HoLens: a visual analytics design for higher-order movement modelling and visualization

Computational Visual Media
https://doi.org/10.1007/s41095-023-0392-y Vol. 10, No. 6, December 2024, 1079–1100

Research Article

HoLens: A visual analytics design for higher-order movement modeling and visualization

Zezheng Feng[1,2], Fang Zhu[2], Hongjun Wang[2], Jianing Hao[3], Shuang-Hua Yang[2,4,5] (✉), Wei Zeng[3], and Huamin Qu[1]

Abstract Higher-order patterns reveal sequential multistep state transitions, which are usually superior to origin-destination analyses that depict only first-order geospatial movement patterns. Conventional methods for higher-order movement modeling first construct a directed acyclic graph (DAG) of movements and then extract higher-order patterns from the DAG. However, DAG-based methods rely heavily on identifying movement keypoints, which are challenging for sparse movements and fail to consider the temporal variants critical for movements in urban environments. To overcome these limitations, we propose *HoLens*, a novel approach for modeling and visualizing higher-order movement patterns in the context of an urban environment. *HoLens* mainly makes twofold contributions: First, we designed an auto-adaptive movement aggregation algorithm that self-organizes movements hierarchically by considering spatial proximity, contextual information, and temporal variability. Second, we developed an interactive visual analytics interface comprising well-established visualization techniques, including the H-Flow for visualizing the higher-order patterns on the map and the higher-order state sequence chart for representing the higher-order state transitions. Two real-world case studies demonstrate that the method can adaptively aggregate data and exhibit the process of exploring higher-order patterns using *HoLens*. We also demonstrate the feasibility, usability, and effectiveness of our approach through expert interviews with three domain experts.

Keywords data visualization; movement modeling; state sequence visualization; movement visualization; urban visual analytics

1 Introduction

Many complex systems use network structures to represent interactions between entities, e.g., to construct a global shipping network to represent ship movements between ports [1] or to build a neuron graph to simulate neuron transmission in the brain [2]. Each entity is represented as a node of the network, and the flows are encoded as edges. Such a network model implicitly assumes that the current status is dependent on only its precedent, i.e., first-order dependency as in a Markov process [3], Fig. 1(B). However, this model cannot represent scenarios in which the current status may not depend entirely on its first-order precedent. For example, when humans browse the Internet, subsequent mouse-clicking behavior may not depend on only one previous behavior [4].

Higher-order dependency analysis, which can

1 Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong, China. E-mail: Z. Feng, zfengak@connect.ust.hk; H. Qu, huamin@cse.ust.hk.

2 Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China. E-mail: F. Zhu, 11711623@mail.sustech.edu.cn; H. Wang, wanghj2020@mail.sustech.edu.cn; S.-H. Yang, shuang-hua.yang@reading.ac.uk (✉).

3 Thrust of Computational Media and Arts (CMA), The Hong Kong University of Science and Technology (Guangzhou), Guangzhou 511458, China. E-mail: J. Hao, jhao768@connect.hkust-gz.edu.cn; W. Zeng, weizeng@hkust-gz.edu.cn.

4 Department of Computer Science, University of Reading, Berkshire RG6 6AH, UK.

5 Shenzhen Key Laboratory of Safety and Security for Next Generation of Industrial Internet, Southern University of Science and Technology, Shenzhen 518055, China.

Manuscript received: 2023-07-18; accepted: 2023-11-13

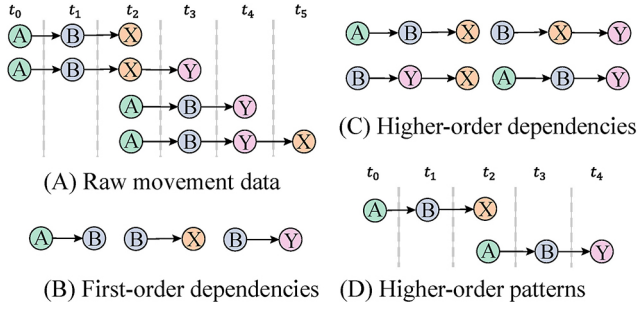


Fig. 1 Illustration of higher-order patterns: (A) Raw movement data describe the change of positions over time, modeled as trajectories. We can extract (B) first-order and (C) higher-order dependencies from these movement data. Moreover, a set of first-/higher-order dependencies can form (D) higher-order patterns.

be traced back to Shannon's high-order memory model [5], can alleviate this problem. This analysis is crucial to many real-world applications e.g., the analysis of animal behavior [6], rumor spread [7]. For example, biologists gathered data from sensors mounted on animals and extracted higher-order dependencies to analyze behavioral patterns and activities [6, 8]. Species invasions were investigated and predicted by tracing multiple higher-order pathways in global ship data [1].

When analyzing geospatial movements, higher-order patterns that reveal sequential multi-step state transitions may depict insights different from origin-destination (OD) patterns. For example, three OD patterns (Fig. 1(B)) are retrieved from the raw movements in Fig. 1(A) can show the relations between two states, whereas higher-order patterns (Fig. 1(C)) can identify more detailed sequential relations. In addition, in urban areas, analyzing the multistep behavior of moving objects is helpful [9]. For example, urban planners can plan public facilities and manage traffic when they understand the multistep moving behavior of citizens, and epidemiologists can trace the possible propagation path and forecast potential routes when making sense of citizens' multistep moving pattern [10]. These studies rely heavily on identifying the functional attributes of each state (i.e., what do citizens do in this area) before forming higher-order patterns. Although clustering-based (e.g., Refs. [11–13]) and aggregation-based techniques (e.g., Refs. [14–16]) have been proposed for keypoint identification, these methods may not be consistent with the contextual information that depicts the underlying mechanisms for movements. Furthermore, existing visual analytics have been proposed to explore higher-order movement patterns (e.g., Refs. [17, 18]). However, these studies seldom considered the influence of temporal attributes on higher-order movement patterns. As an example, Fig. 1(D) illustrates that two higher-order patterns exist at different periods, i.e., A→B→X in period $[t_0 - t_2]$ vs. A→B→Y in period $[t_2 - t_4]$.

This study focused on exploring higher-order movement patterns, specifically in the context of an urban environment. We present a list of design considerations as follows.

- Contextual information. Contextual information is vital for reasoning the underlying mechanisms when analyzing movements because they are related to numerous factors, e.g., activity distributions and household demographics.
- Multi-scale self-organization. The movements shall be aggregated adaptively according to the static contextual information and dynamic movements. Furthermore, the aggregation method shall support consistency when exploring under different scales.
- Temporal variability. Higher-order movement patterns in an urban area are highly regular and temporal variability [19], e.g., people commute to business districts in the morning and back to residential districts in the evening.

We introduce a visual analytics approach, namely, **H**igher-**o**rder **Lens** (*HoLens*). *HoLens* makes twofold contributions in terms of modeling and visualizing higher-order movement patterns in urban areas. First, we designed a self-organization aggregation method that forms regional clusters by considering the spatial proximity of movements, contextual information, and access frequency to each location. We then constructed a directed acyclic graph (DAG) with the centroids of cluster regions as nodes and the interactions between per-pair cluster regions as edges. Consequently, we can extract higher-order movement patterns by considering the spatial and temporal characteristics. Second, we proposed a multi-view visual analytics system that enables the exploration of higher-order movement patterns from spatial and temporal dimensions. The effectiveness of *HoLens* was demonstrated through case studies conducted on real-world movements in New York City and through expert feedback.

The main contributions of this study are as follows:

- A spatial proximity- and contextural-based self-

organizing method for adaptively aggregating urban movement data at different scales, which can avoid the spareness of the urban movement data.
- An interactive visual analytics approach with a newly developed user interface named *HoLens*, consisting of a set of newly designed visualizations to assist analysts in exploring the higher-order patterns in urban areas.
- Two real-world case studies and expert interviews to demonstrate the feasibility, usability, and effectiveness of our proposed approach.

2 Related work

Below, we summarize the related studies on movement modeling and visualization (Section 2.1), state sequence (Section 2.2), and higher-order dependency visualization (Section 2.3).

2.1 Movement modeling and visualization

Movement data describe the change in position over time for moving objects [20]. Understanding movement pattern is important in many domains, e.g., animal ecology [21], social media [22], and urban transportation [23]. Andrienko et al. [24] categorized three approaches for exploratory and analytical visualization of movement data: *direct depiction*, *summaries*, and *pattern extraction*. *Direct depiction* of each movement record (e.g., Ref. [25]) can facilitate the extraction of noteworthy patterns. However, this approach can result in cluttering issues. Appropriate *summaries* can address these issues by using spatial generalization and aggregation methods (e.g., Refs. [26, 27]) or visualization methods (e.g., Refs. [28, 29]).

Nevertheless, owing to their dynamic and heterogeneous properties, *direct depictions* or *summaries* of a large amount of movement data are nontrivial. A more general pipeline is to first conduct *pattern extraction* on movement data, followed by visualization of the movement patterns. Advanced data mining techniques (e.g., Refs. [30, 31]) can be applied to pattern extraction. Lee et al. [13] proposed a partition-and-group framework for clustering trajectories based on ordinary trajectory clustering algorithms to identify common sub-trajectories. These methods focus on localized movement patterns while neglecting higher-order connections among regions. Zeng et al. [32] showed that different movement rhythms were derived when the order of the movement pattern was increased. Alternatively, movements can be modeled as graphs [33], and graph-based methods can be employed to extract higher-order patterns (e.g., Refs. [16, 34, 35]). However, these works usually neglect considering contextual information and fail to support adaptive multi-scale modeling.

In this study, we proposed a novel dynamic and adaptive movement modeling method that considers both spatial proximity and contextual information of urban movement data.

2.2 State sequence visualization

State sequence visualization aims to analyze behavior that generally exhibits some form of symmetry and regularity, e.g., complex computer-based systems [36] and chess playing [37]. The basic approach to state sequence visualization is to place events along a horizontal time axis, as done by Lifelines [38] and CloudLines [39]. However, the large number of nodes and edges hinders analysis and insight discovery. To reduce the load, a common solution is to first aggregate state sequences and then use visualizations such as the Sankey diagram [40]. Other studies (e.g., Refs. [41–45]) consolidated common subsequences and extracted and visualized tree-like representation from state sequence data. These representations may ignore some low-frequency states in the visual summary.

Visualizing movement data as a state sequence is more challenging than conventional state sequence visualization because the position must be considered as a constraint. Using graphs to visualize state transition sequences may render the spatial context invisible. Nevertheless, researchers have developed new visual designs to depict state sequences using spatial information, e.g., interchange circos diagram for urban traffic [46], state transition graphs in observational time-series [17], and designs for animal movements [6, 47, 48]. However, these methods cannot produce a higher-order state and an increasing number of nodes leads to visual complexity.

A novel visualization design called the *higher-order state sequence chart* was developed to avoid visual clutter while helping to analyze higher-order patterns. This design can express both the characteristics of higher-order state sequences and the temporal features of higher-order movement patterns.

2.3 Higher-order dependency visualization

Higher-order dependency, where a sequence of the

preceding status exerts influence on the present state, is prevalent in complex systems, e.g., object movements and web clickstreams [2]. For the higher-order dependency, the historical sequence is considered [49].

Some conventional solutions consolidate states into a graph. The state transition diagram [41] shows all routes between the nodes. Lu et al. [50] visualized the sentiment trend of a time series using stacked area charts. However, these approaches cannot scale well to highly connected graphs. Sankey diagrams [51] have been the choice for many studies [52–54] to show transition pathways. In addition, matrix-based representations [55, 56] have been designed to avoid visual clutter caused by dense edges in Sankey diagrams. However, matrix-based representation introduces usability issues.

