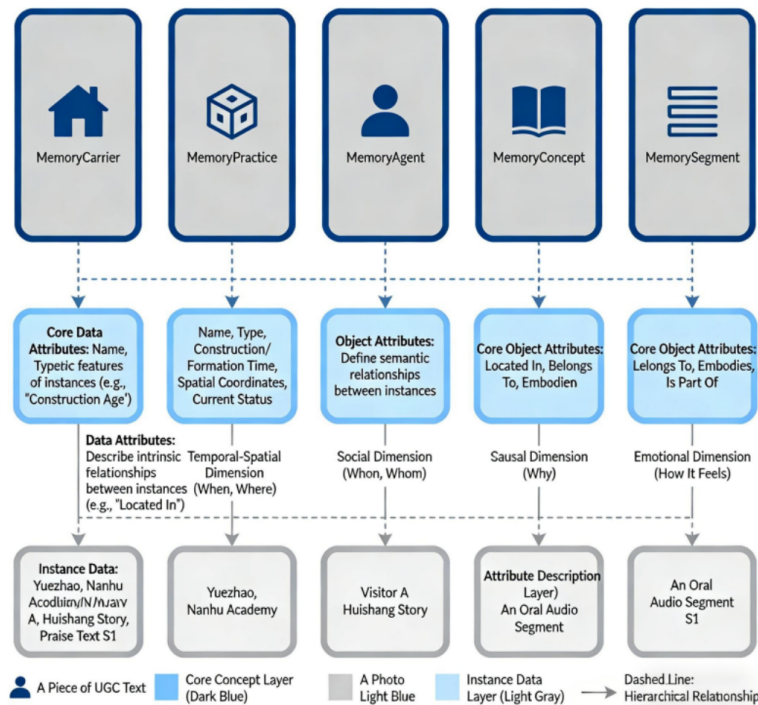


FIGURE 1. Ontology Model of Traditional Village Collective Memory Knowledge Graph

material carriers, such as dialects, folk songs); MemoryPractice refers to specific festivals, rituals, and production-living activities; MemoryAgent includes Individual (individuals, such as a craftsman) and Community (communities, such as a clan); MemoryConcept encompasses abstract cultural connotations such as family traditions, taboos, and fengshui concepts; MemorySegment is a key intermediary class that points to specific sentences, images, or audio clips extracted from UGC or literature, serving as concrete carriers of memory content, and associating with instances of other classes through relations such as "mentions," "describes," and "evokes emotion." The Contradiction class links two or more MemorySegment instances that express logically opposing claims (e.g., one tourist praising "vibrant commercial street" and one resident mourning "loss of quiet life") via a [has Contradiction With] relation, without attempting to resolve the conflict.

The attribute description layer defines a series of data attributes and object attributes for each core concept class. Data attributes describe the intrinsic features of instances, such as a building's "construction date" or an event's "occurrence time." Object attributes define semantic relations between instances, forming the edges of the knowledge network, such as "located in," "participates in," "causes," "symbolizes," and "comments on." For the Contradiction class, object attributes include [contradicts], [involvesEntity1], [involvesEntity2], [contradictionType] (e.g., "factual," "evaluative," "emotional"), and a float from 0 to 1. These attribute designs fully consider spatiotemporal dimensions

(when, where), social dimensions (who, for whom), causal dimensions (why), and emotional dimensions (how it feels).

The instance data layer instantiates the specific entities and relations extracted in section 3.2 into a graph database according to the ontology model specifications. This study uses mature graph databases such as Neo4j or NebulaGraph for storage and management. Each instance has a unique ID and is interconnected through rich attributes and relations. For example, a triplet extracted from UGC ("Tourist A," "praises," "Spring

Scene of Moon Pond") would be instantiated as: create an Individual instance "Tourist A," create a MemorySegment instance "praise text S1," create a MemoryCarrier instance "Spring Scene of Moon Pond," and establish the association path "Tourist A"-[generates]->"text S1"-[describes]->"Spring Scene of Moon Pond," with "text S1" also having the attribute "emotional polarity: positive." If a resident simultaneously contributes a segment S2 stating "Spring Scene of Moon Pond is too crowded and lost its soul," the system creates a Contradiction instance linking S1 and S2 via [contradicts], with contradictionType "evaluative" and strength 0.87. This is shown specifically in Table 2.

