LiberRoad: Probing into the Journey of Chinese Classics through Visual Analytics

Yuhan Guo Yuchu Luo Keer Lu Linfang Li Haizheng Yang Xiaoru Yuan

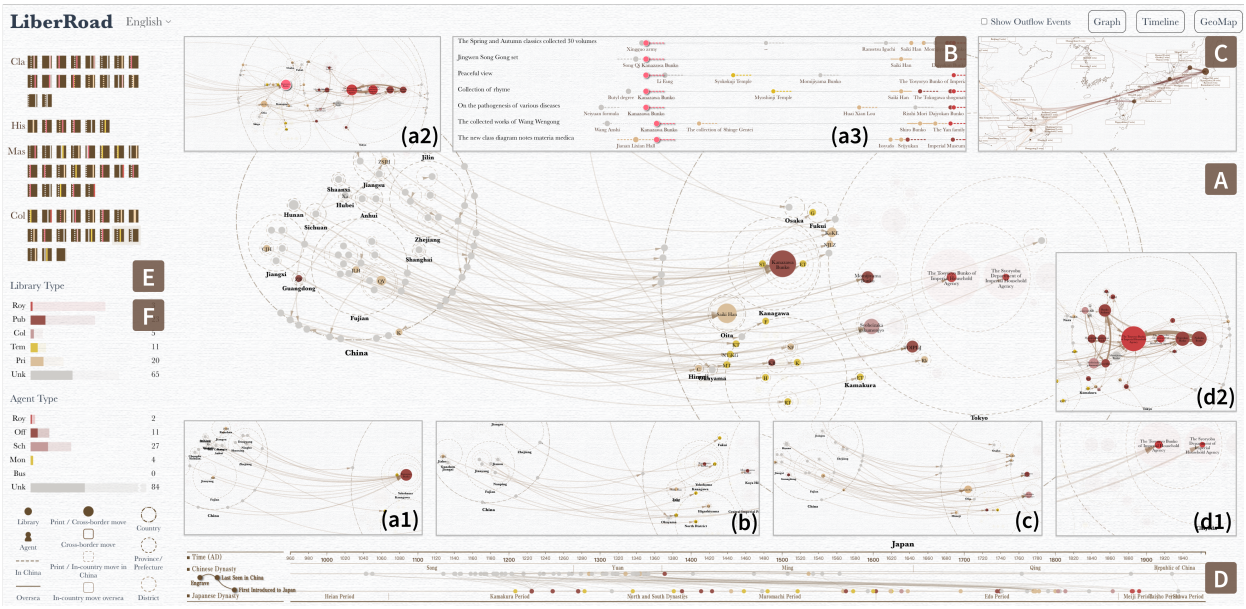

Fig. 1. The interface of LiberRoad consists of three main views and auxiliary views. The main view transforms between the Location Graph (A), the Event Timeline (B), and the Geomap (C). The other views are the time axis (D), the book list (E), and legends (F). In a usage case that analyzes the circulation of Chinese classics from China to Japan through history, experts filter on different time periods. (a) the Kamakura period, (b) the Muromachi period, (c) the Edo period, (d) the Meiji period.

Abstract— Books act as a crucial carrier of cultural dissemination in ancient times. This work involves joint efforts between visualization and humanities researchers, aiming at building a holistic view of the cultural exchange and integration between China and Japan brought about by the overseas circulation of Chinese classics. Book circulation data consist of uncertain spatiotemporal trajectories, with multiple dimensions, and movement across hierarchical spaces forms a compound network. LiberRoad visualizes the circulation of books collected in the Imperial Household Agency of Japan, and can be generalized to other book movement data. The LiberRoad system enables a smooth transition between three views (Location Graph, map, and timeline) according to the desired perspectives (spatial or temporal), as well as flexible filtering and selection. The Location Graph is a novel uncertainty-aware visualization method that employs improved circle packing to represent spatial hierarchy. The map view intuitively shows the overall circulation by clustering and allows zooming into single book trajectory with lenses magnifying local movements. The timeline view ranks dynamically in response to user interaction to facilitate the discovery of temporal events. The evaluation and feedback from the expert users demonstrate that LiberRoad is helpful in revealing movement patterns and comparing circulation characteristics of different times and spaces.

Index Terms—Visual analytics, digital humanities, spatial uncertainty, trajectory visualization, book movement, historical data

1 Introduction

Books serve as a medium for cultural exchanges between different regions. The spread of Chinese classics in Japan has had a profound impact on Japan's academic and cultural development. The prefaces, postscripts, annotations, and seals upon the aged leaves record the footsteps of the classics as they traveled from China to Japan, traversing time and space, which have become vital clues to trace the spread and integration of discrepant civilizations through history. Traditional humanities studies focus on specific books or collections, meticulously examining every aspect of their circulation and preservation, thereby shedding light on the cultural dynamics within particular temporal and spatial contexts in Japan. Nowadays, there is a growing desire among humanities scholars to piece together the scattered exchange episodes, taking a panoramic view of the cultural history of both countries.

This work involves joint efforts between visualization researchers and humanities scholars specializing in overseas Chinese classics. Our goal is to assist Chinese classics experts to develop a complete under-

- *Yuhan Guo, Yuchu Luo, Keer Lu, and Xiaoru Yuan are with Key Laboratory of Machine Perception (Ministry of Education), School of AI, Peking University. E-mail: {yuhan.guo, luoyuchu, xiaoru.yuan}@pku.edu.cn, keer.lu@outlook.com. Xiaoru Yuan is also with National Engineering Laboratory for Big Data Analysis and Application, Peking University. Xiaoru Yuan is the corresponding author.*
- *Linfang Li and Haizheng Yang are with the Center for Ancient Chinese Classics & Archives and the Department of Chinese Language & Literature, Peking University. E-mail: {llfang, haizheng}@pku.edu.cn.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

standing of the spatiotemporal circulation data in a visual manner and to explore new research questions regarding the history of China-Japan cultural exchanges. Our humanities collaborators have gathered the circulation trajectories of sixty-four representative classics collected in the Syoryobu Department of Imperial Household Agency, which cover a long time span from the 10th century to the 20th century.

The study of book circulation is gaining momentum among humanities scholars around the world and has increasingly applied digital techniques. A good example is the establishment of the MEI [1], a database that records the provenance of the book trade in the fifteenth century. The database developed the BookTracker [61] to present the spatial movement of queried books on a scalable map. While BookTracker emphasizes geographic trajectories and data retrieval ability, we found that traditional map visualizations are not sufficient for visualizing the multi-faceted circulation data to support in-depth analysis such as pattern identification. On the one hand, the 2D map does not show temporal attributes, which is important for reflecting the cultural characteristics of different eras. On the other hand, the historical book movement data is unique in its significant uncertainty and heterogeneity. Each institution along the trajectories bears distinct cultural meanings, requiring a fine-grained and accurate presentation of movement patterns, which is hard to be achieved on a map due to clutter caused by non-even spatial distribution. To our knowledge, there is no existing literature that provides comprehensive visual analytics approaches for the specific and unique book circulation data.

To fill this gap, we propose LiberRoad from the interdisciplinary perspective of humanities and visualization. Domain experts are concerned about the detailed circulation history of a single book, the relation and communication between entities, as well as collection features in different eras, locations, and institutions, which embody rich scholarly features and cultural exchange. Based on the tasks identified from user interviews with domain experts, we propose the triple view design that equips experts with both temporal and spatial perspectives. The Event Timeline view and the Geomap view stress the temporal and spatial attributes respectively. While these common event sequence and trajectory visualization methods provide an intuitive presentation, it is difficult to observe the overall patterns (e.g., convergence, divergence, and propagation) on the detailed timeline or highly aggregated map. The Location Graph view is proposed to reveal circulation patterns taking into consideration the spatial uncertainty at different levels, where institutions are represented with circles packed into hierarchical locations or on the border of the minimum level of certain regions. Flexible selection and filtering, together with a smooth transition between the three views, enable experts to explore multiple facets interactively.

We introduce two cases with humanities research significance to demonstrate the effectiveness of LiberRoad in supporting high-level domain research tasks, which is also proved by the interview with domain experts. A user experiment is conducted to figure out the effectiveness of each view in various tasks as well as verify the system's ability to help users from both humanities backgrounds and technical backgrounds understand the basic characteristics of the circulation data. The main contributions of this paper include:

- The design space of book circulation studies through visual analytics, providing a new perspective for the domain research.
- The Location Graph, a novel uncertain-aware visualization method for geospatial data via an abstractive hierarchical graph.
- LiberRoad, a visual analytics system that equips users with comprehensive spatial and temporal perspectives by combining the Location Graph view, the Event Timeline view, and the Geomap view with smooth transition.

2 Related Work

In this section, we review the book circulation study and trajectory visualization. Since the boook circulation entails temporal event sequences with significant uncertainty, we also review event sequence visualization and uncertianty visualization.

[1]The 15cBOOKTRADE Project website (accessed in June 2023). https://15cbooktrade.ox.ac.uk/.