For interactive exploration, Chen et al. [22] developed a visual analytics approach to analyze the movement patterns across cities using social media data. Blaas et al. [17] proposed a smooth curved line to visually support exploring the transition between states in an observational time series at a macro scale; however, it is not suitable for visualizing various higher-order dependencies that exist on one physical node together. Rosvall et al. [18] proposed a glyph that directly bridges two consecutive nodes by using the current node. Theoretically, this method can integrate different dependencies through the current node to build the correspondence; however, the area of the current node limits the scalability of various dependencies. To handle such issues, HoNVis [1] visualizes higher-order dependencies and supports exploration from an overview to a fine-grained level, while ignoring the temporal characteristics of higher-order dependency. *HoLens* aims to intuitively represent the higher-order dependencies, especially revealing the temporal features that can further help experts analyze the movement pattern.

3 Requirement and method overview

In this section, we summarize the requirements (Section 3.1) and provide an overview (Section 3.2) of the proposed solution.

3.1 Requirement analysis

In the early stages of this research, we held regular meetings with target users and two domain experts (**E.A** and **E.B**). Target users focused on emergency response for traffic management and urban planning. They aim to find the law of human movement and determine what people tend to do (what happened) in a certain region at a certain time. Their daily work, which involves analyzing higher-order movement patterns, has two goals: (1) segmenting the urban area into functional regions and (2) optimizing traffic. We summarize the requirements for this research in terms of two aspects: higher-order movement modeling (R.1, R.2, R.3) and higher-order pattern visualization (R.4, R.5).

R.1 context-aware aggregation. To avoid sparsity of the movement data in an urban area, aggregating the data into clusters and constructing a DAG based on these clusters is essential for analysis. Furthermore, when aggregating clusters, considering urban contextual information (e.g., the function of the block and the integrity of the buildings) can facilitate the analysis of higher-order patterns, in which more movements with similar features are gathered together.

R.2 multi-scale self-organization. Analysts usually explore higher-order patterns at different levels, such as district and city levels. Thus, the principle of merging the small regions into a larger one and maintaining the consistency of the regions at different hierarchies profoundly influences the result of analyzing the higher-order pattern.

R.3 higher-order pattern extraction with temporal variability. Human activities typically depict temporal variability. Some activities exist only in certain periods and some may implicitly exhibit peak and non-peak features. Analyzing higher-order patterns can help understand traveling behavior and inspire them to make proper decisions, such as in urban planning. Therefore, a mining method for extracting these higher-order patterns with temporal constraints would be helpful.

R.3.1 global higher-order pattern extraction. Factors such as daily commuting and weekend rest often reflect higher-order patterns, e.g., from housing estates to business districts and then back on weekdays. Therefore, global movement trend between different regions must be explored, and higher-order patterns between different regions within the overall scope of the city must be mined.

R.3.2 local higher-order pattern extraction. In contrast to extracting higher-order patterns

globally, local higher-order pattern extraction was performed for a given region. As a result, these patterns are more specific, and the temporal variability is more detailed. However, the amount of movement data in the given region may sometimes decrease dramatically owing to small ranges. Therefore, the flow of higher-order patterns must be considered to avoid overfitting during extraction.

R.4 higher-order pattern visualization. Representing a higher-order pattern for a global region or a given local region is similar to visualization. A good visualization design for higher-order patterns can help summarize the travel mode. In addition, for a node in a DAG, its meaning in a higher-order pattern differs from its physical meaning at the DAG. Typically, one node in a DAG may represent different orders of many higher-order patterns. Therefore, the visualization of the nodes in the DAG and higher-order aspects is essential.

R.5 higher-order pattern comparison. For higher-order patterns at different temporal and geographical dimensions, it is helpful to determine the existence time of different higher-order patterns, their geographical distribution on the map, and the movement composition at each node.

3.2 Method overview

HoLens consists of two phases, including movement modeling (Section 4) and interactive higher-order pattern visual exploration (Section 5). The movement modeling phase supports adaptive, hierarchical movement aggregation, and higher-order pattern extraction. In the interactive visual exploration phase, a novel visual analytics interface is designed for interactive analysis.

In the movement modeling phase, *HoLens* takes movement data and POI categories as inputs (Fig. 2(A)). After cleaning the data, *HoLens* first estimates the stay time of each moving object at each location. Next, *HoLens* aggregates the movements in the "regions" and distinguishes the region's dominant functions. Then, the basic "regions" are self-organized according to their entropy by the proposed aggregation algorithm (Fig. 2(B)). In addition, regions can be organized at different scales. Subsequently, a DAG is constructed with each region as a vertex and the flow between the regions as an edge (Fig. 2(C)). In accordance with the DAG, higher-order patterns were extracted, with consideration to temporal variability (Fig. 2(D)).

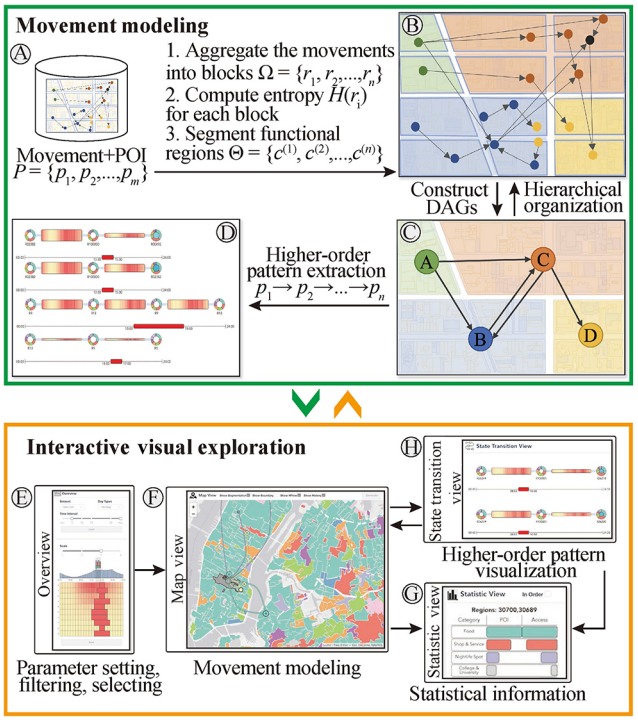


Fig. 2 The overview of *HoLens*. *HoLens* consists of two stages: movement modeling (A, B, C, and D) and interactive visual exploration (E, F, G, and H).

An interactive visual analytics interface (Fig. 3) was developed for the interactive visual exploration phase. It is a web-based multi-view interface that provides interactive visual analytics for higher-order pattern exploration. Data collection and preprocessing were performed offline on a GPU server with four GeForce GTX 2080Ti graphics cards. *HoLens* runs on an Apple M1 MacBook Pro with 16 GB memory. The backend is supported by Flask, and the frontend is implemented by Vue.js and D3.js. In the overview (Fig. 2(E)), users can observe the distribution of global higher-order patterns in the temporal dimension. In the map view (Fig. 2(F)), users can visualize the results of the movement aggregation and the spatial distribution of higher-order patterns. Finally, the state transition view (Fig. 2(H)) visualizes the higher-order patterns in detail and compares different higher-order patterns. Moreover, the statistic view (Fig. 2(G)) displays the statistics from both map and state transition views.

4 Model

In this section, we introduce the construction of

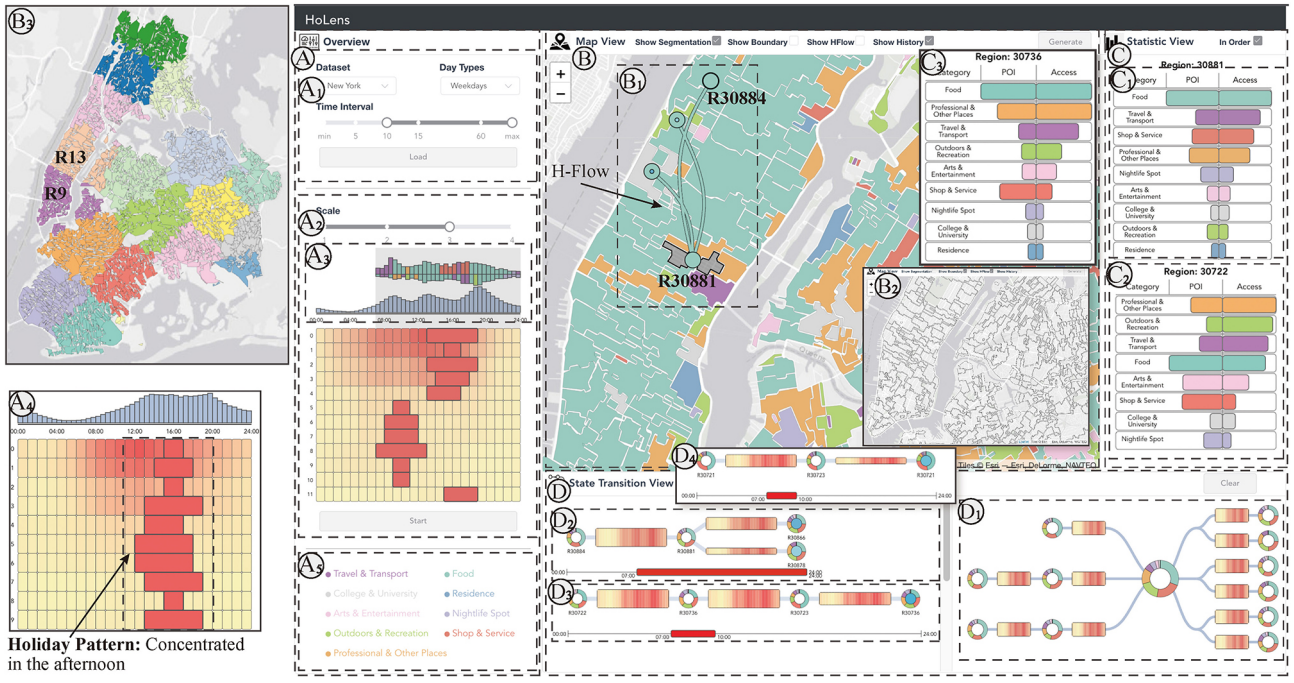


Fig. 3 *HoLens* supports visually exploring the urban higher-order patterns, which consists of four views: (A) the overview provides a configuration panel and a global representation to the end-users, (B) the map view displays the result of movement aggregation and supports visualizing the higher-order patterns in a spatial dimension, (C) the statistic view shows the detailed statistical information of the selected region, and (D) the state transition view provides the function for the further exploration the higher-order patterns.

the data structure (Section 4.1). Subsequently, a novel method for data aggregation (Section 4.2.1) and hierarchical movement self-organization (Section 4.2.2) is introduced. Finally, the method for higher-order pattern extraction in this study (Section 4.3) is introduced.

4.1 Data preparation

Urban movement data usually exhibit three essential attributes: the ID of the moving object, the position of the spatial dimension, and the timestamp of the temporal dimension. In map visualization, the geographical data from the open street map (OSM)[①] were also utilized. Although each movement data record contains a timestamp that can reflect the time point of a moving object existing in a certain position, it cannot provide a period in which each moving object remains in the same position. Thus, in this study, the shortest path-based method was applied to estimate the period of a moving object's stay in a certain position. First, we formed the sequence of each unique moving object's position according to the corresponding timestamp in one day to generate their movement trajectory. Next, the trajectory with significant time intervals is divided into independent sub-trajectories, in which two consecutive points with large intervals have little association, and then the sub-trajectory with only a single point is filtered out. Subsequently, the travel time of two adjacent movement points is computed in accordance with the traffic conditions and the optimal path using a mainstream commercial map application (i.e., the API from Google Maps[②]). Finally, the stay time of the moving object at each position is determined by subtracting the travel time between two adjacent points.