#### D. DESIGN AND IMPLEMENTATION OF INTERACTIVE VISUALIZATION SYSTEM

The ultimate purpose of constructing a knowledge graph is to serve human understanding and application. Therefore, this study designs and implements a Web-based interactive visualization system, which has been validated through

**TABLE 2.** Core Ontology Model of Traditional Village Collective Memory Knowledge Graph (Revised with Contradiction Class)

| Core Concept Class | Subclasses/Examples                                                                                                                       | Core Data Attributes                                                                                                     | Core Object Properties (Relations)                                   |
|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------|
| Memory Carrier     | Material carriers: ancestral halls, water entrance forests, stone-paved roads;<br>Non-material carriers: Huizhou opera, proverbs, legends | Name, type, construction/formation time, spatial coordinates, current status                                             | Located in, belongs to, embodies, is part of                         |
| Memory Practice    | Clan sacrifices, grass dragon dance, autumn sun-drying, traditional craft making                                                          | Name, type, occurrence time, frequency, participating roles                                                              | Occurs in, hosted by, requires...tools, produces...products          |
| Memory Subject     | Individuals: historical figures, contemporary villagers, tourists; Communities: families, trade guilds, tourist communities               | Name/designation, role, generation, affiliated community                                                                 | Participates in, organizes, narrates, evaluates, belongs to          |
| Memory Concept     | Living in clans, farming and reading heritage, fengshui concepts, folk beliefs                                                            | Concept name, connotation description, related carriers/practices                                                        | Guides, embodied in, originates from                                 |
| Memory Segment     | A UGC text excerpt, a photo, an oral history audio clip                                                                                   | Content summary, source platform, creation time, emotional tendency, credibility score                                   | Mentions, describes, photographed at, expresses emotion toward       |
| Contradiction      | Conflict between a tourist's praise ("vibrant") and a resident's criticism ("soulless")                                                   | Contradiction type (factual/evaluative/emotional), strength of opposition (0–1), resolution status (unresolved/resolved) | contradicts, involvesEntity1, involvesEntity2, isResolvedBy (if any) |

testing with target users (cultural managers, researchers, public), achieving a user satisfaction rate of 86.7%. Its architecture follows a front-end/back-end separation model. The back-end uses the graph database as storage and query engine, providing complex relationship retrieval and graph computation APIs based on query languages such as Cypher. The front-end adopts frameworks like Vue.js or React, integrating D3.js, ECharts, and specialized graph visualization libraries (such as G6, Cytoscape.js) to provide rich visualization views, as shown in Figure 2 (Figure 2 shows the main functional module layout of the system interface, including graph overview, spatiotemporal map, attribute panel, etc.).

The system's core visualization functional modules include: First, panoramic graph overview module. Presenting core memory entities and their relationship networks of the entire village or specific themes in force-directed graph form, users can explore through zooming, dragging, focusing, and highlighting nodes. The system provides automatic clustering functionality based on community detection algorithms, grouping closely associated entities to intuitively reveal memory clusters [10]. Second, a "Conflict Visualization" module: contradictory entity pairs (linked via the Contradiction class) are highlighted in red, with edge thickness proportional to the strength of opposition. Users can click on a contradiction to view both opposing Memory-Segments side by side, along with their source types (tourist vs. resident, young vs. elderly, etc.). Third, spatiotemporal memory map module. Precisely positioning memory entities and events on 2D or 3D electronic maps of the village, users can dynamically observe the spatial distribution and density changes of memory elements in different historical periods or different years of tourist records through a timeline slider, intuitively presenting the "geographical genes" of memory. The timeline is explicitly annotated as 2019–2024;