2.1 Book Circulation Study

In this work, the term "Chinese classics" refers to the classics that are located outside of China, such as extraterritorial editions, and ancient Chinese books that have been disseminated abroad, which have promoted the dissemination of Chinese culture [35]. In general, the research subjects of the circulation of Chinese classics involve important works such as the Analects [33] and the Records of the Grand Historian [62, 64], extending to other book categories. The scope of the circulated Chinese classics ranges from East Asia to Europe [34, 63], North America [65], and other regions. Baed on the dissemination of Chinese classics, researchers also pay attention to academic history and cultural history issues, such as the relationship between the dissemination of Chinese classics and the development and changes of academic thought, and their cultural influences [52]. Important concepts such as "the road of Chinese classics" have been proposed [45, 53]. The lack of standardized databases and effective analytical tools has been hindering domain experts from in-depth analysis of large-scale circulation from a holistic view. In this work, we introduce visual analytics techniques for the first time into the Chinese classics circulation domain.

With the development of digital technology, scholars in various fields have become increasingly aware of the importance of collecting data on the dissemination of historical books. Worldwide, databases such as MEI [11] have been established. To assist domain experts in browsing book movement data, BookTracker [61] has visualized the user-queried book trajectories in the MEI database with a 2D map emphasizing the geographic perspective of the movement data. In contrast, our work targets specific domain problems and proposes high-level analysis tasks, such as discovering movement patterns and comparing features across different spatiotemporal dimensions.

2.2 Trajectory Visualization

Trajectory visualization is widely studied in the past decades. Many different kinds of methods and techniques are proposed to tackle different level of tasks [5, 17]. According to Andrienko [2], trajectory visualization methods can be taxonomized into three categories, direct depiction, summarization, and pattern extraction. The direct depiction method is to visualize the trajectory directly in 2D or 3D space [28, 29, 48]. Many visualization and interaction techniques can be used to augment the direct depiction of trajectories, like animation [24], texture [56], lens [47]. While direct depiction suffered from clutter and other scalability problems, it is intuitive and suitable for inspecting details, which is adopted in our system providing the detail view.

The summarization methods aggregate the trajectory data. like spatial aggregation [67], flow aggregation [23], and path aggregation [1]. In our task, episodic [6] trajectory data and the factor that scholars pay more attention to the conceptual geographical location rather than precise geographical position push us to choose more abstract aggregation methods, the geo-network [44]. Easy-to-use tools, such as Vistorian [7], have been developed for humanities scholars to create geo-network visualizations. More abstractly, the trajectories can be viewed as dynamic hypergraphs [19, 49]. However, the uncertainty and hierarchical structure of the spatial information make it challenging to design the aggregation visualization for trajectories.

The pattern extraction methods distill higher-level semantic information from data. For example, important events or places [3, 4] are detected before being displayed to users. Or crucial derived attributes are calculated and demonstrated to users like proximity between objects [14] or curvature and turning significance of routes [55]. While pattern extraction methods lose many details in data and require a predefined analysis pipeline, they are not suitable for our task, which has an immature research pipeline with heavy dependency on data details and background knowledge.

2.3 Visual Event Sequence Analysis

An event sequence is a series of discrete events in the temporal order. Visualization techniques have been applied to tell the story of specific event sequences [8] or facilitate analysis of multiple sequences such as identifying patterns, predicting events, and comparing features. A straightforward way to visualize event sequences is to draw directly

on timelines [43], which could be enhanced by flexible interactions to adapt to analysis tasks. Concerning the sequence causality, Lifelines align the same events to observe the factors related to certain diseases [50]. Lens magnification supports Focus+Context exploration on temporal sequences with large and dynamic event data [31]. When the number of sequences gets large, aggregation is used to provide an overview and reveal patterns. Sequence-level aggregation organizes the timelines into hierarchies, [10, 60], which has been improved with sophisticated algorithms and interactions [21, 38]. Event-level aggregation focuses on the subsequence patterns, usually grouping the events with similar types together and showing the transfer between different states, such as the Sankey diagram [42] and its extensions [59]. Besides Sankey diagrams, MatrixWave [68] adopts zig-zag matrices to visualize transfers between massive states, and Di Bartolomeo et al. [16] aligned sentinel events to present high-level patterns. Another way of clustering is based on the automatic calculation of similarity between sequences to shed light on potential patterns [25].

Though effective in discovering patterns, the above aggregation methods do not apply to our data in that the book movement sequences are highly heterogeneous, which cannot be simply aggregated according to the roughly classified types. For example, libraries of the same type are different in academic characteristics. Books that have been collected by the same type of libraries or agents probably have distinct cultural impacts, which should be distinguished. Storylines, which bend the timelines and portray the relationship between entities over time while preserving individuality, are more adapted to our data [32, 39, 46]. However, considering the relatively large scale of books and accessibility to humanities scholars, we decided to use the straight timelines enhanced by dynamic ranking to present relationships.

2.4 Uncertainty Visualization

Early in the last century, there has been a lot of discussion on and approaches to uncertainty in scientific visualization, which comes from data acquisition (e.g., measurements), transformation (e.g., sampling), and visualization (e.g., rendering) [41]. Uncertainty in data can seriously affect analysis and decision-making [57], which should be considered in visualization seriously [9, 15]. In humanities study, data tends to encompass inherent uncertainty due to the loss of historical evidence. Recently, Panagiotidou et al. [40] have proposed a taxonomy for uncertainty origins and classification of coping strategies in digital humanities visualization. Uncertainty in book circulation involves dubious spatiotemporal information. The multi-degree and non-statistical nature of uncertainty in humanities data makes it challenging to directly apply existing methods in information visualization, such as visualization for uncertain trajectories [12, 54] and temporal uncertainty [22], resulting in the development of domain-oriented visual designs [30, 58] and visual reasoning [66].

In book circulation data, most of the time and space are uncertain with a wide range of possible values, making it difficult to estimate statistical models [30] or encode on maps [37]. The significance of spatial uncertainty motivates us to design the Location Graph based on the circular treemap. Circular treemaps have shown the potential to be modified to express uncertain hierarchical data in Bubble Treemaps [20], where leaf values are unknown. In our case, the values are certain, but the tree structure is incomplete.

3 Book Circulation Data

In this section, we introduce the collection and uniqueness of historical book circulation data. These data contain profound cultural significance but also possess uncertainty and multi-faceted attributes, posing challenges for visualization design.

3.1 Data Collection

One of the most important research materials among the Chinese books circulated overseas are the Song-Yuan editions, the earliest printed versions in China, which served as a significant carrier for the mass dissemination of Chinese culture through large-scale printing. However, with the long passage of time, only a few well-preserved Song-Yuan editions remained today, with only slightly over 5,000 copies existing. The Syoryobu Department of the Imperial Household Agency in Japan is the most representative institution that holds the largest and high-quality collection. Our domain collaborators manually collated the transmission events of 64 books collected in the Syoryobu Department by carefully examining inscriptions, postscripts, and bookplates within the printed versions of Song-Yuan editions of Chinese books, combined with relevant historical knowledge.

As is shown in Figure 2, each book contains basic information such as the title, category, author, version, place and time of printing. At the same time, each book includes a sequence of transmission events, with each event including the order, time, location, book repository, intermediary, and printed materials used to infer the attributes of the transmission event. Since many records in the books are incomplete and much historical information is difficult to verify or investigate due to the distant period, even with our collaborator possessing profound professional knowledge, accurate attributes of many transmission events cannot be inferred, resulting in significant uncertainty in the data.

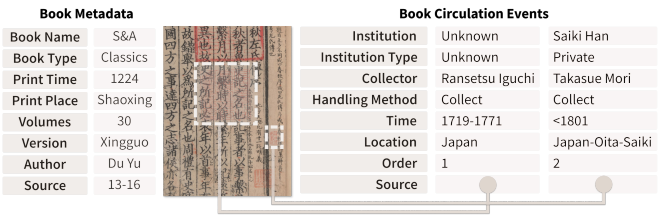

Fig. 2. The book circulation data is spatiotemporal trajectories with multiple dimensions (e.g., institutions, collectors, and their types).

3.2 Data Features

The book circulation data exhibits distinctive spatiotemporal trajectories, and its uniqueness lies in its structure, features, and significance.

Compound graph structure. The spatial perspective of the trajectories is expressed by the administrative regions and the holding institutions, rather than precise geographic coordinates. Therefore, the discrete track points falling on the institutions can be organized into a hierarchical structure based on administrative divisions. On this hierarchy, the movement of books between libraries forms a network, resulting in a compound graph structure.

Non-uniform distribution. The trajectories are uneven in time and space. Affected by political and cultural backgrounds, the circulation of books varies greatly over historical periods, with some carrying little movement while others observe frequent circulation. In terms of spatial distribution, a large number of institutions are located in limited geographical regions in Japan, while other areas are relatively sparse.

Significant uncertainty. Not every collector would stamp their seal, nor would every library leave behind a complete catalog. Even if there are some records, the attributes might be missing or ambiguous. The time periods are mostly inferred intervals, either bounded or unbounded. Some book seals belong to collectors who are not well known in history, making it difficult to infer their identity and location.