4.2 Aggregation and hierarchical organization

In this section, we delve into the procedure of data aggregation (**R.1**) and the hierarchical self-organization of movement data (**R.2**). Our aim is twofold: (1) to consolidate movement points based on their spatial proximity while considering urban contextual information, and (2) to facilitate the adaptive formation of regions in response to user demand. Subsequently, we define the concept "region" following the work from Ref. [57], which employed connected component labeling (CCL) on a map

① https://www.openstreetmap.org/

② https://www.google.com/maps/

Table 1 Summary of major variables

Notation	Description
d	Number of distinct POI categories
D_i	Density vector for region r_i
g	R tree function mapping POIs to regions
$H(r_i)$	Entropy of region r_i
I	Sequence of vertices comprising the higher-order dependency
$i^k_{\Delta T}$	k-th vertex in the DAG for a specific time period ΔT
i^{k+1}	Vertex in the constructed DAG
O_i	Set of POIs projected into region r_i
p_j	j-th point of interest (POI)
X^{k+1}	Vertex in the DAG at time step $k+1$
α	Aggregation threshold
Ω	Set of spatial regions
$\omega(i^k_{\Delta T} \to i^{k+1})$	Edge weight indicating connection intensity in period ΔT
Θ	Set of POI centroids
ΔT	Selected time period

divided by the road network. Moreover, entropy can be used to identify higher-order patterns, as higher-order patterns are patterns that involve diverse Points of Interest (POIs) visited. For example, a higher-order pattern in human mobility data may be a sequence of visits to different types of POIs, such as a home, a workplace, and a restaurant. By combining entropy, we can easily achieve this. The scale of a region can be modified manually by selecting different types of roads (e.g., freeway, primary, secondary). These regions serve as fundamental building blocks for both data aggregation and point organization.

4.2.1 Data aggregation

Our primary focus in this research is to gain insights into events occurring within POIs located within specific regions. To achieve this, we first map the movement data originally projected onto roads to the nearest region. In the context of data aggregation, entropy can be used to measure the uncertainty of the POIs within a region. A region with high entropy contains a diverse set of POIs, whereas a region with low entropy contains a more homogeneous set of data points. Subsequently, we aggregate the movement data within these corresponding regions and calculate the entropy of each region using the following steps.

Spatial projection: Formally, let r_i denote the ith region obtained by CCL, and $\forall r_i \in \Omega$ is the set of spatial regions. Then, we project each POI p into the spatial region through the R-tree [58] with $g = \text{Rtree}(\Omega)$, where $g : p \to r$ is a function of the R-tree that returns the located region of p. Thus, the projected process for region r_i can be written as

$$\mathrm{O}_i = \left\{ c^{(j)} \mid g(p_j) = r_i \right\}, \quad \forall p_j \in \Theta \tag{1}$$

where Θ denotes the set of POI.

Entropy calculation: Subsequently, we calculate the entropy for region r_i,

$$H(r_i) = -\sum_{a_i \in D_i} P(a_i) \log_e P(a_i) \tag{2}$$

where $D_i \in \mathbb{R}^d$ is the density vector calculated by O_i, where d represents the number of distinct POI categories obtained from external sources, such as the Foursquare API[①](e.g., *Arts & Entertainment, Shop & Service, and Outdoors & Recreation*). For example, $D_{i,j}$ indicates the density probability of the jth POI category in r_i, thereby satisfying $\sum_j D_{i,j} = 1$. Entropy H serves as the degree of confusion within the region because we aim to ignore human mobilities that have many consistent attributions. For example, such higher-order dependencies $Tr : p_1 \to p_2 \to \cdots \to p_n$ containing multiple repetitive POI types are undesirable. This step is described in Algorithm 1.

Algorithm 1 Region entropy computation

Input: spatial region set $\Omega = \{r_1, r_2, ..., r_n\}$, set of GPS points $P = \{p_1, p_2, ..., p_m\}$.
Output: $\Theta_r = \{c^{(1)}, c^{(2)}, ..., c^{(n)}\}$

```
g = Rtree(Ω)    // ST query
for r_i ∈ Ω do
    O_i = {c^(j) | g(p_j) = r_i}, ∀p_j ∈ P,
    H(r_i) = −Σ_{a_i∈D_i} P(a_i) log_e P(a_i)
end for
for r_i ∈ Ω do
    Θ ← Θ+ BFS(r_i, V)
end for
return Θ
```

4.2.2 Hierarchical movement organization

We introduced an innovative algorithm for hierarchical self-organization of movement data. This algorithm facilitates the clear visualization and exploration of higher-order dependencies at different levels while ensuring the consistency of regions across these levels. Essentially, it refines the spatial layout and reduces the spatial entropy, as mentioned in Section 4.2.1. Technologically, we concatenate the regions that present lower entropy after aggregating.

① https://api.foursquare.com/v2/venues/categories/

Given the set $\Omega = \{r_1, r_2, ..., r_n\}$, a breadth-first search (BFS) is employed to search the associated neighbors for each region. The aggregated condition for every pair of regions is defined as

$$\frac{1}{2}(H(r_i) + H(r_j)) \geqslant \alpha H(r_i + r_j) \quad (3)$$

where α denotes the aggregating threshold. Intuitively, the smaller α, the easier the aggregation. Algorithm 2 gives the pseudocode for generating spatial proximity, returning $\Theta = \{c^{(1)}, c^{(2)}, ..., c^{(n)}\}$, the set of renewed spatial regions. $c^{(i)}$ represents the cluster index corresponding to the region r_i.

As in **R.2**, *HoLens* aims to provide a multi-level region scale to systemically understand higher-order movement patterns. The first-order merged procedure was introduced above; thus, for next-level aggregation, only the regions incorporated in the previous iteration are labeled as new regions, and their inner regions are removed accordingly. This strategy is consistent with different level region scales because high-level region styles always come from lower-level regions.

4.3 Higher-order dependency extraction

After aggregating the movements into regions, we construct a DAG with the centroid of each region as the vertex and the in/outflow between regions as the directed edges to extract higher-order patterns (as illustrated in Fig. 2(C) and Fig. 2(D)). In this study, we considered temporal variability in higher-order patterns. Thus, we first modify the transition probability for the first-order Markov model by adding a temporal constraint (**R.3**):

$$P(X^{k+1} = i^{k+1} | X_T^k = i_{\Delta T}^k) = \frac{\omega(i_{\Delta T}^k \to i^{k+1})}{\sum_j \omega(i_{\Delta T}^k \to j)} \quad (4)$$

where $i_{\Delta T}^k$ indicates the kth vertex in the constructed DAG and ΔT denotes the selected period. The edge weight $\omega(i_{\Delta T}^k \to i^{k+1})$ indicates the intensity of the connection in a certain period ΔT, which, in our scenario, can be assigned as the sum of pairwise connections $(i^k \to i^{k+1})$ in period ΔT.

For higher-order pattern extraction, many existing methods [4, 59–61] focus on a fixed order, but we focus on the number of orders that are not fixed. Thus, we refer to the method of Xu et al. [2] and modify the method by adding temporal constraints

$$P(X^{k+1} = i^{k+1} | X_T^k = (i_{\Delta T}^k | I)) = \frac{\omega((i_{\Delta T}^k | I) \to i^{k+1})}{\sum_j \omega((i_{\Delta T}^k | I) \to j)} \quad (5)$$

where I indicates a sequence of vertices comprising the higher-order dependency, e.g., if a higher-order dependency exists at vertex D, that is, $[\mathrm{D}|\mathrm{C}, \mathrm{B}, \mathrm{A}]$. I represents the sequence $\mathrm{C}, \mathrm{B}, \mathrm{A}$ and $i_{\Delta T}^k | I$ indicates that D with a higher-order dependency $[\mathrm{D}|\mathrm{C}, \mathrm{B}, \mathrm{A}]$ exists in ΔT.

We take advantage of the Kullback–Leibler divergence (KLD) [62] to recursively measure whether we should enlarge the current order; in other words, to control the order number of the higher-order pattern. For example, we measure the transition probability of the current n-order dependency and its $(n+1)$-order dependency by KLD; if it significantly changes, it indicates that the dependency should be increased from the current order to $(n+1)$-order.

Algorithm 2 BFS

Input: spatial region set $\Omega = \{r_1, r_2, ..., r_n\}$, entropy H, visited set V, $\Theta_r = \{\}$.
Output: $\Theta_r = \{c^{(i)}, c^{(j)}, ...\}$

```
if r_i ∉ V then
    for r_j ∈ r_i.Neighbors do
        if (1/2)(H(r_i) + H(r_j)) ⩾ αH(r_i + r_j) then
            r_i ← merge(r_i, r_j)
            r_i.Neighbors ← r_j.Neighbors ∪ r_i.Neighbors
            V ← V + r_j
            Θ_r ← Θ_r + c^(j)
        end if
    end for
    for r_j ∈ r_i.Neighbors do
        BFS(r_i, V)
    end for
end if
return Θ_r
```

5 Visualizations

This section introduces the designs of *HoLens*, which helps analysts interactively analyze higher-order patterns.

5.1 Overview

This overview aims to help analysts configure parameters and control the analytical process from a temporal dimension. Analysts can select the dataset, pre-configure the parameters, control the temporal process, and obtain the temporal distribution of higher-order patterns. The overview consists of a configuration panel (Fig. 3(A_1)), global controlled temporal distribution matrix (Fig. 3(A_2)), and legend

board (Fig. 3(A_5)).

The configuration panel (Fig. 3(A_1)) supports dataset selection, analytical day types selection (weekday/weekend), and time interval filter. The global controlled temporal distribution matrix comprises a controlled timeline (Fig. 3(A_3)) and a temporal distribution matrix (Fig. 3(A_4)) (**R.4**). The controlled timeline supports controlling the period and visualizing the flow from a temporal dimension (**R.3**). It (Fig. 3(A_3)) is composed of a time axis, single-directional histogram, bidirectional histogram, and sliding window. The single-directional histogram reflects the tendency of active movements distributed in one day. The sliding window aims to enhance interaction when analyzing detailed higher-order patterns. When analysts select an AOI in the map view (Fig. 3(B_1)), a bi-directional histogram (Fig. 3(B_3)) is generated to visualize the in-/out-flow of the selected region. The color scheme of the bar corresponds to the POI category, indicating the function of the selected region during this period. The height of the bi-directional histogram reflects the volume of the flow in detail.

The temporal distribution matrix (Fig. 3(A_4)) reflects the global higher-order patterns distribution at a temporal dimension. It jointly shares the same time axis with the controlled timeline. Each row of the matrix is a heatmap that reflects the flow of higher-order dependencies. In each row, the colored bar is located at the corresponding position to visualize the city-scale higher-order pattern and is placed in descending order of flow size. When analysts select a higher-order pattern in the temporal distribution matrix, a higher-order state sequence chart (Fig. 7(B)) is generated (Fig. 3(D)).

5.2 Map view

The map view (Fig. 3(B)) was designed to support exploring higher-order patterns in a spatial dimension. Analysts can interact with other views (overview and state transition view) through the map view, and the result in the spatial dimension can be displayed in the map view.