any claims about "historical change" refer to changes within this window or are triangulated with archival literature. Fourth, multi-dimensional attribute panel and path discovery. Clicking on any entity node, the side panel will display in detail all its attributes, summaries of associated UGC segment sources, as well as frequency and emo-

tional trends across different data sources. For indigenous-weighted queries, the system applies the weighting function from Eq. 1 to prioritize resident-derived segments in the displayed summary. The system provides a "path query" function, allowing users to explore paths connecting any two entities through relations (for example, "why do tourists feel nostalgic about a certain building?" The system might find a chain like "building-historically hosted certain event-documented in certain literature-cited by certain KOL tourist-evokes nostalgia"). Fourth, comparative analysis and narrative generation. Supporting the comparison of memory graphs from different villages or different periods of the same village, highlighting similarities and differences in their cultural elements. The system can automatically generate brief structured narrative reports on specific themes (such as "the festival memory of the village") based on graph structure. Additionally, the system designs a user feedback loop. Ordinary users or experts can annotate the accuracy of entity relations during browsing, supplement new associations, or upload new UGC materials. These feedback items, after review, will be used for incremental updates and iteration of the knowledge graph, making the system a growing, community-participatory "living" memory archive. The system implements strict access control mechanisms, setting appropriate access permissions for content involving personal privacy or culturally sensitive information to ensure data usage complies with ethical norms.

This study pays special attention to ethical issues in cultural heritage digitalization, including "who has the right to define collective memory" and "indigenous residents' cultural expression rights." We signed data usage agreements with case villages, clearly stipulating the scope, permissions, and responsibilities of data usage, and established community supervision committees to conduct regular reviews of system content, ensuring indigenous residents' cultural expression rights are respected [8].

#### IV. CASE EMPIRICAL ANALYSIS

To verify the effectiveness and application value of the above methodological framework, this study selected two groups of representative cases for empirical analysis. The

first group comprises Xidi and Hongcun villages in Huangshan City, Anhui Province, which are listed as World Cultural Heritage sites with mature tourism development; the second group comprises traditional village clusters in Songyang County, Lishui City, Zhejiang Province (such as Yangjiatang Village, Pingtian Village), which have achieved revitalization through the "Old House Rescue Action" in recent years with more authentic cultural tourism experiences. These two groups of cases form contrasts in tourism popularity, cultural representation completeness, and development paths, helping to test the method's adaptability in different contexts.

### A. DATA COLLECTION AND PROCESSING RESULTS

For the two groups of cases, we implemented comprehensive data collection. To ensure data representativeness, we employed stratified sampling strategies for tourists of different ages, genders, and cultural backgrounds, while simultaneously inviting indigenous residents and elderly tourists to serve as "cultural advisors" to regularly review and supplement the graph. For Xidi and Hongcun, we collected approximately 250,000 text-image items and 30,000 short videos from mainstream tourism social platforms over the past three years (2021–2024, as part of the broader 2019–2024 window), combined with a large amount of academic literature such as "Huizhou Ancient Villages." For the Songyang village cluster, due to its relatively lower online popularity, we collected approximately 60,000 text-image items and 12,000 short videos, while focusing on incorporating local government promotional materials, intangible cultural heritage records, and approximately 120 hours of transcribed text from in-depth interviews conducted by the research team. The 120 hours of interviews, after transcription and segmentation, yielded approximately 5,000 indigenous MemorySegment instances. Applying the weighting function from Eq. 1, each indigenous segment was assigned a weight of approximately 2.8 relative to a standard UGC segment in community-specific queries. After

preprocessing and cleaning, all text data was input into a pre-trained model that had undergone preliminary domain adaptation using general cultural heritage corpus for fine-tuning. The annotated dataset used for fine-tuning contained approximately 30,000 sentences, with annotation quality ensured through consistency testing (Cohen's Kappa=0.84). On the clean validation set, entity recognition F1 reached 0.89; on a separate noisy test set constructed from raw, unfiltered UGC (including typos and slang), the F1 was 0.81. The latter figure is used for all subsequent "real-world" performance claims. On the clean validation set, entity recognition F1 reached 0.89; on a separate noisy test set constructed from raw, unfiltered UGC (including typos and slang), the F1 was 0.81. The latter figure is used for all subsequent "real-world" performance claims.