Bearing profound cultural meaning. Books are the crystallization of civilization, and the spread of books is an extension of culture. Each node in the trajectory, be it a famous collector or an unknown institution, signifies the diffusion of culture and serves as an important entry point for scholars to delve into the history of academic development as well as cultural exchange, therefore unneglectable.

4 Design Rationale

In this section, we introduce the identification of design requirements and tasks through interviews with domain experts. Based on the established tasks, we propose the triple view design of LiberRoad to support analysis tasks from various perspectives.

4.1 Preliminary Interview

We conducted interviews with five domain experts to identify the design requirements. Two of them (P1 and P2) are our collaborators, and two (P3 and P4) were researchers who study the movement of

Chinese classics. Notably, P3 is from Japan and provided valuable insights into the cultural context. Another participant (P5) is a Ph.D. student majoring in this field. During the interviews, we asked the participants the following questions. (1) What is the traditional workflow for studying the circulation of Chinese classics? (2) What aspects of the multiple book circulation data are of interest? (3) What are the research questions in the field of Chinese classics circulation? (4) How should the uncertainty of data be approached? (5) What are the design requirements you expect for a visualization system? The findings on design requirements are summarized as follows.

Emerging field embracing visual illumination. Traditional domain studies focus on specific books or a few books, tracing their movement paths. Regarding large amounts of data on book collections, researchers indicated that there is currently no mature research on large-scale Chinese classic circulation, nor is there an established workflow. As P4 said, *"Even in the humanities field, research on the history of dissemination and collection is new. Although the current project exists in the form of digital humanities, it is still new in terms of research content."* Therefore, our system should support the examination of details for individual books in the traditional workflow, while also providing informative overviews that enable the discovery of patterns and provoke potential research directions.

Emphasis on cultural significance behind the data. Humanities studies put emphasis on exposition and conclusions. Domain scholars draw upon their broad background knowledge to make inferences and deduce the historical influence of Chinese classics on the regional culture of Japan. Each facet, such as locations, institutions, and individuals, reflects an aspect of culture, which might evoke numerous associations. Different periods, places, and individuals have their own characteristics, and the introduction of Chinese classics, as well as the influence it generates, can vary significantly. It is required to present data from multiple perspectives and support scholars in exploring data from various angles for probing into cultural differences.

4.2 Task Identification

Numerous analyses and research directions emerged from rounds of discussions and interviews with the domain experts, from both the perspective of navigating data and that of new possibilities for overseas Chinese classics studies. Consequently, the visual analytics system should provide an effective visualization of the complex book circulation data as well as ample space for exploration and inspiration, while appealing intuitive and user-friendly for humanities researchers.

T1. Providing an overview. An overview of the basic statistics about the data is appreciated. Domain experts are interested in the overall collection of classics at the Imperial Household Agency, namely the number of books, the proportion of different book categories, and their printing time (P1, P4). The distribution of the libraries and the time these classics arrived in Japan is important as well.

T2. Inspecting single book circulation. All interviewers emphasized the requirement of inspecting the detailed trajectory of the book they are interested in. Each movement is of great importance. Experts will investigate the starting point, when and where it is collected, as well as the libraries and collectors it encountered along the journey. Presenting the details would *"benefit the general humanities scholars studying literature or history"* (P5) as well, supplementing their studies.

T3. Searching with facets. P2 put forward the demand for complex retrieval, hoping to query data from multiple perspectives (book type, time, location, library, agent) and combined attributes (e.g., books move from one wanted library to another). Experts not only care about a few specific books, but also a group of books of the same category, the collection of a person, the collection by libraries of the same type, the circulation within a specific period of time, and so on.

T4. Identifying patterns. Visualization offers opportunities for uncovering the intricate intersection of time, location, and collectors (P2). In visualization, the intersecting relationships can be understood as patterns of trajectories [17]. When examining book circulation, the main focus is on compound patterns, including divergence, convergence, and propagation. Books may be *"concentrated in one place during a certain period"* (P1), or they might disperse to various locations.

T5. Comparing features. The academic characteristics of different periods and regions are reflected in the categories of books and the types of libraries that collected them. Comparison of the features of book movement in different time and space *"might drive a more fine-grained temporal and regional division"* (P2) regarding cultural history.

T6. Presenting uncertainty and provenance. In view of the ambiguity of the source data, the visualization system should present where the uncertainty lies. Moreover, *"it's these unclear areas that are where to find problems and solve problems"* (P2, P3). At the same time, provenance information is required, including the original data attributes and page images from which the circulation event is extracted, in order to facilitate scholars' examination and correction.

4.3 System Overview

LiberRoad provides a triple-view design to enable a comprehensive analysis of the multi-angle, multi-level, and multi-granularity tasks. Figure 3 illustrates the design of the three views. Each view targets a specific perspective and supports low-level tasks distinctively. The smooth and seamless transition between these views, bringing together the spatial and temporal perspectives, allows the combination of top-down and bottom-up exploration, thus supporting users to accomplish high-level tasks through continuous and multi-faceted analysis.

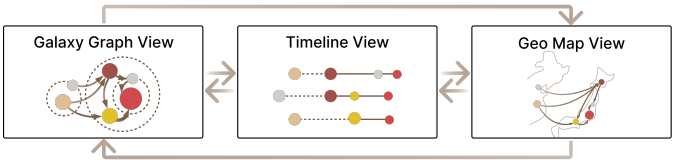

Fig. 3. Users can switch between the three views of LiberRoad with smooth transition.

The Location Graph view is a novel visualization method proposed in this work. This view aims to provide an uncertain-aware overview of the institutional-level movement (**T1**), which is powerful in revealing movement patterns such as the spatial concentration of multiple books into a collecting institution (**T4**). The Event Timeline view presents the detailed temporal circulation sequence of each book (**T2**), which is enhanced by interactive dynamic ranking to show temporal relations of movement events (**T4**). The Geomap view equips users with an intuitive geospatial perspective, providing an overview of movement between geographical regions (**T1**) as well as detailed single book trajectories (**T2**). The interactable visual elements, including editions, event nodes, institution nodes, locations, collectors, and paths, support focus on different facets (**T3**). The combination of the spatial perspective and temporal perspective allows the discovery of high-level patterns, e.g., the convergence of books into an institution or a location within a short time period (**T4**). With further selection and filtering, users can compare different patterns to analyze the collection and cultural features of different time and space (**T5**). The Event Timeline view and Geomap view display provenance information on demand, and all three views present uncertainty with elaborate design (**T6**).

It is common and intuitive to visualize trajectories with a map and visualize temporal event sequences by a timeline, which we showed to the domain experts during the initial design stages. Domain experts found the map necessary, which they are familiar with, thus corresponding to their spatial cognition and *"easily triggering the association with related domain knowledge"* (P2). The timeline is also appreciated, which represents the trajectories in a detailed manner so that users can *"clearly inspect the specific movement of each book"* (P5). However, since the unevenly distributed origins and destinations are cluttered and aggregated on the map, directly visualizing book trajectories on the map is not sufficient for revealing institutional-level circulation patterns. At the same time, the detailed and sparse timelines do not provide an overview of the overall circulation. This deficiency of traditional map and timeline visualization motivated the proposal of the Location Graph, which organizes institutions into regional hierarchies and places uncertain locations on the circle borders. The transition between the three views not only bridges the perspectives but also helps experts

establish psychological connections between traditional methods and new visualization designs, enabling an easy and wide exploration and analysis space for humanities scholars.

5 LiberRoad System

This section presents the design and methodological details of the three main views as well as auxiliary widgets of the LiberRoad system.

5.1 Location Graph View

The Location Graph provides an overview of spatial movement between institutions and locations. Due to the spatial uncertainty and disparity in the area of regions, it is non-trivial to visualize the institutional-level movement while preserving the overall context. On a map, it is necessary to zoom in extensively to see specific institutions. The locations of many historical collecting institutions are unknown, making it difficult to place them on the map in a way that produces little misinformation. While node-link diagrams directly show the nodes and edges, institutions in the same location would be dispersed in different positions on the screen, making it difficult to associate the institutions with geospatial relations. In the context of geographical information, scholars prioritize the concept of regions over real spatial distance. As a result, we abstract the movement between discrete locations into a graph, on which the locations are packed into hierarchical circles. The positions of these circles maintain the relative geographical directions. The institutions with uncertain accurate locations are placed on the border of the most precise known area.

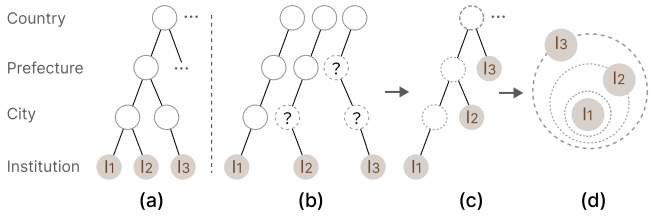

Fig. 4. Construction of the Location Graph. (a) actual locations (b) inferred locations (c) extracted hierarchy which represents different levels of uncertainty (d) certain locations are packed into nested circles and the uncertain institutions are placed on the border of the parent node.