Region segmentation. The map view supports the segmentation of urban areas, providing an intuitive representation of movement aggregation on the map. As mentioned in Section 4.2, "region" is used as the basic unit; thus, the segmented region is composed of these basic "regions". In the map view, all the single unit "regions" are aggregated in the same segmented region, and the in-between roads and boundaries are eliminated for a more intuitive visual effect. The region segmentation function in the map view reveals the dominant POI category for the selected period (selected in the controlled timeline (Fig. 3(A_3)) in the overview) at each segmented region, enabling analysts to understand what people tend to do in the segmented region.

Comparison & multi-region exploration. The map view supports comparison (**R.4**) and multi-region exploration, helping analysts compare movement aggregation at different periods and explore higher-order patterns based on user demand. The comparison function in the map view differs from that in the state transition view (Fig. 3(D)). In the map view, the comparison is used to compare the aggregation results at different periods. In Section 4.2.1, the method enables adaptive aggregation, where the aggregation result changes over time. Therefore, a grid design (Fig. 8(D–G)) is used to compare the historical selection. When analysts adjust the sliding window (Fig. 3(A_3)). Movement aggregation result changes in the map view, but the previously selected region remains (Fig. 8(D–G)) for the comparison in spatial dimensions. When analyzing higher-order pattern, analysts sometimes need to analyze a continuous group of regions. Therefore, users can click one or a group of regions for analysis, and *HoLens* can view these regions as a whole to extract higher-order patterns.

Higher-order pattern visualization. Inspired by Kelpfusion [63], a novel flow visualization (Fig. 3(B_1)) called **H**igher-order **Flow** (H-Flow) is proposed to visualize higher-order dependencies from the spatial dimension (**R.5**). The details of the H-flow are as follows.

Construction scheme (Fig. 4(A)): The first node of H-Flow is designed as a bold circle, representing the beginning of the higher-order pattern (Fig. 4(A_1)). Two adjacent orders in the higher-order dependency are linked by curve (Fig. 4(A_1)): First, we rotate the midpoint p at an angle θ and then connect the two nodes. The rotated point (p') is sequentially controlled by the Bézier curve function [64]. We regulated a deflection in the same direction (clockwise/counter-clockwise) to represent the direction. Similar to Ref. [1], we provide a

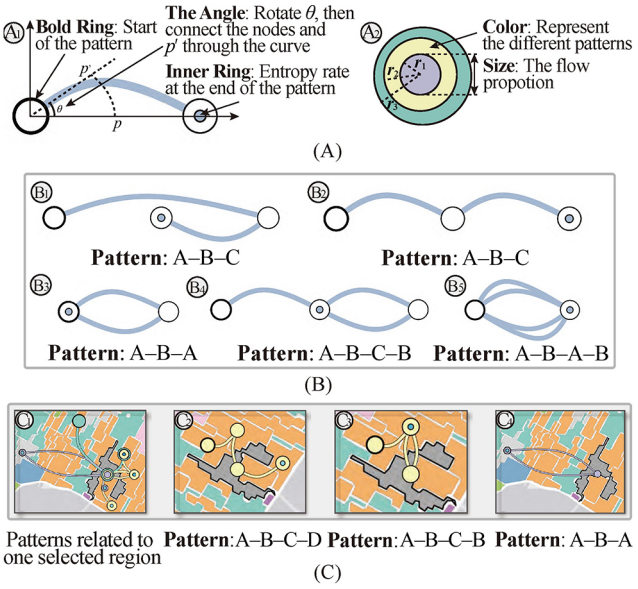


Fig. 4 Illustration of H-Flow. (A) The design scheme of the H-Flow, which aims to represent the higher-order dependencies in (B) the higher-order pattern from the (C) spatial dimension.

representation of the entropy rates [2, 18] of each extracted higher-order pattern, indicating certainty, that is, the more certain, the lower entropy rates. To identify the end state of the pattern, we nested a small circle (Fig. 4(A_1)) in the last state of the pattern, that is, color saturation, to indicate the certainty of the entire pattern. Additionally, to avoid the loss of information caused by stacking at the junction when an intersection exists, we use the radius to indicate the proportion of the flow, which is the larger flow on the lower layer and is distinguished by different colors (Fig. 4(A_2)).

Details of the H-Flow (Fig. 4(B)): This research focuses on exploring second-/third-order patterns (more discussion in Section 7.2). Therefore, according to real cases (Section 6.1) and feedback from expert interviews (Section 6.3), we summarized several higher-order pattern modes to regulate the design principles for H-Flow. The first mode is the linear mode, e.g., patterns A − B − C and A − B − C − D. The second is the annular mode, in which the higher-order contains a cyclicity, and the destination of the cyclicity is the end of the pattern, e.g., pattern A − B − A. Here, we obey the construction scheme for these two situations; examples are shown in Fig. 4(B_1, B_2, and B_3). However, for other situations, e.g., patterns A − B − A − C and A − B − A − B, the destination of the cyclicity is not the end of the pattern. Thus, to distinguish the different flows emitted from the node with the same direction, the second rotation at an angle of $2/3\theta$ is made. Thus, it can distinguish between different flows in higher-order patterns, e.g., Fig. 4(B_4, B_5). Furthermore, the corresponding higher-order patterns are generated in the state transition view (Fig. 3(D)) in the form of a higher-order state sequence chart (Fig. 3(D_2)) when clicking on the node of H-Flow. Moreover, to avoid interference from multiple colors that represent the different POI categories, we added the function of "show boundary". The map view retains the boundary of the segmentation region (Fig. 3(B_2)).

5.3 Statistic view

The statistic view aims to show the detailed statistical information of the selected region intuitively (**R.5**). In *HoLens*, we use the tornado diagrams (Fig. 3(C_1)) to represent the proportion of POI categories in the selected region. Given that the number of POIs and the access frequency are of different orders of magnitude, these two parts were normalized to make the comparison more intuitive. Analysts can sort according to different demands (POI descending or access descending) for further analysis. Furthermore, in the statistic view, two tornado diagrams represent statistical information from different views to provide an intuitive visual comparison. The upper (Fig. 3(C_1)) part displays the statistical information of the selected region in the map view. In contrast, the lower (Fig. 3(C_2)) part displays the information of the selected circular glyph in the state transition view, which is convenient for analysts to compare different regions in one pattern.

5.4 State transition view

The state transition view Fig. 3(D) aims to show the details of higher-order patterns (**R.4**) and support comparison between different higher-order patterns (**R.5**). In this view, we designed a higher-order state sequence chart (Fig. 5) that can not only represent the higher-order state but also reveal the flow change between the two states.

The higher-order state sequence chart is composed of nodes and edges. The node is a circular glyph (Fig. 5(A_1)), and each node represents one region. The sector of the circular glyph is the distribution of the flow for each POI category. The edge (Fig. 5(A_1)) between two adjacent nodes represents the flow between two regions. Compared

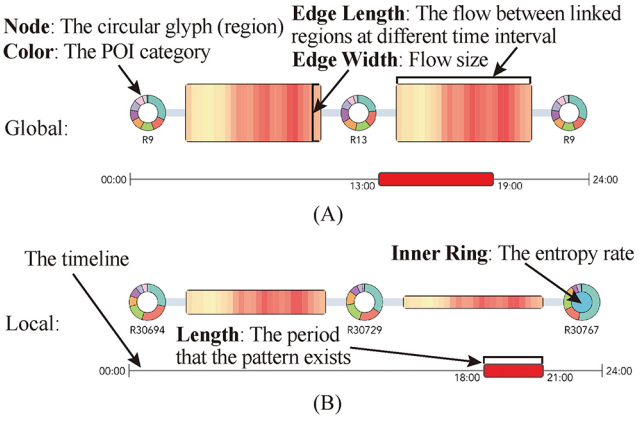


Fig. 5 Illustration of the higher-order state sequence chart. The higher-order state sequence chart includes two types: (A) the global and (B) the local higher-order state sequence chart.

with the conventional Sankey diagram [40], we first use a rectangle (Fig. 5(A)) located at the center between two adjacent nodes instead of at the edge of the Sankey diagram to represent the flow. The rectangle width indicates the overall flow between two adjacent nodes. In accordance with Zeng et al. [65], a heatmap was applied to represent the flow change between the two regions in each time interval. It can also reveal the difference between first-order dependence and higher-order patterns at the temporal dimension. Given the temporal feature of the higher-order pattern, we add a timeline (Fig. 5(B)) under the higher-order state sequence chart to represent the temporal dimension in which the current higher-order pattern occurs. The design of the two pattern types in the state transition view is similar. The only difference is that, at the end of the local higher-order pattern, we add a colored dot (Fig. 5(B)) to show the entropy rate of the local higher-order pattern, which is consistent with the design of H-Flow (Fig. 3(B_1)).

Specifically, the H-Flow shows all the local higher-order patterns related to the selected regions as a collection of higher-order patterns. Correspondingly, the collection of the higher-order state sequence chart (Fig. 3(D_1)) is generated in the right of the state transition view when the collection of local higher-order patterns related to the selected region is visualized on the map. Analysts can select the higher-order pattern they want to explore by clicking the corresponding circular glyph in this collection. A higher-order sequence chart can be generated in two ways. The first is clicking the corresponding period of the higher-order patterns extracted in the grid matrix in the overview. The other is to click the corresponding node of H-Flow in the map view for further exploration. The state transition view also supports the comparison mode between different higher-order patterns. A higher-order state sequence chart is generated in the state transition view whenever the analyst selects a higher-order pattern. Analysts can simultaneously visualize multiple higher-order patterns to compare their flow distribution, POI composition of the regions participating in the higher-order pattern, formation time of the higher-order pattern, and temporal distribution of the flow in two adjacent regions.

5.5 Alternative designs

We modified and iterated the design by communicating closely with domain experts. Figure 6(A) is the first attempt for visualizing the global higher-order pattern. We referred to the design of a conventional Sankey diagram [40] and modified it to satisfy our requirements. We aggregated the same regions at the same order level and visualized them as one node. A rectangle located at the center between two adjacent nodes was used to represent the flow. The red rectangle at the first edge represents the period of the pattern. Figure 6(A_2) shows the design of local higher-order patterns, in which a circular glyph is used to show the detail of the POI composition in each region. However, the global and local higher-order patterns are essentially state transitions. That is, to move from one state (region) to another with sequence and temporal variability. Therefore, this design may be redundant and many higher-order patterns will generate visual clusters owing to the intersection of the links. Thus, we chose

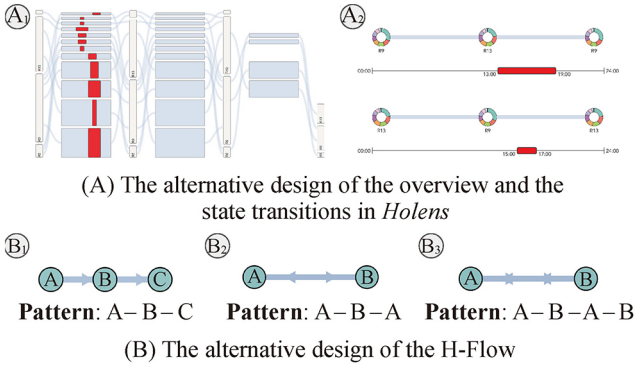


Fig. 6 The alternative designs. (A) is the overview and the global higher-order state sequence chart to visualize the higher-order patterns. (B) is the initial design of the H-Flow.

the current design in Fig. 7(A) as the final design.

The initial design (Fig. 6(B)) for the H-Flow is to link the adjacent nodes directly through a straight line and places an arrow one-third from the destination to indicate the flow direction. However, during the interviews with domain experts, we found that the higher-order pattern may sometimes exist in the form of a cyclicity, e.g., A − B − A, A − B − A − B, or A − B − A − C, in which the origin and destination of the higher-order pattern are the same. Therefore, when visualizing patterns such as A − B − A, the two directions of the H-Flow coincide and too many arrows may cause visual clutter. Therefore, a curve is used instead of a straight line to represent the flow between adjacent nodes (Fig. 4). Details of the final version design of the H-Flow are provided in Section 5.2.