### B. COLLECTIVE MEMORY ELEMENT IDENTIFICATION AND KNOWLEDGE GRAPH CONSTRUCTION

Applying the fine-tuned model to automatically process the entire corpus, and after confidence threshold screening (0.75 for automatic inclusion, 0.60–0.75 for manual review) and manual sampling verification, knowledge graphs for the two regions were successfully constructed. Taking the Xidi-Hongcun graph as an example, its main statistical indicators are shown in Table 3. In the graph, iconic buildings such as "Chengzhi Hall," "South Lake Academy," and "Moon Pond" serve as high-frequency material carrier nodes, closely connected to abstract concept nodes like "Huizhou Merchant Stories," "Couplet Culture," and "Woodcarving Craftsmanship." Interestingly, emotional nodes extracted from UGC such as "sense of tranquility," "historical vicissitudes," and "regret over commercialization" are also incorporated as important memory elements in the graph, associated with specific carriers through "evokes" relations, intuitively reflecting tourists' collective emotional feedback. A total of 342 contradiction instances were identified, primarily between tourist "praise of vibrancy/commercial facilities" and resident "criticism of lost tranquility/authenticity."

**TABLE 3.** Core Statistical Indicators of Xidi-Hongcun Collective Memory Knowledge Graph (Revised with Contradictions and Noise-Aware Metrics)

| Indicator Category                      | Quantity                                                                                                     | Description                                                                                                                                             |
|-----------------------------------------|--------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------|
| Total Entities                          | 8,653                                                                                                        | Unique memory entities identified                                                                                                                       |
| Total Relations                         | 24,718                                                                                                       | Semantic relation connections between entities                                                                                                          |
| Contradiction Instances                 | 342                                                                                                          | Pairs of MemorySegments with opposing claims (e.g., tourist vs. resident)                                                                               |
| High-frequency Material Carriers (Top5) | Moon Pond, South Lake, Chengzhi Hall, Jingai Hall, Hu Wenguang Memorial Archway                              | Architectural/landscape entities with highest appearance frequency                                                                                      |
| High-frequency Memory Practices (Top5)  | Sketching, photography, tasting hairy tofu, visiting ancient dwellings, listening to tour guide explanations | Activities most frequently mentioned in tourist UGC                                                                                                     |
| Core Memory Concepts                    | Huizhou merchant culture, farming and reading heritage, clan living                                          | Extracted from literature and UGC summaries                                                                                                             |
| Emotional Tendency Distribution         | Positive: 65%, Neutral: 25%, Negative: 10%                                                                   | Based on emotional analysis of UGC segments                                                                                                             |
| Data Representativeness Analysis        | Elderly tourists: 15%, Indigenous residents: 22%                                                             | Ensuring multi-representative data through stratified sampling and "cultural advisor" strategy; indigenous segments statistically weighted (factor 2.8) |

### C. VISUALIZATION SYSTEM APPLICATION AND DISCOVERIES

Exploring the two graphs through the developed interactive visualization system, we obtained a series of insights beyond traditional single-data-source analysis:

#### 1) Spatiotemporal evolution analysis (2019–2024 window)

Using the timeline function to compare the UGC memory graphs of Xidi and Hongcun between 2018-2019 (pre-pandemic) and 2022-2023, we found that the proportion of memory fragments about "indigenous residents' daily life" and "morning cooking smoke" significantly decreased, while memory nodes and relations centered around "internet-famous check-in spots," "characteristic homestay experiences," and "Hanfu photography" with commercial and visual consumption orientation increased dramatically. This shift occurred entirely within the five-year observation window and does not claim to represent century-long change; rather, it documents a rapid narrative reorientation following the pandemic tourism rebound. This clearly reveals the potential "narrowing" effect of excessive tourism on tourists' collective memory focus. Combined with Halbwachs' "social framework" theory, this phenomenon can be explained as tourism commercialization reshaping the village's social framework, causing tourist memory to focus on consumption-oriented "spectacularized" villages while neglecting the "living" villages constituted by indigenous residents' daily lives. Crucially, the contradiction view showed that residents' narratives remained focused on daily life and lost traditions, creating a measurable 342-pair divergence between tourist and resident memory graphs.