Extracting uncertain spatial hierarchy. The spatial attributes of each trajectory include ancient locations, modern locations, and holding libraries. Due to changes in ancient administrative divisions and difficulties in standardizing their expression, we established the hierarchical structure based on modern locations. Figure 4a shows the ideal hierarchy, where the libraries are attached to the cities where they are located. However, the inferred locations have different levels of uncertainty (Figure 4b). For example, for some libraries, we know the prefecture but don't know the exact city, while for some other libraries, the prefecture is unknown as well. Figure 4c shows the reorganized hierarchy according to the known spatial information. Nodes with higher uncertainty have lower depth.

Positioning nodes. Nodes on the extracted tree can be classified into three categories: *region nodes*, *certain library nodes*, and *uncertain library nodes*. The *region nodes* refer to the non-leaf nodes (dashed circles in Figure 4c), representing the known administrative regions. The *certain library nodes* refer to the leaf nodes without non-leaf siblings, representing libraries with known locations. In contrast, the *uncertain library nodes* refer to the leaf nodes with non-leaf siblings, representing libraries with unknown locations. The *region nodes* and *certain library nodes* are positioned by circular packing, and the *uncertain library nodes* are placed on the border of their parent nodes (Figure 4d). We refer to the *region nodes* and the *certain library nodes* as *inner nodes,* and the *uncertain library nodes* as *border nodes* when introducing the layout algorithm. Taking the hierarchical constraints, graph readability, and geographic similarity into consideration, the node layout is divided into two stages, which is illustrated in Figure 5. In the first stage, the approximate positions of *inner nodes* are determined, with the goal of approximating geographic positions while maintaining node compactness. In the second stage, the positions of *border nodes* are placed with the goal of reducing path length.

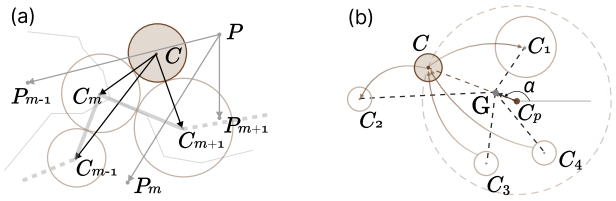

Fig. 5. The two-stage node layout algorithm. (a) In the first stage, the circles are placed in sequence. Each new circle is tangent to two adjacent circles on the front-chain [51]. The adjacent circles are selected according to the loss of the geographical angles. (b) In the second stage, the border nodes are placed on the ray from the center of its parent node to the barycenter of its adjacent nodes.

Stage 1. Place inner nodes. The *inner nodes* at the same level with the same parent are depicted as a group of circles represented by $\{(C_i, P_i, r_i)|i = 1, 2, \cdots, n\}$, where C_i denotes the center of the i-th circle, P_i denotes the geographical coordinates, r_i denotes the radius, and n is the number of nodes. We revised the circle packing algorithm proposed by Wang et al. [51] to maintain the relative angular relationships between nodes in the layout. Wang's circle packing algorithm considered both the compactness of the layout and the constraint that nodes should not overlap. In the layout process, a front chain, i.e., the sequence of outermost nodes, is retained, and new nodes are added one by one. At each time, the algorithm identifies node C_m that is nearest from the center of the given canvas and positions the new node C so that C is tangent to C_m and C_{m+1}. We modified the strategy for selecting C_m. A score function is applied to measure the average deviation of the angle between the node to be placed and the already positioned nodes. Each time, the node with the maximum score is selected as the tangent circle. The score function is defined as:

$$score(C_m) = \sum_{i=1}^{n} \frac{< C_i - C, P_i - P >}{\|C_i - C\| \cdot \|P_i - P\|} \cdot r_i,$$

where C denotes the center of the circle tangent to C_m and C_{m+1}. While in the classical circle packing algorithm, the area of the parent node maps the quantity, and the child nodes adapt to the parent node, we emphasize the role of institutions and pack the nodes in a bottom-up manner. The area of the leaf node is determined by the number of movement events targeted at it.

Stage 2. Tune border nodes. The goal of optimizing the *border nodes* is to minimize the edge length. As Figure 5 (b) illustrates, the border node is placed on the ray that connects the center of its parent node and the barycenter of its adjacent nodes. For the *border node* C, let $\{(C_i, r_i)|i = 1, 2, \cdots, n\}$ represents its neighbors, where C_i denotes the center of the i-th circle, r_i denotes the radius, and n denotes the number of neighbors. The barycenter G of the border circle C is the weighted mean of the position of the neighbors.

$$G(C) = \frac{\sum_{i=1}^{n} r_i C_i}{\sum_{i=1}^{n} r_i}.$$

Let α denotes the angle of the vector $G(C) - C_p$, where C_p refers to the center of the parent of node C. If the barycenter $G(C)$ coincides with C_p, we set α as zero. The position of circle C is calculated as $(r_p \cos\alpha, r_p \sin\alpha)$. If C intersects with previously placed border siblings, we keep adding small offsets to α until there is no intersection.

Visual Encoding. The *library nodes* are represented by filled circles, the size of which represents the number of book collections by the library, and the color represents the type of the library. The *region nodes* are dashed circles, representing the administrative boundaries and do not map the quantities of movements. The styles of the dashed borders distinguish the hierarchical levels, such as countries, prefectures, and cities. The edges are drawn with quadratic Bezier curves and the width

of the edge between two nodes encodes the number of books directly flowing from the source to the target.

5.2 Event Timeline View

The timeline view emphasizes the chronological distribution and sequential arrangement of events pertaining to each classic circulation. In this view, each book is allocated a row, movement events are represented by circles on the timeline, and the degree of uncertainty is represented by bold lines.

Time Reasoning. In the collected circulation data, the temporal uncertainty lies in two aspects: the order of events and the time of each event. We adopt the order speculated by domain experts and the circulation of a book can be represented as an event sequence $< e_0, e_2, \cdots, e_{N-1} >$. The time of the event might be a certain year, a time period (e.g., the Yuan Dynasty), or before/after a certain year (e.g. $>$ 1891). We represent the temporal attribute of event e_i as a tuple $(L_i, H_i, l_i, h_i, t_i)$, where L_i and H_i denotes the given lower bound and upper bound, l_i and h_i denotes the normalized lower bound and upper bound, and t_i denotes the estimated time. The values of L and H implicitly express four types of uncertainty in the raw data: (1) *certain,* if a certain year is given, in which case $L = H$; (2) *bounded,* if an interval is given, in which case $L < H$; (3) *semi-bounded,* if the given time is before or after a certain year, in which case one of L and H is infinity; (4) *unbounded,* if the temporal information can not be inferred, in which case both L and H is infinity. The speculated time of each event is obtained separately based on information such as the collection institution and is mostly *semi-bounded* or *bounded.* In fact, taking the constraints of the event order into consideration, we can narrow down the uncertain range, which is represented by l and h.

- *Sequential constraints.* The basic constraint for the normalized bounds is that the preceding events must be earlier than the subsequent events. Thus, l and h should satisfy $l_i \leq l_j, h_i \leq h_j, \forall i < j$. Also, l and h are bounded within the temporal scope, i.e, from 960 (the beginning of the Song Dynasty) to 1960 (the latest time that the Syoryobu Department introduced Song-Yuan editions).
- *Assumption on the uncertainty levels.* We assume that the four types of uncertainty can be arranged in ascending order of uncertainty degree as follows: certain, semi-bounded, bounded, and unbounded. Though, intuitively, the *semi-bounded* uncertainty is more significant than the *bounded* uncertainty, domain experts suggested that the *semi-bounded* time can be set as the known bound since the real time is more likely to be close to the bound.
- *Iterative estimation.* We calculate the time in the order of the uncertainty level. The time of the *certain* type and the *semi-bounded* type is set by the known bound. If event e_i has *certain* time, $t_i = L_i$; If it is *semi-bounded*,
$$t_i = \begin{cases} l_i + 1, & |l_i| < \infty \\ h_i - 1, & |h_i| < \infty \end{cases} .$$
Then, for *bounded* event e_j, the compact bounds (l_j, h_j) is updated and the estimated time is the midpoint of the interval: $t_j = (l_j + h_j)/2$. Lastly, we update the bounds of the *unbounded* event e_k and distribute the unknown time points evenly in the interval.
$$t_k = l_{k-m} + \frac{m}{m+n}\left(h_{k+n} - l_{k-m}\right),$$
where $k-m$ is the order of the earliest event among the consecutive unbounded events to which e_k belongs, and $k+n$ is the last one.
- *Visual encoding.* Bold line segments are used to encode uncertain times. When L_i or H_i is infinity, the line is represented with dots, otherwise a solid line is used.

Interaction. Selecting a singular book of interest, the system exhibits a comprehensive exposition of its associated details, including its version and circulation events. The system also customizes the user's experience by dynamically sorting books based on their interests. By default, the books were sorted in temporal sequence according to the earliest introduction date because of experts' emphasize on such key events. When a particular collector or institution is selected, books are sorted chronologically based on the time they first passed through it.