5.6 Interaction

The interaction of *HoLens* can be summarized as follows.

Filtering and modifying to support the exploration details-on-demand. Users can filter and modify the data, enabling them to focus the information according to their interests. For example, a bi-directional slider can be used to filter the period when people stay in the POI and modify the scale of clustered movements on the map. *HoLens* supports exploration information on demands at different levels, at different granularities, and in different regions. Users can select and operate in every view, hover the AOI to explore the details, and brush the sliding window to filter according to their demands. *HoLens* also supports the simultaneous exploration of global and local higher-order patterns.

Linking and highlighting visualizations. In *HoLens*, views are linked to each other. For example, when an analyst hovers over the circular glyph of a higher-order state sequence chart in the state transition view, the corresponding region in the map view is highlighted automatically, which facilitates the users' understanding of the spatial information.

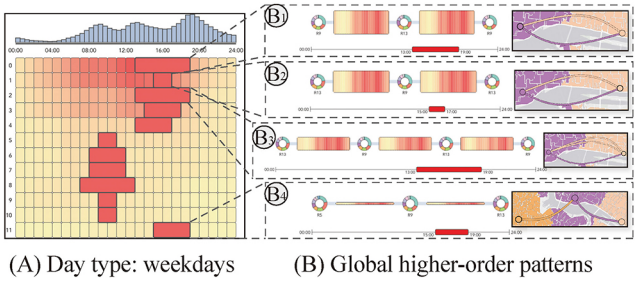


Fig. 7 From the exploration of global higher-order patterns in (A) the global controlled temporal distribution matrix, we summarize two modes: the annular higher-order pattern (e.g., B_1, B_2, and B_3)) and the linear higher-order pattern (e.g., B_4).

6 Evaluation

This section conducts the case studies (Section 6.1 and Section 6.2) manipulated by one of the collaborators (**E.A**) and the feedback from the interview of domain experts. The real-world check-in dataset [66, 67] in New York City consists of 978,054 check-in records from 3rd, Apr. 2012 to 29th, Jan. 2014.

6.1 Study 1: Adaptive organization and hierarchical multi-scale aggregation

This case aims to analyze how *HoLens* aggregates urban movements. **E.A** first set the weekdays as the analytical day type. He then filtered the stay period for every movement as a default setting. After clicking the loading button, he found that the extracted higher-order patterns for the global scale were concentrated in the morning and the evening from (Fig. 7(A)). Furthermore, except for these two periods, he also found from (Fig. 3(A_3)) that there are many active people at noon. Moreover, he found that higher-order patterns were mostly concentrated in **R9** and **R13**, which correspond to midtown and downtown Manhattan. In this case, he selected **R9** and **R13** as the Area of Interest (AOI) for exploration, and the analytical periods were morning, noon, and evening.

Adaptive organization. First, **E.A** slide the sliding window to the morning (7:00–10:00 on weekdays). The result of the movement aggregation is shown in Fig. 8(A_1). He then arbitrarily selects one region (Fig. 8(A_2)) for further exploration. The movement modeling and region segmentation results are shown in Fig. 8(A). By glancing at the statistic view (Fig. 8(A_3)), **E.A** found that the distribution of POI shows that the dominant POI in this region is *Food*, whereas the results based on access frequency are classified in the category *Travel & Transport*. **E.A** commented that the set period (7:00–10:00 on weekdays), in which people usually commute to work at this time and check in at bus stops/metro stations leads to the access frequency of *Travel & Transport* being much higher than any other POI

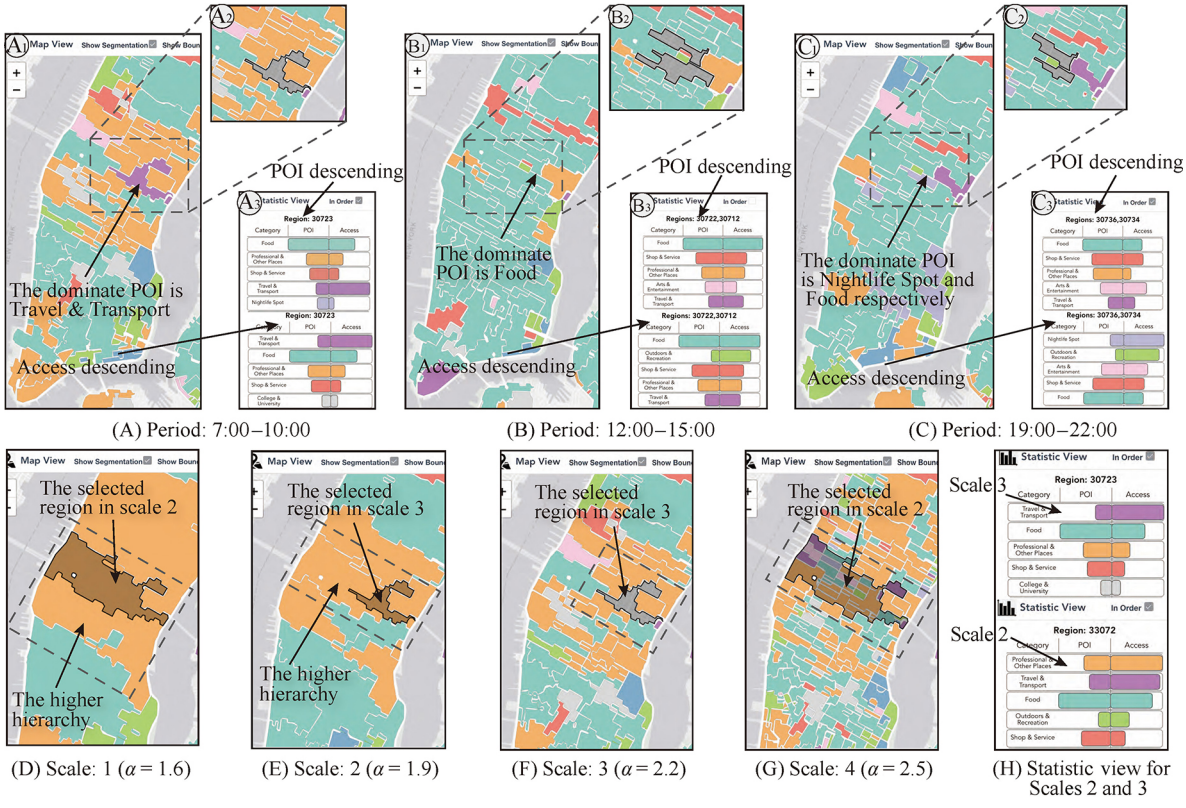


Fig. 8 The movement modeling algorithm supports adaptively organizing the movement data with temporal variability (e.g., (A) 7:00–10:00, (B) 12:00–15:00, and (C) 19:00–22:00). Moreover, the algorithm also supports hierarchical multi-scale modeling (e.g., (D) scale: 1, (E) scale: 2, (F) scale: 3, and (G) scale: 4), which the lower-level inherit the higher one.

category. Subsequently, he moved the sliding window to another period (12:00–15:00). The movement aggregation and region segmentation results are shown in (Fig. 8(B_1)). He found that the result for the segmented regions changed and the regions he previously selected (Fig. 8(A_2)) were no longer the same. The category of this newly selected region is *Food* (Fig. 8(B_3)) by considering the weekday noon, which indicates that people in this region tend to have some food for lunch. Moreover, **E.A** then set the sliding window to the period at night (19:00–22:00). The movement modeling results are shown in Fig. 8(C_1). The result in the related region (Fig. 8(C_2)) is categorized as *Nightlife Spot* (Fig. 8(C_3)). **E.A** explained that according to his living experience in New York, this situation occurs because there are also many activities happening in the nightlife spots in this area after work.

Hierarchical multi-scale movement aggregation. In this exploration, **E.A** first changed the scale of movement aggregation using a switch button in the map view. The movement aggregation result when he fixed the analytical period by the sliding window (07:00–10:00) and set one scale (Scale 3) is shown in Fig. 8(F). Then, **E.A** maintained the period constant and set another scale (Scale 2). The results of the new scale are presented in Fig. 8(E). The higher hierarchy contains the lower hierarchy for the same region he selected. Furthermore, from the statistic view in Fig. 8(H), he found that the proportion of the POI category has changed because of the change of the movement modeling range after changing the scale. As shown in Fig. 8(D) and Fig. 8(G), the new movement aggregation range (higher hierarchy) inherits and expands the original movement aggregation range (lower hierarchy) by changing the other scales under the same period.

E.A commented that the aggregation method could inspire him to make decisions on the planning of urban emergency response facilities. This method considers the contextual information to aggregate movements with more similar functions. He said that with this aggregation method, he could provide an overview of the type of activity in the selected region. **E.A** gave an example: the function in the daytime is a transportation hub and business area, whereas, at

night, the region tends to have more role in nightlife spots. Therefore, according to the information gained, they can pay more attention to fire control and public security at night in such regions to reduce the risk of fire and public safety that recreational activities may bring about. Furthermore, the hierarchical multi-scale region segmentation method can provide intuitive representations to fit their analysis at different demands.

6.2 Study 2: Interactive exploration and comparison

This case demonstrates the effectiveness of higher-order pattern analysis (requirement **R.3**, **R.4**, and **R.5**), which aims to show how analysts explore and analyze higher-order patterns (global/local) to make decisions.

Global higher-order pattern exploration. E.A first selected holidays as the exploration day type, showing that the global higher-order patterns are all concentrated in the afternoon and evening in the temporal dimension (Fig. 3(A_4)). In contrast, when **E.A** selected weekdays, higher-order patterns were distributed in the morning and afternoon (Fig. 3(A_2)). **E.A** commented that human activity in urban areas usually starts early in the working days, whereas people usually choose to relax during holidays. For detailed exploration, when **E.A** clicked the patterns in the temporal distribution matrix one by one, he found that most of the patterns were associated with **R9** and **R13** (Fig. 3(B_3)), where these two regions are downtown and midtown Manhattan. In the state transition view (Fig. 7(B_1)), the statistical result shows that the dominant activities for humans in **R9** are shopping and having meals (the proportion of arc). Furthermore, **R13** is considerably different from **R9** which *Outdoors & Recreation* in **R13** accounts for a significant proportion. **E.A** mentioned that this is because **R13** includes Central Park. After exploring all higher-order patterns in holidays and weekdays, it was found that although most global higher-order patterns were concentrated in Manhattan, the temporal dimensions of these patterns were different. In addition, the details of the functions in these regions varied in different day types. **E.A** said this might increase their confidence to prioritize the planning of Manhattan to make a more effective urban area. Furthermore, **E.A** summarizes two different modes: the annular higher-order pattern (the relationship between two adjacent regions, e.g., Fig. 7(B_1, B_2, and B_3)) and the linear higher-order pattern (e.g., Fig. 7(B_4)). These modes vary because, on the city scale, many people's activities can be completed inside the segmented region. Therefore, the global higher-order pattern mode was relatively simple.