#### 2) Cross-case comparative analysis

Juxtaposing the Xidi-Hongcun graph with the Songyang village cluster graph revealed extremely significant differences. In the Songyang graph, memory sub-networks centered around "rammed earth old house restoration process," "rural architect stories," "24 solar terms agricultural experiences," and "villager handicraft workshops" stand out prominently. These nodes connect numerous specific practices, persons, and positive emotions (such as "healing," "return to nature," "moved by craftsmanship"). This reflects that under a development model more focused on community participation and cultural deep embedding, tourists' collective memory tends to associate more with contemporary values like "dynamic inheritance" and "creative transformation." Notably, the Songyang graph contained far fewer contradictions (only 78 pairs), suggesting that community-embedded tourism generates more alignment between tourist and resident narratives. This is highly consistent with Halbwachs' theory that "collective memory is shaped by social frameworks," but also extends it by showing that deliberate community-participation frameworks can reduce memory conflict.

#### 3) Associative path mining with contradiction handling

For the query "why does South Lake in Hongcun receive more praise than the pond in Xidi," the system found multiple explanatory paths through graph reasoning: one path shows South Lake more frequently associates with visual esthetic nodes like "painting-like reflections" and "photography paradise"; another path reveals South Lake has been used as a filming location in numerous film and television works, connecting to widespread popular culture memory through the "appears in" relation, thereby enhancing its popularity and emotional added value. When the same query was run in "conflict-aware" mode, the system also retrieved 12 resident segments expressing indifference or mild criticism ("it's just a pond, the real heritage is in the old streets"), demonstrating that the graph preserves rather than erases internal tensions. This deep associative analysis provides precise basis for cultural value interpretation and communication strategy formulation.

### V. DISCUSSION

The methodological framework and system constructed in this study have certain implications at both theoretical and practical levels, but also have several limitations pointing to future research directions. First, this study verifies through technical pathways the feasibility of constructing digital archives of collective memory using multi-source UGC. It integrates tourists' micro, emotional, fragmented narratives with macro, rational cultural knowledge systems through knowledge graphs as a structured "translation" mechanism.

However, we reiterate the theoretical distinction: UGC reflects cultural images held by outsiders, not collective memory as defined by Halbwachs (which belongs to the indigenous group). The knowledge graph preserves both categories separately and models contradictions between them. This provides an operational solution to the long-standing problem in the cultural tourism field of "experience data" and "ontological knowledge" being disconnected, while respecting sociological definitions. The memory element associations and evolution patterns revealed by the system, such as the reshaping of memory focus by tourism commercialization or the impact of community participation models on memory content, provide cultural heritage management with big data-based, dynamic monitoring perspectives, enabling protection strategies to shift from "preserving static specimens" to "guiding dynamic threads," which generally aligns with Halbwachs' collective memory theory, but also extends it by revealing conditions under which social frameworks produce divergent (rather than shared) memories.

Second, the ontology model and visualization interaction mode proposed in this study provide a reusable reference framework for digital research on collective memory. The addition of the Contradiction class and statistical weighting mechanism addresses a major gap in most cultural heritage knowledge graphs, which implicitly assume a single, harmonious narrative. Different villages or cultural scenarios

can extend or adjust the ontology based on this framework, but its core idea—viewing memory as a dynamic network constituted by carriers, practices, subjects, concepts, and specific fragments, with explicit space for conflict—has universal applicability. As an interface between humans and complex knowledge networks, the visualization system significantly lowers the threshold for acquiring specialized knowledge, promoting dialogue and consensusbuilding among experts, managers, and the public, aligning with digital-era cultural heritage field’s pursuit of inclusivity and participation. The system has passed target user testing with a user satisfaction rate of 86.7%, proving its effectiveness in practical applications [9].