5.3 Geomap View

The Geomap view provides both an overview of multiple trajectories and a detailed single trajectory view. A Mercator-projected map of East Asia served as the base map. The movement and stop events are represented as circles and lines on the base map.

Multiple book movement. In the overview, trajectories are aggregated to reduce clutter. Due to the missing and uncertainty in the episodic data, we aggregate the location of the atomic events instead of using compound features like the Origin-Destination (OD) movements or sub-paths. Spatial uncertainty makes trajectory clustering challenging. Firstly, clustering locations with different scales of uncertainty might cause misleading. In our data, all locations in book trajectories are described with administrative districts, where higher level administrative districts are of wider range and mean huger uncertainty. An ordinary method is to represent each location as a point on the capital or centroid of corresponding administrative districts. However, with large numbers of uncertain locations, this approach would mislead the users to believe that trajectories are concentrated in the centroid.

To avoid misleading and perverse more accurate uncertainty in the clustering, we adopted a layered clustering process (Figure 6a). We discriminate uncertainty in each location according to the administrative district's division and cluster locations of different levels separately. From the finer level to the coarse level, we start from a `threshold` level according to the screen sizes (*city* by default), where we think the uncertainty under such a level is precise enough under the map scale. We use the FairPair algorithm [18] to take hierarchical clustering on each level. The clustered locations will be rendered with a blur filter at different extents to indicate the scale of uncertainty.

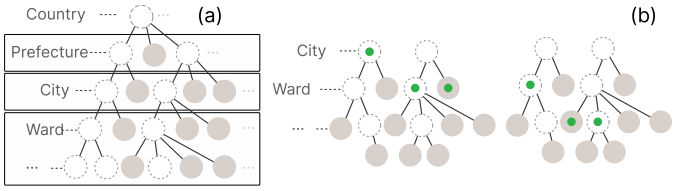

Fig. 6. (a) location clustering according to uncertainty degree (b) selecting representative labels.

The second challenge is the labeling of clustered locations. Due to space limitations, we select several representative place names. Direct methods contain sampling the place names with the highest frequencies or just adopting the place name of the lowest common ancestor on the location tree. The result would suffer from co-exists of place names with containment relationship, while the latter might give over a vague description of the cluster. We propose a dynamic programming-based algorithm (Figure 6b) to find a balance between precision and coverage while avoiding the coexistence of place names with containment relationships. For all locations clustered in a group, the place names of their representative administrative districts form a forest. The problem is formulated as a tree knapsack problem [13], which can be solved in $O(nL)$ time complexity. The multiple trajectory overview is rendered as a geo-network [44] after the location is cluttered. The size of nodes and the width of edges encode the number of stops and movements. We adopt the force-directed edge bundling [27] to decrease visual clutter. The labels are placed with a force-directed label placement algorithm.

Single book movement. In the single trajectory view, users inspect the geographical information on the map as well as data provenance on the side panel. To allow users to view trajectory segments in different scales at one glance, the system automatically recommends magnifying lenses on the map to display fine local movements that are too small to see on a uniformly scaled map. To implement lens recommendations, we detect all point pairs that are too close to draw separately on the map. We use a disjoint set to group these points into clusters, where points in the same cluster are expected to be accommodated by a lens. For each cluster, we compute the magnification of its lens to fit the extent of its points. The position of the lens is determined by searching for the nearest available place on the map that is not occluded by countries related to our data.

5.4 Auxiliary Views

Besides the main view, we provide a book list view (Figure 1E), a legend view (Figure 1F), and a time axis (Figure 1D) to show the distribution of books and institution types, which serve as filters as well. The book list view encodes each book as a rectangular glyph and arranges them according to the four categories. Two vertical lines on the glyph encode the book's printing information (dashed line) and the book's first transmission to Japan (solid line). The position of the lines represents time, and the color represents the type of organization. Different versions of the same book are placed in adjacent positions, which is indicated by a light-colored background. The legend is designed as scented widgets, using a bar chart to display the number of circulation events at different types of libraries and agents. The time axis provides an interactive interface for filtering the time range and selecting specific Chinese or Japanese historical periods. Additionally, the time axis displays significant events in book circulation, including "publication", "last appearance in China", and "first introduction to Japan", encoded using three circles and links between them to represent each book.

5.5 Exploring LiberRoad

The LiberRoad system provides three types of actions, namely selection, filtering, and transforming, allowing users to change their focus and perspectives flexibly. **Selection** targets entities in the view, including six types: books, editions, collectors, locations, institutions, and paths. Clicking multiple library nodes results in the selection of a path, and the trajectories of the books which have gone through the path will be highlighted. **Filtering** targets a set of related books. There are five types of filters: book type, collector type, institution type, time range, and special events. The special events include a book's printing, the last known location in China (which probably indicates where the Japanese obtained the book, reflecting their search strategies), and the first arrival in Japan (which indicates the main cohort that brought Chinese classics to Japan). **Transforming** allows users to change their perspectives and results in a smooth transition from the current view to the desired view. Among the three layouts, the Event Timeline has the lowest level of aggregation, followed by the Location Graph (aggregated by institutions), and the Geomap (by geographical location). When switching from a higher-level aggregated view to a lower-level one, the elements split and move to their corresponding positions.

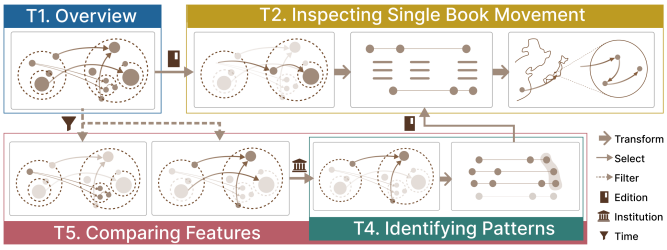

Fig. 7. The pipeline of LiberRoad. T3 and T6 are fulfilled through visual encoding and interaction.

Figure 7 shows the pipeline of exploring LiberRoad. Typical exploration paths are illustrated on the pipeline. The Location Graph provides an overview (**T1**). Users can select an edition to examine the movement of a single book (**T2**). When observing movement patterns (e.g., convergence), users can select the institution and switch to the Event Timeline for validation (**T4**). By filtering the time range, users can compare the features of the trajectories in different periods (**T5**). At each step, they can select the entities (e.g., institutions and editions) for detailed examination (**T3**), represented by the solid arrows in Figure 7.

6 Case Study

We invited the domain collaborators to explore the LiberRoad system in order to address the questions that they were most concerned with. This section introduces the exploration process and findings through two typical cases. The first case is to analyze the characteristics of Japan's acquisition of Chinese classics in different periods. The second case is to analyze the role of different types of collection subjects (institutions and collectors) in the dissemination of Chinese classics.

6.1 Case 1: Japan's Acquisition in Different Periods

The primary concern of the domain experts lies in the characteristics of transnational movement between China and Japan across different historical periods, which serve as reflections of the variations in the political, academic, and cultural contexts throughout history. By filtering for the movement that books were first introduced into Japan, experts obtained an overview of the transnational circulation (**T1**), as shown in Figure 1A. The classics were sourced from regions such as Fujian and Zhejiang in China and subsequently transported to various locations in Japan, including Tokyo, Kyoto, Kanagawa, and Oita (Figure 1C). To examine the circulation of each period, experts brushed on the time axis (**T3**). The earliest collections of Chinese classics in our dataset began during the Kamakura period (1185-1333), as shown in Figure 1a. During this period, books were primarily accumulated in Kanazawa Bunko. By selecting Kanazawa Bunko (Figure 1-a1) and switching to the timeline view (Figure 1-a2), detailed information regarding the chronological order of book acquisitions was examined. The books were sorted based on their arrival at this institution, with the majority arriving after the year 1224. The collection consisted of various types of books, including classics, histories, masters, and collections, with a significant portion directly imported from China. Experts observed that numerous books entered Japan during the Muromachi and Edo periods. In the Muromachi period, most of the introduced books were dispersed among temples (Figure 1b). However, as shown in Figure 1c, during the Edo period, lots of books converged into institutions such as Saiki Han, Momijiyama Bunko, and Syoheizaka Gakumonjyo (**T4**).

The Meiji period witnessed a notable decline in transnational circulation, with few books reaching the Imperial Household Agency (Figure 1-d1). This decrease could be attributed to the active promotion of Westernization policies by the Japanese government following the Meiji Restoration, wherein learning from the West and promoting the modernization of Japanese society became the mainstream of academia and thought, consequently reducing the promotion of Chinese culture. Interested in the domestic circulation in Japan during the Meiji period, experts removed the filter on transnational movement and found that classics continued to circulate in a well-organized manner in Tokyo (Figure 1-d2). A significant number of books collected by major institutions during the Edo period flowed into Naikoku Bunko, and were subsequently transferred to the Imperial Household Agency. Indeed, despite the turbulent political climate during the Meiji period, Japanese Sinology research managed to uphold and inherit the academic traditions established during the Edo period. This continuity paved the way for the development of highly influential modern Sinology research in Japan. Through their exploration, experts have acquired a comprehensive understanding of the circulation patterns and characteristics of each historical period. These findings can be effectively corroborated with the political and cultural contexts that prevailed during each respective period. Through their exploration, experts gained a clear understanding of the circulation patterns and characteristics of each historical period (**T6**). These findings can be effectively corroborated with the political and cultural contexts of each respective period.