Local higher-order pattern exploration. According to global higher-order patterns, **E.A** selected the research area in Manhattan. He set scale 3 as the default setting to explore the local higher-order pattern. He first slid the sliding window to 07:00–24:00 (Fig. 3(A_3)) to explore how the higher-order patterns happened in one working day. He arbitrarily selected one region (Fig. 3(B_1)) in the midtown area of Manhattan. From the bi-directional bar chart (Fig. 3(A_3)), **E.A** found that the majority of people check in for meals in this area, whereas the people who check in at the station dominate during the morning and evening commutes (i.e., commuting by public transportation in the morning and night). According to H-Flow (Fig. 3(B_1)), **E.A** found that all corresponding regions were not connected with the selected regions. **E.A** said this is mainly because of the developed urban transportation, which people can use to commute to their destination far away. He then chose two patterns that share the same starting point (Fig. 3(B_1)). The two patterns start at **R30884** and pass through **R30881**; however, two different destinations appear at the end. The edge (Fig. 3(D_2)) shows that the flow between the two regions is large in the morning and evening, respectively. **E.A** said that given the close distance between the endpoints, in combination with the flow between the two locations, it is an A − B − A pattern. The reason why the endpoints differ is that people rarely check in at home. Users usually check in outside areas such as dining, traveling, and shopping. Therefore, he was inspired by the urban planning task, which can further explore the two regions to optimize commuting time in terms of public transportation scheduling between the two regions, such as increasing public transportation shifts between the two regions during the morning and evening rush hours. **E.A** then changed the period to 07:00–10:00 (Fig. 9(A)). As expected, the segmentation changes and the selected region performs the function *Travel & Transport* (Fig. 9(B)). Most surroundings (Fig. 9(B))exhibit the function of *Professional & Other Places*, which is a

Fig. 9 The controlled timeline (A) and the selected region (B) perform the function of *Travel & Transport* when the period to 07:00–10:00.

business function. In addition, the regions associated with this area are similar. Combined with the inflow from the timeline (Fig. 9(A)), the transportation POI dominates during this period. **E.A** believes that this area reflects more traffic attributes in this period, that is, people come here through public transportation on weekdays and then work around this region.

E.A mentioned that the *HoLens* inspired his work on traffic management and urban planning. From the statistical view, he concluded that the selected region in the case study contains the main functions of *Travel & Transport*, *Food*, and *Professional & Other Places*, and the dominant function changes the time, i.e., it functions as a transportation hub in the morning and evening, mainly serving as an office and catering during the other time. In addition, by exploring higher-order patterns, **E.A** found related areas, such as **R30884** and **R30881**, where people were more willing to go after arriving at this location. As **E.A** is mainly engaged in urban safety and urban traffic planning, it can plan public transport routes in relevant areas in the morning and evening, such as increasing transport capacity and optimizing scheduling, and reasonably plan some urban emergency safety facilities, such as fire hoses and shelters, in the case of disasters.

6.3 Expert interview

We interviewed three independent domain experts to evaluate and provide suggestions for this study. The first expert (**E.A**) was a research assistant professor at University S whose research lies on urban computing. The second expert (**E.B**) was an assistant professor at University T specializing in urban visualization and intelligent traffic. The third expert (**E.C**) was an analyst at a consulting firm with a degree in urban planning. **E.A** and **E.B** have Ph.D. degrees, and **E.C** has an M.Sc. degree and over four years of experience in urban planning. Moreover, **E.B** and **E.C** have living experience in New York, which is the focus of this case study. First, we briefly introduce our research and show them a demo video of the *HoLens*. We collected feedback and summarized it from the aspects of feasibility, usability, and effectiveness. All the experts provided valuable feedback and suggestions based on their backgrounds. A summary of the interview is provided below.

Feasibility & usability. The three experts agreed that the aggregation method in *HoLens* was a good attempt. The visual analytics on higher-order patterns based on the aggregation method enlightens urban planning. This study is beneficial to both research and work. **E.A** commented that "the *HoLens* considers both the spatial proximity and the spatial contextual information when aggregating the data. The aggregation method is self-organized and adaptive, which fits the real demand when analyzing higher-order pattern at different scales in urban area." **E.B** mentioned that "the research considers the temporal feature and segmented functional regions by considering dynamic access frequency rather than only static POI distribution, it is of great research value." They all agreed that our visual analytics on analyzing higher-order patterns had a significant reference value. **E.B** said that "the visual effect of region segmentation on the map view is intuitive, the design of the H-Flow and the higher-order state sequence chart is useful on representing the higher-order patterns and can provide the function on visual comparison. I can follow the scenario easily, and the visual analytical process of the proposed method is feasible, which can truly help in urban decision-making."

Effectiveness. All three experts agreed that the visual analytics approach was effective for their work. **E.B** said that "the *HoLens* is enlightening to my current research on infectious disease prevention and control. Your method can segment the region according to the real access situation, and the range of the region segmentation is dynamic considering the temporal variability. Moreover, we can learn from the extracted higher-order patterns that those people who appear in a specific place at a specific time, where do most of them come from and are more

likely to go." **E.C** shared his project experience for optimizing the taxi route planning strategy. This project aims to provide strategies for taxi companies to reduce the no-load rate of taxis. They applied an algorithm based on the Monte Carlo search tree. However, this method requires a rich mathematical foundation to understand the model. In addition, their model hardly provides strategy under different scales and cannot provide more detailed information on customers, e.g., whether the customer takes a taxi after drinking. Moreover, **E.C** said that *HoLens* could improve the interpretation according to the interaction with the UI and provide an intuitive visualization of what people in the current location tend to do and where they tend to go next.

Suggestions. The experts provided fruitful suggestions on the design and potential usage directions. **E.A** and **E.C** commented that the design of the statistic view could be improved. The initial design of the statistic view only represents the statistical information of the access frequency. Nevertheless, the experts reminded us that visualizing the POI statistic of the selected region is also essential, and we can make a detailed comparison. Furthermore, **E.A** said that we must better distinguish between the origin and destination of higher-order patterns because this is crucial in an annular higher-order pattern mode. Therefore, we used tornado diagrams in the statistic view to represent the statistic information of POI categories and POI access frequency. In addition, a small circle in the last node is nested to represent the entropy rate of the higher-order flow, and the origin node is boldened to better identify the origin and destination of the higher-order pattern. **E.B** combines his current research and gives a possible application scenario. He suggests that we can integrate other spatial-temporal urban data, e.g., telco data, which can be used for epidemiological investigation. The self-organized aggregation method can segment the risk area according to the spatial context and the people's accessibility, not just spatial proximity. In addition, a higher-order pattern exploration pipeline can help analysts predict the potential activity scope of an infectious source and infer the region where the infectious source may have passed. Moreover, all experts said that the computational efficiency of the modeling algorithm could be improved.

7 Discussion

This section discusses parameter selection (Section 7.1) and the determination of the order number (Section 7.2). Then, we point out the limitations from the data perspective (Section 7.3) and provide some inspirations of the multi-variant visualization (Section 7.4).

7.1 Parameter selection

In this research, two parameters determine the performance and visual effect, that is, the aggregating threshold α and the aggregation range β. α controls the entropy threshold when aggregating the regions, which is the scale of the aggregation. Intuitively, the smaller α, the easier the aggregation. β controls the aggregation process within a reasonable spatial range, making the visual effect of the aggregation neither too big nor too small in size. The parameter selection may affect users' decision-making.

Figure 10 is the different combinations of the α and β. We selected one region in **R13** as the AOI and the exploration period was 18:00–20:00. We selected 3–5, 3–7, 3–9, and 5–9 as the candidates of β. As shown in Fig. 10, different combinations exhibit different visual effects. In Fig. 10(A, E, and I), the aggregation effect is not evident when the scale is small (i.e., $\alpha = 2.5$). A similar problem is also in Fig. 10(J), where β is 3–7. Compared with Fig. 10(C, G, and K) ($\beta = 3$–9), it can be concluded that if the upper limit of β is too small, aggregation will not be evident when α is relatively large. However, when the upper limit is reasonable, the lower limit leads to a large aggregation when α is

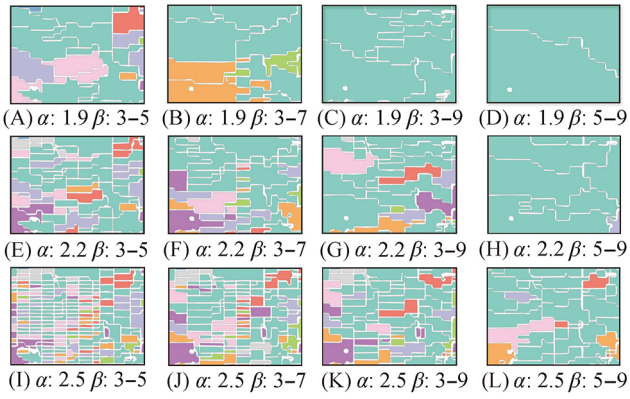


Fig. 10 Parameter selection: (A)–(D) are the aggregating range β with different range with the aggregating threshold $\alpha = 1.9$; (E)–(H) are the aggregating range β with different range with the aggregating threshold $\alpha = 2.2$; (I)–(L) are the aggregating range β with different range with the aggregating threshold $\alpha = 2.5$.

small (e.g., Fig. 10(D)) because when α is large, the aggregation result is relatively large, and the method holds that the larger hierarchy will aggregate on the basis of the previous hierarchy. Moreover, when the difference between the upper and lower limit is too large, the aggregation of the same hierarchy may differ significantly.

The spatial range ensures that the aggregated region does not become extreme because of the setting of α; that is, the area of the aggregated regions varies. This variation may mislead decision-making when the size of the segmented region varies significantly and exists on a map simultaneously. This study eventually selects 3–9 as the β, by considering the visual effects on the selected regions. In the case studies, we focused on exploring midtown and downtown Manhattan, where data are the most abundant. In addition, blocks in this region were divided into similar areas. However, in other city regions, e.g., in Brooklyn, the blocks are not divided regularly; thus, the aggregation range is difficult to control flexibly. Therefore, finding a reasonable parameter β to adapt to regular blocks in Manhattan and irregular blocks in other areas simultaneously is difficult. Moreover, when a region is aggregated at different scales, the larger scale is aggregated based on the smaller scale. Therefore, ensuring that all scales remain stable in terms of visual effect by setting one aggregating range β is also challenging. Thus, $\beta = 3$–9 is not optimal, which seems reasonable in our case studies. An algorithm that can adaptively control the aggregation range according to the size of the aggregation region is necessary in the future; however, it is not involved in this study.

7.2 Order number determination

In this research, the "higher-order" is a multi-step propagation. However, the number of orders is a complicated problem that requires long pathways to distinguish the real effects owing to a lack of data [4]. Previous studies have had many different viewpoints, e.g., Rosvall et al. [18] mentioned that the second-order is statistically significant on ranking and spreading dynamics, whereas Tao et al. [1] claimed that it could depend on up to five. Our research applies the KLD to measure the probability distribution changes and determine whether the extraction of higher-order patterns stops. However, these criteria depend on the threshold setting.

We selected these periods (8:00–10:00, 12:00–14:00, and 18:00–20:00) and several arbitrary regions in **R13** to explore the number of orders. We counted the number of flows in each pattern to observe the statistical effect on different orders. Figure 11 shows the results of average flow statistics of higher-order patterns of different numbers of orders by both the method and the Markov dynamics [1]. We found that when the number of orders increased to four in both methods, a significant decline occurred in the flow in the higher-order pattern. In other words, as the number of orders increases, each higher-order pattern becomes unique. Therefore, the flow with the same trajectory in a certain period decreased. To meet the real situation, *HoLens* also considered the flow volume. If a higher-order pattern contains few trajectories, it still does not consider a higher-order pattern. According to Fig. 11, the order numbers are limited to 3 at most, which means that we focus only on the second- and third-order in *HoLens*. Nevertheless, during this research, an automatic algorithm for determining the number of orders remains in urgent demand for future exploration.