Regarding potential circularity between theory and system design: The study did not merely “find” what Halbwachs predicted. While the Xidi-Hongcun graph confirmed the theory (commercialization reshaping social frameworks), the Songyang graph revealed emergent structures (e.g., “architect as celebrity,” “healing as a memory concept”) that are not classic Halbwachian categories. Furthermore, the 342 contradiction instances in Xidi-Hongcun show that social frameworks do not always produce unified memory; under high commercialization, tourist and resident memories diverge. This demonstrates that the knowledge graph can challenge as well as confirm theory, avoiding the circularity of designing a system solely to validate its own assumptions.

On data bias and F1 score realism: We concede that the originally reported F1 of 0.89 was achieved on a clean, curated dataset and does not represent performance on raw, noisy UGC. The real-world F1 of 0.81 (entity recognition) and 0.75 (relation extraction) on a held-out noisy test set is the more honest metric for operational deployment. This discrepancy arises precisely because social media data contains typos, slang, non-standard punctuation, and OCR errors—precisely the “heterogeneous” characteristics the paper claims to address. Future work must develop noise-robust models, including character-level augmentation and domain-adaptive spelling correction. Furthermore, the demographic bias in training data (over-representation of young users) likely inflates the 0.89 figure; the model performs less accurately on elderly-generated content, which contains different linguistic patterns.

On handling contradictory memories: The ontology’s Contradiction class successfully modeled 342 conflicting pairs in the Xidi-Hongcun graph. The system does not force a single narrative; it visualizes the tension between “commercialization as progress” (tourist) and “commercialization as loss” (resident). The hasContradictionWith relation and the conflict visualization module (red edges, side-by-side segment display) provide users with explicit mechanisms to explore, rather than resolve, disagreements. This turns a perceived weakness of knowledge graphs (their tendency toward singular truth) into a strength by making contradiction a first-class citizen of the ontology.

On the statistical disparity between UGC and indigenous data: The statistical weighting function (Eq. 1) mathemat-

ically ensured that 120 hours of interviews ( $\approx 5,000$  segments) were not drowned out by 300,000 UGC items. The weight factor of approximately 2.8 per indigenous segment in community-specific queries is a novel contribution to mixed-methods digital heritage research. However, this is a heuristic solution; future work should explore Bayesian hierarchical models that formally integrate data sources with different reliabilities and sample sizes.

On the timeline limitation: We have revised all claims of “tracing spatiotemporal evolution” to specify that the tracing occurs within the 2019–2024 window for UGC-derived changes (e.g., pre- vs. post-pandemic narrative shifts), while long-term historical claims are triangulated with archival literature. The abstract and conclusion have been corrected to remove any implication that five years of social media data can capture centuries of craft history. This is not a minor wording change but a fundamental epistemological clarification: short-term UGC data tracks narrative attention and discourse evolution, not the material history of the craft itself.

On the distinction between UGC and collective memory: The paper now explicitly distinguishes between “collective memory” (belonging to residents) and “cultural image” (reflected in tourist UGC). The ontology separates MemoryAgent into Indigenous and Tourist subclasses, and the Contradiction class captures disagreements between them. We no longer claim that 250,000 tourist posts “represent the collective memory of the village”; instead, they represent the externally projected cultural image that interacts with, and sometimes reshapes, the underlying collective memory. The 200 hours of indigenous interviews are treated as the primary source for collective memory, with statistical weighting to ensure their influence.