6.2 Case 2: Role of Different Types of Agents

Different types of institutions and collectors have played distinct roles in the circulation of Chinese classics in different periods. Experts wanted to explore the time periods and locations in which specific types of collectors were active. The overview of the Location Graph provides a good starting point for the exploration (**T1**), showing the obvious bias distribution of temples in Kyoto (Figure 8-1). Experts filtered the institution type on "temple" and switch the view to take a closer inspection. In the Event Timeline view (Figure 8-2), all books related to temple bubbling up shows the active time of temple institution covered from the late Kamakura period to the Muromachi period. The spatial and temporal convergence of the temple reveals the historical factor, where five mountains in Kyoto led the importing of Chinese classics into Japan.

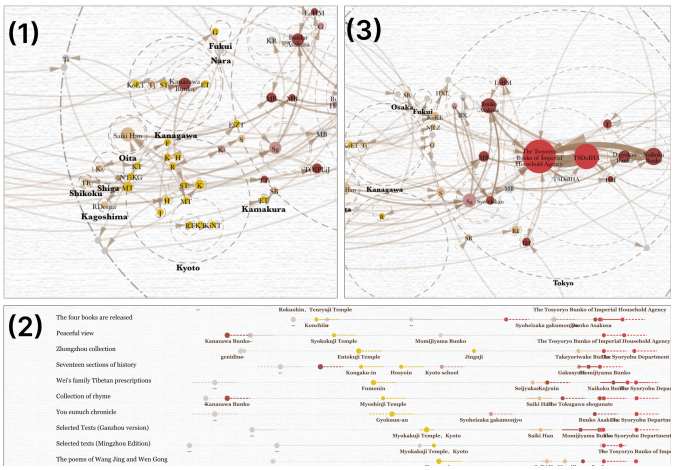

Fig. 8. Analysis of the roles of collection institutions and collectors of different types. (1) Monks of the five mountain literature played a great role in promoting the spread of Chinese classics. (2) They were active from the late Kamakura period to the Muromachi period. (3) Since the Edo period, scholars in the government agency or shogunate agencies become dominators in the introduction and transmission of books.

Another group of collectors that experts were interested in were the scholars. By selecting scholar collectors and filtering on time, experts find scholars became active in circulation from the Edo period (**T3**). A strong relationship between scholar collectors and private and college institutions was detected. From Location Graph (Figure 8-3) showing scholar collector-related circulation, experts found these private institutions are smaller in scale and with less certain information. Experts also observed the imbalanced distribution of related book types, dominated by classic books. This fact confirmed the prevailing of the classic book under the advocate of Neo-Confucian by the Tokugawa shogunate. The Location Graph also shows the important role of *Changping Han School of Learning*, which led to the spread of Confucian Culture, followed by many other colleges, private libraries, and schools (**T4**).

7 Evaluation

The evaluation of LiberRoad system consists of two aspects. We interviewed the domain collaborators to evaluate the effectiveness of LiberRoad in helping completing domain tasks and accelerating their workflow. Moreover, to understand the usability of the system, a user experiment was conducted, providing insights for the future promotion of LiberRoad as a generic tool for humanists.

7.1 Expert Interview

Interviews with domain collaborators were conducted after the implementation of the system prototype. We asked the domain experts to compare their previous workflow that uses Excel tables to inspect data with visualization-assisted analysis. Experts were assigned the task of analyzing the characteristics of circulation from China to Japan through history, which they had proposed in the initial discussions. They were also encouraged to come up with interesting analysis tasks. Overall, they approved that the LiberRoad system provides a fascinating visualization and comprehensive analytical perspective. Apart from the formal interview, our collaborators have been inviting colleagues to explore the system and we continue to hold regular meetings to exchange their feedback. We will introduce some unique insights from expert users, which provoke thinking on future visual design choices.

Embracing novel visual designs. The effectiveness of the Location Graph and Event Timeline in presenting details and data features exceeded experts' expectations. When designing LiberRoad, P1 and P2 preferred the maps that they were familiar with and found the other views a bit abstract to understand. However, presented with the implemented system, experts were surprised at the clarity of the patterns shown in the Location Graph and the detailed circulation in the Event Timeline. While the contradiction between novel visual designs and the conservatism of humanists has been discussed in prior literature [61], we encourage visualization researchers to push beyond the boundaries of conventional visual designs and propose novel visualizations for humanities research.

Supporting analyzing uncertain data. Though one of the experts (P2) is familiar with programming, the significant uncertainty in the data has made it difficult for him to perform statistical analysis or utilize existing map tools to visualize the circulation. As a result, P2 extremely appreciated the design of Location Graph, where uncertain locations were placed on the circle border. *"On the map, points must be given geographical coordinates, but the abstract graph does not present precise locations, thus reducing much misinformation."*

Promoting new research paradigms. LiberRoad received unanimous acclaim among the colleagues of our domain collaborators, considered as a prototype for the new research paradigm in the field of classic circulation studies. Traditionally, scholars need to extensively review the literature and historical materials to determine what is worth studying, which takes great effort and limits them to focus on several classics, lacking a comprehensive understanding of other books. With visualization, however, experts are equipped with a top-down perspective and are able to *"see important nodes including pivots and anomalies in the diagram at once"* (P1).

7.2 User Experiment

We conducted a user study to evaluate whether the system is easy for domain scholars to use, as well as how users interpret the three views during their analysis. Additionally, we were interested in whether there exist differences between humanists and technical background users in the exploration of visual analytics systems, which could provide insights for future visualization design targeting humanities scholars. Therefore, we invited users from diverse backgrounds to take part in the study.

Participants. We recruited 17 participants (5 males and 12 females) for the experiment. 10 participants were humanists (e.g., literature and history), and 7 were from the technical background (e.g., computer science). Participants were asked to specify their experience in charting software and programming languages on a scale of 1 to 5, where 1 denoted no prior knowledge and 5 indicated a high level of familiarity. Most participants demonstrated moderate familiarity with charting software (average rating of 3.47) but lacked familiarity with data analysis or visualization using programming languages (average rating of 1.94).

Procedure. First, we asked participants to fill in their basic information and gave them a short tutorial about LiberRoad. Participants were required to complete two mocking tasks to warm up and verify their understanding of the system. During the study, participants were asked to answer eleven questions we designed according to the tasks identified in this work. The questions vary from low-level inspection to high-level discovery. In order to figure out the usefulness of each view in fulfilling tasks from different aspects, we asked participants to evaluate the helpfulness of each view on completing each task. To compare users' exploration with the design expectation, the representative exploration paths were selected in advance, as shown in Figure 9b. For example, the path for Q8 means starting from the Location Graph, filtering on different time periods, then selecting institutions and switching to the timeline. Then, participants conducted free exploration to observe and analyze various book circulation events. The study concluded with a subjective questionnaire to rate the functionality of LiberRoad and a brief interview with each participant.

Results. Overall, participants got a well understanding of the book circulation through LiberRoad (average accuracy of 83.57%). Figure 9 shows the user's ratings on the three views regarding each task, with the representative exploration paths we designed illustrated in Figure 9b. In general, the views on the exploration paths received high ratings as we anticipated. The timeline view was thought helpful for tasks that require a detailed examination of circulation events (e.g., Q3 and T4), and the Location Graph was considered useful for discovering spatial features (e.g., Q1 and Q11).

- *Difference in perceiving Location Graph.* The most interesting finding is the significant difference in ratings of the Location Graph between humanists and technical background users for Q2, Q7, Q8, and

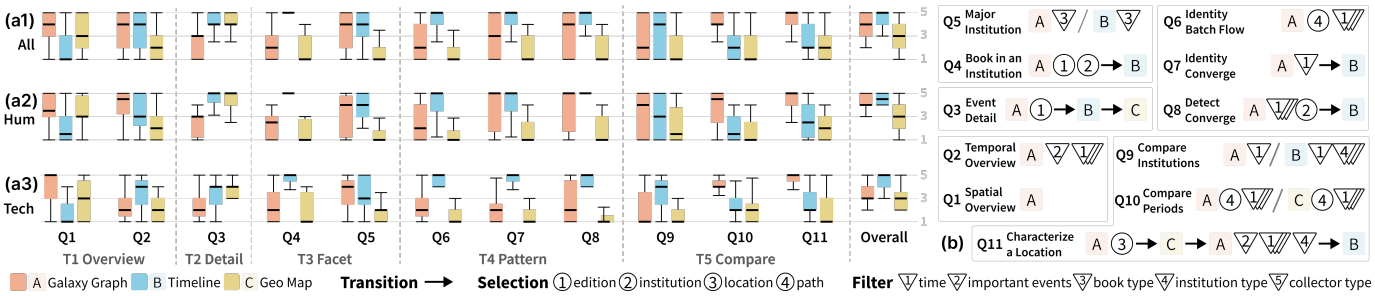

Fig. 9. Users' evaluation of the helpfulness of the three views in different tasks from different perspectives. (a1) all users; (a2) humanity background users; (a3) users with technique backgrounds. (b) shows the exploration paths we proposed when designing the LiberRoad system and the tasks. Overall, the role of each view in multi-faceted tasks corresponds to the expectation.