7.3 Data perspective

This study used check-in data from Refs. [66, 67]. This dataset labels what people are doing at that time in the checked timestamp. It is a digital footprint for the people. In contrast to taxi trajectory data, in which the sensor may sample at regular temporal intervals, human movement check-in data are sampled by humans. Thus, the temporal interval for the data is random, but each sampling point has a real meaning. In this study, the analysts did not conduct a one-day or real-time analysis of traffic management and urban planning tasks. Only two types of days were considered: weekdays and weekends. Thus, when we overlaid these data into two different day

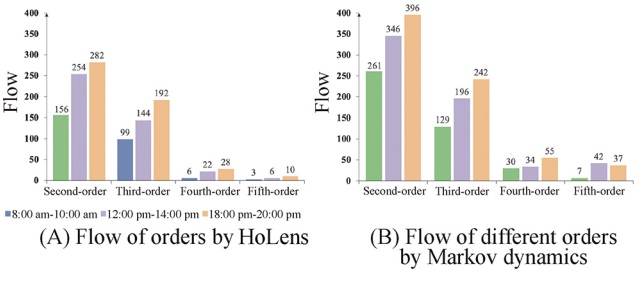


Fig. 11 Results of average flow statistics of higher-order patterns of different numbers of orders by (A) our proposed method and (B) Markov dynamics.

types, the dataset became relatively dense. However, these data form sequences of points and not a series of continuous stay periods. Therefore, we deduced the stay for each person at one location using the shortest path (network distance) between two consecutive check-in points. We ignored the commuting time if two consecutive points were in the same building. Humans have various methods of commuting between two check-in points, i.e., by vehicle or walking, and some may use public transportation. Although one cannot acquire an accurate commuting time when one person moves inside a building, the deduced stay period can be just a reference, which is reasonable to a certain extent and cannot be guaranteed to be exactly correct.

In addition, as shown in Fig. 3(B), we found some blank areas without any POI label in the map view because there was no check-in data from the dataset in these regions, which is a limitation of the data itself and will be further improved with the improvement of data quality. Furthermore, the label of the dataset may have affected the analysis of higher-order pattern in the urban area. Users tend to check in public places, e.g., restaurants, metro stations, gyms, and shops. However, few of them check in at home, or at best, they check in at the station near their home, resulting in difficulty in analyzing the higher-order pattern behavior on human's commuting, e.g., the pattern home → working place → home. We believe that these limitations can be overcome by improving the data richness and quality.

7.4 Multi-variant visualization

HoLens uses a color scheme to represent the POI categories. The dataset contained nine different POIs; thus, it could select nine from the color library to represent different POI categories. However, if there are too many types of POI, the current representation becomes unreasonable. As too many colors were displayed on the map, the interface was chaotic. In addition, increasing colors lead to visual proximity between different colors, thus affecting analysts' decision-making. For such multi-variant visualization problems, several researchers (e.g., Refs. [68–71]) have focused on this. In DFSeer [71], the authors used colors to represent different categories of ML models. However, this study chooses 5 of the 22 models to visualize. Therefore, it only needs to deal with the selected five models and assigns each selected model a color that does not need to be consistent. This is not suitable according to my demand. My solution seems to be relatively suitable for the current dataset and our current demands. Moreover, the multi-variant visualization problem on a map remains a challenge to the VIS community.

8 Conclusions

In this study, we proposed *HoLens*, a comprehensive visual analytics approach for analyzing higher-order patterns in urban areas, including a novel aggregation method and a web-based visual interface. The aggregation method in *HoLens* outperforms conventional urban movement methods in terms of rationality and the visual effect of conventional movement aggregation methods in satisfying the requirement of adaptively aggregating urban movement data by considering spatial proximity and urban contextual information. Moreover, *HoLens* supports hierarchical organizing, indicating that analysts can analyze higher-order pattern at different scales. The visual design of *HoLens* satisfies the requirements of visualizing the movement aggregation, higher-order pattern extraction, visualization, and comparison. *HoLens* supports exploring the higher-order pattern in terms of temporal dimension by improving the algorithm from the conventional approach. *HoLens* consists of several novel designed visualization techniques, i.e., H-Flow and higher-order state sequence chart. In this study, we consider different combinations of parameters and alternative designs and select an optimal selection with good performance on intuitive visualization. Finally, two real-world case studies and feedback from an interview with domain experts demonstrated the feasibility, usability, and effectiveness of *HoLens*.

Acknowledgements

We thank all the domain experts interviewed in this study. We also thank the reviewers for their comments and suggestions. This work was supported in part by the Shenzhen Science and Technology Program (No. ZDSYS20210623092007023), in part by the National Natural Science Foundation of China (No. 62172398), and the Guangdong Basic and Applied Basic Research Foundation (No. 2021A1515011700).

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article. The author Huamin Qu is the Associate Editor of this journal.

References

[1] Tao, J.; Xu, J.; Wang, C.; Chawla, N. V. HoNVis: Visualizing and exploring higher-order networks. In: Proceedings of the IEEE Pacific Visualization Symposium, 1–10, 2017.

[2] Xu, J.; Wickramarathne, T. L.; Chawla, N. V. Representing higher-order dependencies in networks. *Science Advances* Vol. 2, No. 5, e1600028, 2016.

[3] Markov, A. A. *Theory of Algorithms*. Moscow: Academy of Sciences of the USSR, 1954.

[4] Chierichetti, F.; Kumar, R.; Raghavan, P.; Sarlos, T. Are web users really Markovian? In: Proceedings of the 21st International Conference on World Wide Web, 609–618, 2012.

[5] Shannon, C. E. A mathematical theory of communication. *The Bell System Technical Journal* Vol. 27, No. 3, 379–423, 1948.

[6] Grundy, E.; Jones, M. W.; Laramee, R. S.; Wilson, R. P.; Shepard, E. L. C. Visualisation of sensor data from animal movement. *Computer Graphics Forum* Vol. 28, No. 3, 815–822, 2009.

[7] Nekovee, M.; Moreno, Y.; Bianconi, G.; Marsili, M. Theory of rumour spreading in complex social networks. *Physica A*: *Statistical Mechanics and Its Applications* Vol. 374, No. 1, 457–470, 2007.

[8] Kareiva, P. M.; Shigesada, N. Analyzing insect movement as a correlated random walk. *Oecologia* Vol. 56, No. 2, 234–238, 1983.

[9] Andrienko, N.; Andrienko, G. State transition graphs for semantic analysis of movement behaviours. *Information Visualization* Vol. 17, No. 1, 41–65, 2018.

[10] Andrienko, N.; Andrienko, G.; Stange, H.; Liebig, T.; Hecker, D. Visual analytics for understanding spatial situations from episodic movement data. *KI - Künstliche Intelligenz* Vol. 26, No. 3, 241–251, 2012.

[11] Gaffney, S.; Smyth, P. Trajectory clustering with mixtures of regression models. In: Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 63–72, 1999.

[12] Gaffney, S. J.; Robertson, A. W.; Smyth, P.; Camargo, S. J.; Ghil, M. Probabilistic clustering of extratropical cyclones using regression mixture models. *Climate Dynamics* Vol. 29, No. 4, 423–440, 2007.

[13] Lee, J. G.; Han, J.; Whang, K. Y. Trajectory clustering: A partition-and-group framework. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, 593–604, 2007.

[14] Andrienko, G.; Andrienko, N. Spatio-temporal aggregation for visual analysis of movements. In: Proceedings of the IEEE Symposium on Visual Analytics Science and Technology, 51–58, 2008.

[15] Adrienko, N.; Adrienko, G. Spatial generalization and aggregation of massive movement data. *IEEE Transactions on Visualization and Computer Graphics* Vol. 17, No. 2, 205–219, 2011.

[16] Zhou, Z.; Meng, L.; Tang, C.; Zhao, Y.; Guo, Z.; Hu, M.; Chen, W. Visual abstraction of large scale geospatial origin-destination movement data. *IEEE Transactions on Visualization and Computer Graphics* Vol. 25, No. 1, 43–53, 2019.

[17] Blaas, J.; Botha, C.; Grundy, E.; Jones, M.; Laramee, R.; Post, F. Smooth graphs for visual exploration of higher-order state transitions. *IEEE Transactions on Visualization and Computer Graphics* Vol. 15, No. 6, 969–976, 2009.

[18] Rosvall, M.; Esquivel, A. V.; Lancichinetti, A.; West, J. D.; Lambiotte, R. Memory in network flows and its effects on spreading dynamics and community detection. *Nature Communications* Vol. 5, Article No. 4630, 2014.

[19] Zeng, W.; Fu, C. W.; Müller Arisona, S.; Schubiger, S.; Burkhard, R.; Ma, K. L. Visualizing the relationship between human mobility and points of interest. *IEEE Transactions on Intelligent Transportation Systems* Vol. 18, No. 8, 2271–2284, 2017.

[20] Dodge, S.; Weibel, R.; Lautenschütz, A. K. Towards a taxonomy of movement patterns. *Information Visualization* Vol. 7, Nos. 3–4, 240–252, 2008.

[21] Slingsby, A.; van Loon, E. Exploratory visual analysis for animal movement ecology. *Computer Graphics Forum* Vol. 35, No. 3, 471–480, 2016.

[22] Chen, S.; Yuan, X.; Wang, Z.; Guo, C.; Liang, J.; Wang, Z.; Zhang, X.; Zhang, J. Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data. *IEEE Transactions on Visualization and Computer Graphics* Vol. 22, No. 1, 270–279, 2016.

[23] Chen, W.; Guo, F.; Wang, F. Y. A survey of traffic data visualization. *IEEE Transactions on Intelligent Transportation Systems* Vol. 16, No. 6, 2970–2984, 2015.

[24] Andrienko, G.; Andrienko, N.; Dykes, J.; Fabrikant, S. I.; Wachowicz, M. Geovisualization of dynamics, movement and change: Key issues and developing approaches in visualization research. *Information Visualization* Vol. 7, Nos. 3–4, 173–180, 2008.

[25] Kapler, T.; Wright, W. GeoTime information visualization. In: Proceedings of the IEEE Symposium on Information Visualization, 25–32, 2005.

[26] Adrienko, N.; Adrienko, G. Spatial generalization and aggregation of massive movement data. *IEEE Transactions on Visualization and Computer Graphics* Vol. 17, No. 2, 205–219, 2011.

[27] Guo, D.; Zhu, X. Origin-destination flow data smoothing and mapping. *IEEE Transactions on Visualization and Computer Graphics* Vol. 20, No. 12, 2043–2052, 2014.

[28] Scheepens, R.; Willems, N.; van de Wetering, H.; Andrienko, G.; Andrienko, N.; van Wijk, J. J. Composite density maps for multivariate trajectories. *IEEE Transactions on Visualization and Computer Graphics* Vol. 17, No. 12, 2518–2527, 2011.

[29] Feng, Z.; Li, H.; Zeng, W.; Yang, S. H.; Qu, H. Topology density map for urban data visualization and analysis. *IEEE Transactions on Visualization and Computer Graphics* Vol. 27, No. 2, 828–838, 2021.

[30] Giannotti, F.; Nanni, M.; Pinelli, F.; Pedreschi, D. Trajectory pattern mining. In: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 330–339, 2007.

[31] Laube, P.; Purves, R. S. An approach to evaluating motion pattern detection techniques in spatio-temporal data. *Computers, Environment and Urban Systems* Vol. 30, No. 3, 347–374, 2006.

