

HoLens: A Visual Analytics Design for Higher-order Movement Modeling and Visualization

Zezheng Feng^{1,2}, Fang Zhu², Hongjun Wang², Jianing Hao³, Shuang-Hua Yang^{5,4,2}(✉), Wei Zeng³, Huamin Qu¹

© The Author(s)

Abstract Higher-order patterns reveal sequential multi-step state transitions, which are usually superior to origin-destination analysis, which depicts only first-order geospatial movement patterns. Conventional methods for higher-order movement modeling first construct a directed acyclic graph (DAG) of movements, then extract higher-order patterns from the DAG. However, DAG-based methods heavily rely on the identification of movement keypoints that are challenging for sparse movements and fail to consider the temporal variants that are critical for movements in urban environments. To overcome the 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 design an auto-adaptive movement aggregation algorithm that self-organizes movements hierarchically by considering spatial proximity, contextual information, and temporal variability; second, we develop an interactive visual analytics interface consisting of 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 manifest that the method can adaptively aggregate the data and exhibit the process of how to explore the higher-order patterns by *HoLens*. We also demonstrate our approach's feasibility, usability, and effectiveness through an expert interview 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 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, the model cannot represent scenarios where the current status may not entirely depend 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 be traced back to Shannon's high-order memory model [5], can alleviate the problem. The analysis is crucial to many real-world applications *e.g.*, the analysis of animal behavior [6], rumor spread [7]. For example, biologists gathered the data from sensors mounted on animals and extracted higher-order dependencies to analyze behavior patterns and activities [6, 8]. By tracing multiple higher-order pathways in global ship data, species invasions are investigated and predicted [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. Additionally, in urban areas, analyzing the moving objects' multistep behavior is helpful [9]. For example, urban planners can plan public facilities and manage traffic when they understand the multistep moving behavior of the citizens; epidemiologists can trace back the possible propagation path and forecast the potential routes when making sense of the citizens' multistep moving pattern [10]. These works heavily rely 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.*, [11–13]) and aggregation-

Manuscript received: 2023-08-15 / Accepted: 2024-01-15



TSINGHUA
UNIVERSITY PRESS



Springer

based techniques (*e.g.*, [14–16]) have been proposed for keypoint identification, the methods may not be in line with contextual information that depicts underlying mechanisms for the movements. Furthermore, existing visual analytics have been proposed to explore higher-order movement patterns, *e.g.*, [17, 18]. However, these studies seldom consider 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.*, $A \rightarrow B \rightarrow X$ in period $[t_0 - t_2]$ vs. $A \rightarrow B \rightarrow Y$ in period $[t_2 - t_4]$.

This work focuses on exploring higher-order movement patterns, specifically in the context of an urban environment. We distill 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 many 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. Moreover, 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, **Higher-order Lens (HoLens)**. *HoLens* makes twofold contributions in terms of modeling and visualizing higher-order movement patterns in urban areas. First, we design a self-organization aggregation method that forms region clusters by considering the spatial proximity of the movements, contextual information, and access frequency to each location. We then construct a directed acyclic graph (DAG) with centroids of cluster regions as nodes and interactions between per-pair cluster regions as edges. In this manner, we can extract higher-order movement

1 Hong Kong University of Science and Technology. Email: zfangak@connect.ust.hk, huamin@cse.ust.hk

2 Southern University of Science and Technology. E-mail: {11711623, wanghj2020}@mail.sustech.edu.cn

3 Hong Kong University of Science and Technology (Guangzhou). E-mail: jhao768@connect.hkust-gz.edu.cn, weizeng@ust.hk

4 Department of Computer Science, University of Reading.

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

(✉) Shuang-Hua Yang is the corresponding author. E-mail: yangsh@sustech.edu.cn; Shuang-hua.yang@reading.ac.uk

Manuscript received: 2022-01-01; accepted: 2022-01-01

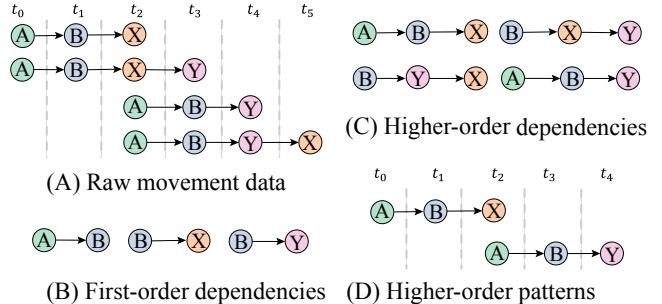


Fig. 1 Illustration of higher-order patterns: (A) Raw movement data describe the change of positions over time, modeled as trajectories. We can extract (B) first-order and (C) higher-order dependencies from these movement data. Moreover, a set of first-/higher-order dependencies can form (D) higher-order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

patterns, considering the spatial and temporal characteristics. Second, we propose a multiview visual analytics system that enables the exploration of higher-order movement patterns from the spatial and temporal dimensions. The effectiveness of *HoLens* is demonstrated through case studies conducted on real-world movements in New York City and through expert feedback.