Q9, which involve feature identification and comparison. Humanists found the Location Graph useful, while technical background users consider it nearly useless. During the interviews, humanist users reflected that the Location Graph offers an intuitive way to grasp circulation patterns. After observing patterns, they would switch to the timeline to examine the details and confirm their findings. In contrast, technical users considered that they could accurately perceive time and quantity through interaction on the timeline without the need to switch between the two views. This finding indicates that humanists might favor intuitive visualization that can directly answer their questions, even if it is novel and unfamiliar to them.

- *Feedback on clutter in Location Graph.* Some participants reported that the Location Graph suffered from edge clutter, which interferes with their viewing of nodes and text labels. Traditional edge bundling methods such as hierarchical edge bundling [26] and force-directed edge bundling [27] do not perform well on our graph where nodes have areas and edges are required not to pass through the nodes. In the current system, the clutter could be reduced by interaction. We are considering the optimization of node layout and edge bundling algorithms in future work.
- *Guidance on exploring the complex system.* Though ratings on the Location Graph for Q7 and Q8 were high among humanists, Q6, a similar question that involves identifying main paths, did not receive a consistent rating. The humanist participants recalled that they were not familiar with path selection when solving this task, and found the Event Timeline easier to use. Despite the tutorial, participants found it difficult to know which elements were interactive when they started to explore the system. To ease the cognitive load of humanists in exploring complex visual analytics systems, embedded manuals and intelligent agents could be investigated in future research.

8 Discussion

This section discusses the comparison between the Location Graph visualization and the map visualization. We also discuss the scalability of the Location Graph, and the role of visualization in digital humanities.

8.1 Scalability of Location Graph

The Location Graph is a generalizable method for representing the book circulation data, as well as other trajectory data with uncertain and hierarchical locations. In this work, the Location Graph is applied to trajectories of 64 books with 389 movements (corresponding to 389 edges). These circulation events involve 107 different administrative regions and 209 institutions, resulting in 316 nodes in the hierarchical structure of the Location Graph. The Location Graph method can be extended to include larger-scale book circulation data encompassing collections beyond those selected in the Imperial Household Agency. The challenges on larger datasets lie in limited space for visualizing all the leaf nodes and significant edge clutter. In order to facilitate the exploration and analysis of numerous locations and institutions, a promising research direction is the Level of Detail (LOD) exploration of the hierarchically packed locations, where the nodes expand dynamically according to users' degree of interest. To reduce edge clutter, optimized node layout and edge bundling can be investigated.

8.2 Location Graph for Geographical Data

While existing tools for geographical data are usually built on maps that present precise locations, an abstract form can be an effective alternative. Especially, in humanities studies, the location data tends to be significantly uncertain and scholars would like to analyze features of fine-grained locations distributed in a vast space. Humanists' emphases on region concepts rather than accurate positions and distances allow us to express locations abstractly. The Location Graph provides hierarchical conceptual representations of locations and presents different levels of uncertainty, thus being able to reveal institutional circulation patterns. In addition to book circulation, we are extending the form of the circular map to present other data with geographical attributes in the field of Chinese classics, and have seen the vast potential of visualizing location data with packed circles to be applied to various scenarios.

8.3 The Role of Visualization in Digital Humanities

In the initial interviews, some experts mentioned that they saw the role of visualization more in terms of presentation and education, and true analysis relied on the thinking of humanists. However, when using LiberRoad, they were surprised by its analytical capabilities. In recent years, visualization has gained intense attention in digital humanities research, involving researchers from various domains, i.e., humanities, visualization, and art [36]. While an increasing number of humanities scholars have been utilizing visualization tools for data analysis, visualization methods targeting domain-specific tasks do not occupy a large proportion in the literature associated with digital humanities. The collaboration on book circulation has made us realize that there might exist numerous valuable research problems in the humanities domain that require human-machine collaboration. The unique characteristics of humanities data, such as uncertainty, necessitate the development of new visualization methods. Therefore, we encourage visualization scholars to actively engage in communication with humanists, with a keen sense of research opportunities and novel visual designs. We hope such deep interdisciplinary collaboration will drive the paradigm shift in humanities research, as well as stimulate intriguing methods in the visualization community.

9 Conclusion

In this work, we propose LiberRoad, an effective visual analytics system supporting experts to better explore Chinese classic circulation data. With tight collaboration with domain experts, we identified the requirements and tasks. The LiberRoad includes three views under a multi-facet analysis framework. Location Graph, a novel visualization method, was proposed to visualize the unique circulation data with significant uncertainty. Two cases as well as the expert interview and user study proved the effectiveness of the design of Location Graph and LiberRoad. We position this work as the first attempt to introduce visual analytics into the Chinese classics circulation study, and to propose a prototype for the visualization research of historical book movement.

Acknowledgments

The authors thank the participants for attending the interviews and the user study, as well as the anonymous reviewers for their valuable comments. This work was supported by NSFC No. 62272012.

REFERENCES

[1] G. Andrienko, N. Andrienko, S. Rinzivillo, M. Nanni, D. Pedreschi, and F. Giannotti. Interactive visual clustering of large collections of trajectories. In *Proceedings of IEEE Symposium on Visual Analytics Science and Technology*, pp. 3–10. IEEE, 2009.

[2] G. L. Andrienko, N. V. Andrienko, J. Dykes, S. I. Fabrikant, and M. Wachowicz. Geovisualization of dynamics, movement and change: Key issues and developing approaches in visualization research. *Information Visualization.*, 7(3-4):173–180, 2008.

[3] G. L. Andrienko, N. V. Andrienko, C. Hurter, S. Rinzivillo, and S. Wrobel. From movement tracks through events to places: Extracting and characterizing significant places from mobility data. In *Proceedings of IEEE Conference on Visual Analytics Science and Technology*, pp. 161–170. IEEE Computer Society, 2011.

[4] G. L. Andrienko, N. V. Andrienko, C. Hurter, S. Rinzivillo, and S. Wrobel. Scalable analysis of movement data for extracting and exploring significant places. *IEEE Transactions on Visualization and Computer Graphics*, 19(7):1078–1094, 2013.

[5] N. Andrienko and G. Andrienko. Visual analytics of movement: An overview of methods, tools and procedures. *Information Visualization*, 12(1):3–24, 2013.

[6] N. V. Andrienko, G. L. Andrienko, H. Stange, T. Liebig, and D. Hecker. Visual analytics for understanding spatial situations from episodic movement data. *Künstliche Intell.*, 26(3):241–251, 2012.

[7] B. Bach, N. H. Riche, R. Fernandez, E. Giannisakis, B. Lee, and J.-D. Fekete. Networkcube: bringing dynamic network visualizations to domain scientists. In *Poster Proceedings of the Conference on Information Visualization*, 2015.

[8] M. Brehmer, B. Lee, B. Bach, N. H. Riche, and T. Munzner. Timelines revisited: A design space and considerations for expressive storytelling. *IEEE Transactions on Visualization and Computer Graphics*, 23(9):2151–2164, 2016.

[9] K. Brodlie, R. Allendes Osorio, and A. Lopes. A Review of Uncertainty in Data Visualization. In *Expanding the Frontiers of Visual Analytics and Visualization*, pp. 81–109. Springer, London, 2012.

[10] M. Burch, F. Beck, and S. Diehl. Timeline trees: Visualizing sequences of transactions in information hierarchies. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, pp. 75–82, 2008.

[11] C. CERL. Material evidence in incunabula, https://data.cerl.org/mei/, 2015.

[12] S. Chen, X. Yuan, Z. Wang, C. Guo, J. Liang, Z. Wang, X. Zhang, and J. Zhang. Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):270–279, 2015.

[13] G. Cho and D. X. Shaw. A depth-first dynamic programming algorithm for the tree knapsack problem. *INFORMS J. Comput.*, 9(4):431–438, 1997.

[14] T. Crnovrsanin, C. Muelder, C. Correa, and K.-L. Ma. Proximity-based visualization of movement trace data. In *Proceedings of IEEE Symposium on Visual Analytics Science and Technology*, pp. 11–18. IEEE, 2009.

[15] S. Deitrick and R. Edsall. The influence of uncertainty visualization on decision making: An empirical evaluation. In *Proceedings of International Symposium on Spatial Data Handling*, pp. 719–738. Springer, 2006.

[16] S. Di Bartolomeo, Y. Zhang, F. Sheng, and C. Dunne. Sequence braiding: Visual overviews of temporal event sequences and attributes. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1353–1363, 2020.

[17] S. Dodge, R. Weibel, and A.-K. Lautenschütz. Towards a taxonomy of movement patterns. *Information Visualization*, 7(3-4):240–252, 2008.