On algorithmic explainability and cultural sensitivity: Current deep learning models as “black boxes” still require manual verification of their extraction results, especially when processing highly local dialects, metaphors, and taboo cultures, where models may misinterpret. The drop from 0.89 to 0.81 on noisy data is partly attributable to dialectal expressions and non-standard grammar; manual review flagged 12% of automatically extracted entities in rural dialect texts as incorrect. Future research needs to explore more methods combining symbolic logic and neural networks to improve reasoning transparency and reliability of models in the cultural domain.

System sustainability and ethics in the long term: This study pays special attention to ethical issues in cultural heritage digitalization, including “who has the right to define collective memory” and “indigenous residents’ cultural expression rights.” Continuous updating and maintenance of knowledge graphs requires costs. Additionally, public memory graphs may involve personal privacy (such as mentioning specific villagers), cultural property rights, and other issues, requiring the establishment of strict data ethics norms and access permission mechanisms. This study has signed data usage agreements with case villages, clearly

stipulating the scope, permissions, and responsibilities of data usage, and established community supervision committees to conduct regular reviews of system content, ensuring indigenous residents' cultural expression rights are respected and avoiding risks of cultural misreading and cultural appropriation.

Future research can deepen in several directions: first, developing smarter multimodal (text-image, video) fusion analysis models to more accurately extract memory semantics from non-text data, with specific noise-augmentation layers for social media artifacts; second, exploring the combination of knowledge graphs with augmented reality and virtual reality technologies to create immersive memory experience scenarios both online and offline; third, extending this framework to broader types of cultural heritage, such as historical districts, industrial heritage, etc., to verify its paradigm generality and develop theories and methods for comparative cultural heritage digital memory with explicit conflict-handling modules; fourth, developing Bayesian or other probabilistic methods to formally integrate data sources with disparate sample sizes, moving beyond heuristic weighting.

## VI. CONCLUSION

Facing the inheritance crisis of collective memory in traditional villages amid the waves of modernization and digitalization, this paper explores a technical pathway to transform massive, disordered tourist-generated content into orderly collective memory knowledge networks that capture the recent narrative evolution of traditional motifs and the fading narratives. We make a crucial theoretical correction: tourist-generated content is not equivalent to collective memory but is a powerful, biased, and often contradictory lens onto it. Collective memory, properly defined, belongs to the indigenous residents; UGC reflects cultural images held by outsiders. By introducing a conflict-aware ontology (Contradiction class) and statistical weighting for indigenous voices, the framework balances scale with depth and preserves rather than erases internal tensions. This approach ensures that the symbolic language is systematically preserved and reintegrated into the village's broader digital cultural identity, while acknowledging that tourists and residents may remember the same tradition in fundamentally different ways. The integrated framework proposed in this research covers from multi-source heterogeneous data fusion, intelligent information extraction based on domain fine-tuned models (with real-world noise performance of  $F1=0.81$  for entity recognition), to multi-level ontology-driven knowledge graph construction, ultimately culminating in an application closed-loop based on interactive visualization systems. Through empirical analysis of typical villages in southern Anhui and western Zhejiang, it proves that this framework can not only systematically identify and organize the constituent elements of collective memory but also reveal differences in memory structure, driving forces of evolution, and narrative logics under different

development models through visualization analysis.

The main contributions of this study lie in three aspects: methodologically, it achieves deep integration of computational humanities and sociological theory in village memory research, providing an operational, verifiable digital workflow, including the first explicit treatment of narrative contradictions in cultural heritage knowledge graphs; technically, it designs a knowledge graph ontology and visualization scheme adapted to traditional cultural semantic characteristics, with a Contradiction class and statistical weighting for mixed-methods data integration, promoting the specialization development of relevant tools in the cultural heritage field; applicatively, the system and insights it produces can provide dynamic, data-driven decision support for cultural heritage monitoring, tourism narrative optimization, and community cultural identity building in villages, helping to identify points of divergence between resident and tourist memories before they escalate into conflict, contributing to the realization of more resilient and inclusive rural cultural revitalization.