[32] Zeng, W.; Fu, C. W.; Müller Arisona, S.; Schubiger, S.; Burkhard, R.; Ma, K. L. A visual analytics design for studying rhythm patterns from human daily movement data. *Visual Informatics* Vol. 1, No. 2, 81–91, 2017.

[33] Deng, Z.; Weng, D.; Liu, S.; Tian, Y.; Xu, M.; Wu, Y. A survey of urban visual analytics: Advances and future directions. *Computational Visual Media* Vol. 9, No. 1, 3–39, 2023.

[34] Huang, X.; Zhao, Y.; Ma, C.; Yang, J.; Ye, X.; Zhang, C. TrajGraph: A graph-based visual analytics approach to studying urban network centralities using taxi trajectory data. *IEEE Transactions on Visualization and Computer Graphics* Vol. 22, No. 1, 160–169, 2016.

[35] Zeng, W.; Shen, Q.; Jiang, Y.; Telea, A. Route-aware edge bundling for visualizing origin-destination trails in urban traffic. *Computer Graphics Forum* Vol. 38, No. 3, 581–593, 2019.

[36] Pretorius, A. Visualization of state transition graphs. Ph.D. Dissertation. Eindhoven, the Netherlands: Eindhoven University of Technology, 2008.

[37] Lu, W. L.; Wang, Y. S.; Lin, W. C. Chess evolution visualization. *IEEE Transactions on Visualization and Computer Graphics* Vol. 20, No. 5, 702–713, 2014.

[38] Plaisant, C.; Milash, B.; Rose, A.; Widoff, S.; Shneiderman, B. LifeLines: Visualizing personal histories. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 221–227, 1996.

[39] Krstajic, M.; Bertini, E.; Keim, D. CloudLines: Compact display of event episodes in multiple time-series. *IEEE Transactions on Visualization and Computer Graphics* Vol. 17, No. 12, 2432–2439, 2011.

[40] Schmidt, M. The sankey diagram in energy and material flow management. *Journal of Industrial Ecology* Vol. 12, No. 1, 82–94, 2008.

[41] Ham, F. V.; van de Wetering, H.; van Wijk, J. J. Interactive visualization of state transition systems. *IEEE Transactions on Visualization and Computer Graphics* Vol. 8, No. 4, 319–329, 2002.

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

[43] Monroe, M.; Lan, R.; Lee, H.; Plaisant, C.; Shneiderman, B. Temporal event sequence simplification. *IEEE Transactions on Visualization and Computer Graphics* Vol. 19, No. 12, 2227–2236, 2013.

[44] Shen, Z.; Wei, J.; Sundaresan, N.; Ma, K. L. Visual analysis of massive web session data. In: Proceedings of the IEEE Symposium on Large Data Analysis and Visualization, 65–72, 2012.

[45] Liu, Z.; Kerr, B.; Dontcheva, M.; Grover, J.; Hoffman, M.; Wilson, A. CoreFlow: Extracting and visualizing branching patterns from event sequences. *Computer Graphics Forum* Vol. 36, No. 3, 527–538, 2017.

[46] Zeng, W.; Fu, C. W.; Arisona, S. M.; Qu, H. Visualizing interchange patterns in massive movement data. *Computer Graphics Forum* Vol. 32, No. 3pt3, 271–280, 2013.

[47] Slingsby, A.; van Loon, E. Exploratory visual analysis for animal movement ecology. *Computer Graphics Forum* Vol. 35, No. 3, 471–480, 2016.

[48] Ware, C.; Arsenault, R.; Plumlee, M.; Wiley, D. Visualizing the underwater behavior of humpback whales. *IEEE Computer Graphics and Applications* Vol. 26, No. 4, 14–18, 2006.

[49] Zhang, J.; Guo, H.; Yuan, X. Efficient unsteady flow visualization with high-order access dependencies. In: Proceedings of the IEEE Pacific Visualization Symposium, 80–87, 2016.

[50] Lu, Y.; Steptoe, M.; Burke, S.; Wang, H.; Tsai, J. Y.; Davulcu, H.; Montgomery, D.; Corman, S. R.; Maciejewski, R. Exploring evolving media

discourse through event cueing. *IEEE Transactions on Visualization and Computer Graphics* Vol. 22, No. 1, 220–229, 2016.

[51] Riehmann, P.; Hanfler, M.; Froehlich, B. Interactive sankey diagrams. In: Proceedings of the IEEE Symposium on Information Visualization, 233–240, 2005.

[52] Wongsuphasawat, K.; Gotz, D. Exploring flow, factors, and outcomes of temporal event sequences with the outflow visualization. *IEEE Transactions on Visualization and Computer Graphics* Vol. 18, No. 12, 2659–2668, 2012.

[53] Gotz, D.; Stavropoulos, H. DecisionFlow: Visual analytics for high-dimensional temporal event sequence data. *IEEE Transactions on Visualization and Computer Graphics* Vol. 20, No. 12, 1783–1792, 2014.

[54] Perer, A.; Gotz, D. Data-driven exploration of care plans for patients. In: Proceedings of the CHI Extended Abstracts on Human Factors in Computing Systems, 439–444, 2013.

[55] Zhao, J.; Liu, Z.; Dontcheva, M.; Hertzmann, A.; Wilson, A. MatrixWave: Visual comparison of event sequence data. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 259–268, 2015.

[56] Perer, A.; Sun, J. MatrixFlow: Temporal network visual analytics to track symptom evolution during disease progression. *Annual Symposium Proceedings* Vol. 2012, 716–725, 2012.

[57] Yuan, J.; Zheng, Y.; Xie, X. Discovering regions of different functions in a city using human mobility and POIs. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 186–194, 2012.

[58] Beckmann, N.; Kriegel, H. P.; Schneider, R.; Seeger, B. The R*-tree: An efficient and robust access method for points and rectangles. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, 322–331, 1990.

[59] Schwarz, G. Estimating the dimension of a model. *The Annals of Statistics* Vol. 6, No. 2, 461–464, 1978.

[60] Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* Vol. 19, No. 6, 716–723, 1974.

[61] Van der Heyden, M. J.; Diks, C. G. C.; Hoekstra, B. P. T.; DeGoede, J. Testing the order of discrete Markov chains using surrogate data. *Physica D*: *Nonlinear Phenomena* Vol. 117, Nos. 1–4, 299–313, 1998.

[62] Kullback, S.; Leibler, R. A. On information and sufficiency. *The Annals of Mathematical Statistics* Vol. 22, No. 1, 79–86, 1951.

[63] Meulemans, W.; Riche, N. H.; Speckmann, B.; Alper, B.; Dwyer, T. KelpFusion: A hybrid set visualization technique. *IEEE Transactions on Visualization and Computer Graphics* Vol. 19, No. 11, 1846–1858, 2013.

[64] Baydas, S.; Karakas, B. Defining a curve as a Bezier curve. *Journal of Taibah University for Science* Vol. 13, No. 1, 522–528, 2019.

[65] Zeng, W.; Fu, C. W.; Müller Arisona, S.; Erath, A.; Qu, H. Visualizing waypoints-constrained origin-destination patterns for massive transportation data. *Computer Graphics Forum* Vol. 35, No. 8, 95–107, 2016.

[66] Yang, D.; Qu, B.; Yang, J.; Cudre-Mauroux, P. Revisiting user mobility and social relationships in LBSNs: A hypergraph embedding approach. In: Proceedings of the World Wide Web Conference, 2147–2157, 2019.

[67] Yang, D.; Qu, B.; Yang, J.; Cudré-Mauroux, P. LBSN2Vec: Heterogeneous hypergraph embedding for location-based social networks. *IEEE Transactions on Knowledge and Data Engineering* Vol. 34, No. 4, 1843–1855, 2022.

[68] Cao, N.; Gotz, D.; Sun, J.; Qu, H. DICON: Interactive visual analysis of multidimensional clusters. *IEEE Transactions on Visualization and Computer Graphics* Vol. 17, No. 12, 2581–2590, 2011.

[69] Yang, J.; Hubball, D.; Ward, M. O.; Rundensteiner, E. A.; Ribarsky, W. Value and relation display: Interactive visual exploration of large data sets with hundreds of dimensions. *IEEE Transactions on Visualization and Computer Graphics* Vol. 13, No. 3, 494–507, 2007.

[70] Post, F. J.; van Walsum, T.; Post, F. H.; Silver, D. Iconic techniques for feature visualization. In: Proceedings of the Visualization, 288–295, 1995.

[71] Sun, D.; Feng, Z.; Chen, Y.; Wang, Y.; Zeng, J.; Yuan, M.; Pong, T. C.; Qu, H. DFSeer: A visual analytics approach to facilitate model selection for demand forecasting. In: Proceedings of the CHI Conference on Human Factors in Computing Systems, 1–13, 2020.

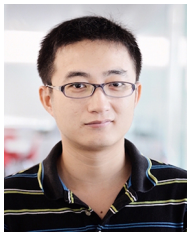
Zezheng Feng received his B.E. degree from Northeastern University (NEU), China, in 2017, his M.S. degree (with distinction) from Loughborough University, UK, in 2018, and his Ph.D. degree from the Hong Kong University of Science and Technology (HKUST), Hong Kong, China, in 2023. His recent research interests include visualization and visual analytics, explainable artificial intelligence (XAI), and urban computing. For more information, please visit `https://jerrodfeng.github.io/`

Fang Zhu received his B.E. degree in computer science and technology from the Southern University of Science and Technology, China, in 2022. He is currently working toward an M.S. degree at the Southern University of Science and Technology. His research interests include visual analytics and explainable AI.

Hongjun Wang is working toward an M.S. degree in computer science and technology from the Southern University of Science and Technology, China. He received his B.E. degree from Nanjing University of Posts and Telecommunications, China, in 2019. His research interests include machine learning, urban computing, explainable AI, data mining, and data visualization.

Jianing Hao received her B.E. degree in computer science and technology from Shandong University, China, in 2022. She is currently a Ph.D. student at the Hong Kong University of Science and Technology (Guangzhou). Her research interests include human–AI collaboration and visual analytics.

Shuang-Hua Yang received his B.S. degree in instrument and automation and M.S. degree in process control from the China University of Petroleum (Huadong), Beijing, China, in 1983 and 1986, respectively, and his Ph.D. degree in intelligent systems from Zhejiang University, Hangzhou, China, in 1991. He is currently the director of the Shenzhen Key Laboratory of Safety and Security for the Next Generation of Industrial Internet at the Southern University of Science and Technology, China, and the head of the Department of Computer Science at the University of Reading, UK. His research interests include cyber-physical systems, Internet of Things, wireless network-based monitoring and control, and safety-critical systems. He is a Fellow of IET and InstMC, UK. He is also an Associate Editor of *IET Cyber-Physical Systems*: *Theory and Applications*.



Wei Zeng is an assistant professor at the Hong Kong University of Science and Technology (Guangzhou). He received his Ph.D. degree in computer science from Nanyang Technological University and worked as a senior researcher at the Future Cities Laboratory, ETH Zurich, and as an associate researcher at Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences. His recent research interests include visualization and visual analytics, computer graphics, AR/VR, and HCI. For further information, please visit `https://zeng-wei.com/`

Huamin Qu is a professor in the Department of Computer Science and Engineering (CSE) at Hong Kong University of Science and Technology (HKUST) and the director of the Interdisciplinary Program Office (IPO) of HKUST. He obtained his B.S. degree in mathematics from Xi'an Jiaotong University, China, and his M.S. and Ph.D. degrees in computer science from Stony Brook University. His main research interests are visualization and human–computer interaction, with a focus on urban informatics, social network analysis, E-learning, text visualization, and explainable artificial intelligence (XAI). For more information, please visit `http://huamin.org/`