[18] D. Eppstein. Fast hierarchical clustering and other applications of dynamic closest pairs. In *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 619–628. ACM/SIAM, 1998.

[19] M. T. Fischer, A. Frings, D. A. Keim, and D. Seebacher. Towards a survey on static and dynamic hypergraph visualizations. In *Proceedings of IEEE Visualization Conference*, pp. 81–85. IEEE, 2021.

[20] J. Görtler, C. Schulz, D. Weiskopf, and O. Deussen. Bubble treemaps for uncertainty visualization. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):719–728, 2017.

[21] D. Gotz, J. Zhang, W. Wang, J. Shrestha, and D. Borland. Visual analysis of high-dimensional event sequence data via dynamic hierarchical aggregation. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):440–450, 2020.

[22] T. Gschwandtnei, M. Bögl, P. Federico, and S. Miksch. Visual encodings of temporal uncertainty: A comparative user study. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):539–548, 2015.

[23] D. Guo and X. Zhu. Origin-destination flow data smoothing and mapping. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2043–2052, 2014.

[24] H. Guo, Z. Wang, B. Yu, H. Zhao, and X. Yuan. Tripvista: Triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection. In *Proceedings of IEEE Pacific Visualization Symposium*, pp. 163–170. IEEE, 2011.

[25] S. Guo, K. Xu, R. Zhao, D. Gotz, H. Zha, and N. Cao. Eventthread: Visual summarization and stage analysis of event sequence data. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):56–65, 2018.

[26] D. Holten. Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):741–748, 2006.

[27] D. Holten and J. J. Van Wijk. Force-directed edge bundling for graph visualization. In *Proceedings of Computer Graphics Forum*, vol. 28, pp. 983–990. Wiley Online Library, 2009.

[28] T. Ilägcrstrand. What about people in regional science. *Regional Science Association*, 24, 1970.

[29] M.-J. Kraak. The space-time cube revisited from a geovisualization perspective. In *Proceedings of International Cartographic Conference*, pp. 1988–1996. Citeseer, 2003.

[30] F. Kräutli and S. B. Davis. Known unknowns: Representing uncertainty in historical time. *Electronic Visualisation and the Arts*, pp. 61–68, 2013.

[31] M. Krstajic, E. Bertini, and D. Keim. Cloudlines: Compact display of event episodes in multiple time-series. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2432–2439, 2011.

[32] S. Liu, Y. Wu, E. Wei, M. Liu, and Y. Liu. Storyflow: Tracking the evolution of stories. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2436–2445, 2013.

[33] Y. Liu. *From Manuscripts to Block-prints: Research on Chinese and Japanese Edition of "Analects"*. Peking University Press, 2013.

[34] Y. Liu. A study on the Baojuan collected in the bodleian library at Oxford. *The Documentation*, 170(6):77–87, 2018.

[35] J. Lu and Y. Wang. *The Circulation and Influence of Chinese Classics in Japan*. Hangzhou University Press, 1990.

[36] Y. Luo, H. Wu, Y. Guo, S. Tan, C. Liu, R. Jiang, and X. Yuan. Visualization in digital humanities. *Big Data Research*, 8(6):74–93, 2022.

[37] A. M. MacEachren, R. E. Roth, J. O'Brien, B. Li, D. Swingley, and M. Gahegan. Visual Semiotics & Uncertainty Visualization: An Empirical Study. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2496–2505, 2012.

[38] J. Magallanes, T. Stone, P. D. Morris, S. Mason, S. Wood, and M.-C. Villa-Uriol. Sequen-c: A multilevel overview of temporal event sequences. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):901–911, 2021.

[39] R. Munroe. Movie narrative charts, http://xkcd.com/657/, March 2013.

[40] G. Panagiotidou, H. Lamqaddam, J. Poblome, K. Brosens, K. Verbert, and A. Vande Moere. Communicating uncertainty in digital humanities visualization research. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):635–645, 2023.

[41] A. T. Pang, C. M. Wittenbrink, S. K. Lodha, et al. Approaches to uncertainty visualization. *The Visual Computer*, 13(8):370–390, 1997.

[42] P. Riehmann, M. Hanfler, and B. Froehlich. Interactive sankey diagrams. In *Proceedings of IEEE Symposium on Information Visualization*, pp. 233–240. IEEE, 2005.

[43] D. Rosenberg and A. Grafton. *Cartographies of Time: A History of the Timeline*. Princeton Architectural Press, 2013.

[44] S. Schöttler, Y. Yang, H. Pfister, and B. Bach. Visualizing and interacting with geospatial networks: A survey and design space. In *Proceedings of Computer Graphics Forum*, vol. 40, pp. 5–33. Wiley Online Library, 2021.

[45] X. Sun. The circulation of han documents in ancient east asia and the formation of the passage of han books. *Journal of Social Science Front*, (11):113–121, 2017.

[46] Y. Tanahashi and K. Ma. Design considerations for optimizing storyline visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2679–2688, 2012.

[47] A. Thudt, D. Baur, and S. Carpendale. Visits: A spatiotemporal visualization of location histories. In *Proceedings of Eurographics Conference on Visualization*. Eurographics Association, 2013.

[48] C. Tominski, H. Schumann, G. Andrienko, and N. Andrienko. Stacking-based visualization of trajectory attribute data. *IEEE Transactions on*

Visualization and Computer Graphics, 18(12):2565–2574, 2012.

[49] P. Valdivia, P. Buono, C. Plaisant, N. Dufournaud, and J.-D. Fekete. Analyzing dynamic hypergraphs with parallel aggregated ordered hypergraph visualization. *IEEE Transactions on Visualization and Computer Graphics*, 27(1):1–13, 2019.

[50] T. D. Wang, C. Plaisant, A. J. Quinn, R. Stanchak, S. Murphy, and B. Shneiderman. Aligning temporal data by sentinel events: Discovering patterns in electronic health records. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 457–466, 2008.

[51] W. Wang, H. Wang, G. Dai, and H. Wang. Visualization of large hierarchical data by circle packing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 517–520, 2006.

[52] Y. Wang. *Studies on the History of Chinese and Japanese Classic Exchanges*. Hangzhou University Press, 1992.

[53] Y. Wang. *Research on the "Book Road" between China and Japan*. Beijing Library Press, 2003.

[54] Z. Wang, T. Ye, M. Lu, X. Yuan, H. Qu, J. Yuan, and Q. Wu. Visual exploration of sparse traffic trajectory data. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1813–1822, 2014.

[55] Z. Wang and X. Yuan. Urban trajectory timeline visualization. In *Proceedings of International Conference on Big Data and Smart Computing*, pp. 13–18. IEEE, 2014.

[56] C. Ware, R. Arsenault, M. Plumlee, and D. Wiley. Visualizing the underwater behavior of humpback whales. *IEEE Computer Graphics and Applications*, 26(4):14–18, 2006.

[57] D. Weiskopf. Uncertainty visualization: Concepts, methods, and applications in biological data visualization. *Frontiers in Bioinformatics*, 2, 2022.

[58] F. Windhager, S. Salisu, and E. Mayr. Exhibiting Uncertainty: Visualizing Data Quality Indicators for Cultural Collections. *Informatics*, 6(3):29, 2019.

[59] K. Wongsuphasawat and D. Gotz. Exploring flow, factors, and outcomes of temporal event sequences with the outflow visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2659–2668, 2012.

[60] K. Wongsuphasawat, J. A. Guerra Gómez, C. Plaisant, T. D. Wang, M. Taieb-Maimon, and B. Shneiderman. Lifeflow: Visualizing an overview of event sequences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1747–1756, 2011.

[61] Y. Xing, C. Dondi, R. Borgo, and A. Abdul-Rahman. A design study of visualizing historical book movement. In *EuroVis (Short Papers)*. 2022.

[62] H. Yang. *Research on "Records of the Grand Historian" in Japanese*. Zhonghua Book Company, 2017.

[63] H. Yang. Aw ross and his sinology and translation of "The Book of Han". *Peking University Center for Ancient Chinese Documentation Studies Journal*, 21(1):256–263, 2020.

[64] H. Yang. *A Compilation of Engraved Versions of "Records of the Grand Historian" in Japan*. Tianjin People's Publishing House, 2020.

[65] H. Yang. Caleb Cushing and Chinese collection in North America. *Library Tribune*, 42(1):131–136, 2022.

[66] W. Zhang, S. Tan, S. Chen, L. Meng, T. Zhang, R. Zhu, and W. Chen. Visual reasoning for uncertainty in spatio-temporal events of historical figures. *IEEE Transactions on Visualization and Computer Graphics*, 2022.

[67] J. Zhao, P. Forer, and A. S. Harvey. Activities, ringmaps and geovisualization of large human movement fields. *Information Visualization.*, 7(3-4):198–209, 2008.

[68] J. Zhao, Z. Liu, M. Dontcheva, A. Hertzmann, and A. Wilson. Matrixwave: Visual comparison of event sequence data. In *Proceedings of the Annual ACM Conference on Human Factors in Computing Systems*, pp. 259–268, 2015.