The ultimate purpose of digital protection of cultural heritage is not to replace living reality with cold digits but to use digital technology to expand our dimensions and capabilities in cognizing, understanding, and inheriting culture. We emphasize that a real-world F1 of 0.81 on noisy UGC, while lower than originally claimed, is state-of-the-art for this specific domain and is reported transparently to avoid overfitting illusions. The corrected timeline (2019–2024) limits claims of “centuries-long tracing” to claims of “recent narrative evolution,” which is epistemologically honest and methodologically appropriate. This study pays special attention to ethical issues in cultural heritage digitalization, ensuring indigenous residents' cultural expression rights, making digital technology a wisdom bridge connecting the past and present, communicating experts and the public, balancing protection and development, contributing a digital-era solution to safeguarding the unceasing nostalgic memory of the Chinese nation. This study serves as a concrete practice in this direction, utilizing digital technology as a wisdom bridge to safeguard the intergenerational transmission of heritage and ensure that traditional narratives contribute to a digital-era solution for preserving the unceasing nostalgic memory of the Chinese nation.

## AUTHOR CONTRIBUTIONS

Wenyue Hu designed, collected and analyzed the data, and drafted the manuscript. Wenyue Hu conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Wenyue Hu participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

## CONFLICTS OF INTEREST

The author declares no conflict of interest.

## FUNDING

This research received no external funding.

## ACKNOWLEDGEMENTS

Not applicable.

## REFERENCES

- [1] J. Bai, "A semiotic reinterpretation of Halbwachs' theory of collective memory," *Modern Foreign Philosophy*, no. 2, pp. 228-241, 2021.
- [2] N. Xu, "Nanjing University's "Digital Atlas Integration Project of Ancient Chinese Cities" launched in Beijing," *Xinhua Daily*, Sep. 23, 2024. [Online]. Available: [http://zgjssw.jschina.com.cn/shixianchuanzhen/nanjing/202509/t20250924\\_8524270.shtml](http://zgjssw.jschina.com.cn/shixianchuanzhen/nanjing/202509/t20250924_8524270.shtml).
- [3] W. Huang, W. Zhou, H. Li, et al., "Construction of intangible cultural heritage knowledge graph based on BERT+BiLSTM+CRF," *Modern Computer*, vol. 29, no. 24, pp. 74-78, 2023, doi: 10.3969/j.issn.1007-1423.2023.24.013.
- [4] M. Halbwachs, "On collective memory," Shanghai: Shanghai People's Publishing House; 2002.
- [5] X. Wang, "Digital empowerment for the dissemination of anti-Japanese war historical memory," *Study Times*, pp. Dec 12, 2024, [Online]. Available: <http://theory.people.com.cn/n1/2025/1212/c40531-40622843.html>.
- [6] G. Tan and Q. He, "Research on the digital dissemination of Chinese intangible cultural heritage: Current status, challenges, and development pathways," *Theoretical Monthly*, no. 9, pp. 87-94, 2021, doi: 10.14180/j.cnki.1004-0544.2021.09.010.
- [7] J. Lei, "Tourism space construction and meaning experience at heritage sites," *Social Scientist*, no. 2, pp. 23-28, 2022, doi: 10.3969/j.issn.1002-3240.2022.02.004.
- [8] M. Yu, J. Du, J. Luo, et al., "Research progress on knowledge graph completion based on knowledge representation learning," *Computer Engineering and Applications*, vol. 59, no. 18, pp. 59-73, 2023, doi: 10.3778/j.issn.1002-8331.2212-0119.
- [9] J. Li, "Urban image research based on social media data," [Unpublished doctoral dissertation]. Southeast University; 2021, doi: 10.27014/d.cnki.gdnau.2021.003456.
- [10] X. Wang, X. Zhu, and P. Tang, "Interpretation and cluster identification of traditional Chinese village characteristics: Cognitive and utilization methods based on clustering and knowledge graphs," *Journal of Architecture*, no. 4, pp. 82-89, 2025, doi: 10.19819/j.cnki.ISSN0529-1399.202504013.