The main contributions of this work are as follows:

- a spatial proximity- and contextual-based self-organizing method for adaptively aggregating urban movement data at different scales, which can avoid the sparseness 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 related studies in movement modeling and visualization (Sect. 2.1), state sequence (Sect. 2.2), and higher-order dependency visualization (Sect. 2.3).

2.1 Movement Modeling and Visualization

Movement data describe the change of positions 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.*, [25]) can facilitate the extraction of noteworthy patterns. However, the approach can easily cause cluttering issues. Appropriate *summaries* can address the

issues using spatial generalization and aggregation methods (*e.g.*, [26, 27]) or visualization methods (*e.g.*, [28, 29]).

Nevertheless, due to the 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.*, [30, 31]) can be applied for pattern extraction. Lee et al. [13] proposed a partition-and-group framework for clustering trajectories based on ordinary trajectory clustering algorithms to find common sub-trajectories. These methods focus on localized movement patterns, whilst neglecting the higher-order connections among the regions. Zeng et al. [32] showed that different movement rhythms are derived when the order of movement pattern is increased. Alternatively, movements can be modeled as graphs [33] and graph-based methods can be employed to extract higher-order patterns (*e.g.*, [16, 34, 35]), but these works usually neglect considering contextual information and fail to support adaptive multi-scale modeling.

In this study, we propose 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 huge amount of nodes and edges hinders analysis and insight discovery. To reduce the load, a common solution is to aggregate state sequences first, and then use visualizations like Sankey diagram [40]. Other studies (*e.g.*, [41–45]) consolidate common subsequences, extract and visualize 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 since it needs to consider the position as a constraint. Simply using graphs to visualize state transition sequences may cause the spatial context invisible. Nevertheless, researchers have come up with new visual designs to depict state sequences with 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, the methods cannot show a higher-order state and the increasing nodes will lead to visual complexity.

A novel visualization design named *higher-order state sequence chart* is developed to avoid visual clutter meanwhile helping analyze higher-order patterns. The design can express both characteristics of higher-order state sequences and temporal features of higher-order movement pattern.

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 click-streams [2]. In the higher-order dependency, the historical sequence is considered [49].

Some conventional solutions consolidate the states into a graph. The state transition diagram [41] shows all routes between nodes. Lu et al. [50] visualize the sentiment trend of time series with stacked area charts. However, these approaches can not scale well to highly-connected graphs. Sankey diagrams [51] is the choice for many studies [52–54] to show the transition pathways. In addition, matrix-based representations [55, 56] are 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 movement patterns across cities through 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 scope, while it is not suitable for visualizing various higher-order dependencies exist on one physical node together. Rosvall et al. [18] proposed a glyph that bridges the two consecutive nodes directly by the current node. In theory, this method can integrate different dependencies through the current node to build the correspondence, but the area of the current node limits the scalability of various dependencies. To handle such issues, HoNVis [1] visualizes the higher-order dependencies and supports the exploration from an overview to a fine-grained level while ignoring the temporal characteristics of the 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 (Sect.3.1) and provide an overview (Sect.3.2) of our solution.

3.1 Requirement Analysis

In the early stage of this research, we held regular meetings with the target users and two domain experts (**E.A** and **E.B**). The target users are 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, including 1) segmenting the urban area into functional regions and 2) optimizing the traffic. We summarize the requirements for this research from 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 the clusters, considering the 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 bigger one and keeping the regions' consistency at different hierarchies profoundly influences the result of analyzing the higher-order pattern.

R.3 Higher-order Pattern Extraction with Temporal Variability. Human activities usually depict temporal variability. Some activities only exist in certain periods, and some may implicitly exhibit peak and nonpeak features. Analyzing the higher-order patterns can help understand the traveling behavior and inspire them to make proper decisions, such as urban planning. Therefore, a mining method for extracting these higher-order patterns with temporal constraints is 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 estate to business district and then back on weekdays. Therefore, the global movement trend between different regions must be explored, and the 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 is 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 due to small ranges. Therefore, the flow of higher-order patterns must also 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 one node in a DAG, its meaning in a higher-order pattern differs from its physical meaning at the DAG. Usually, 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 the higher-order patterns at different temporal and geographical dimensions, it is helpful to know 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 (Sect.4) and interactive higher-order pattern visual exploration (Sect.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.

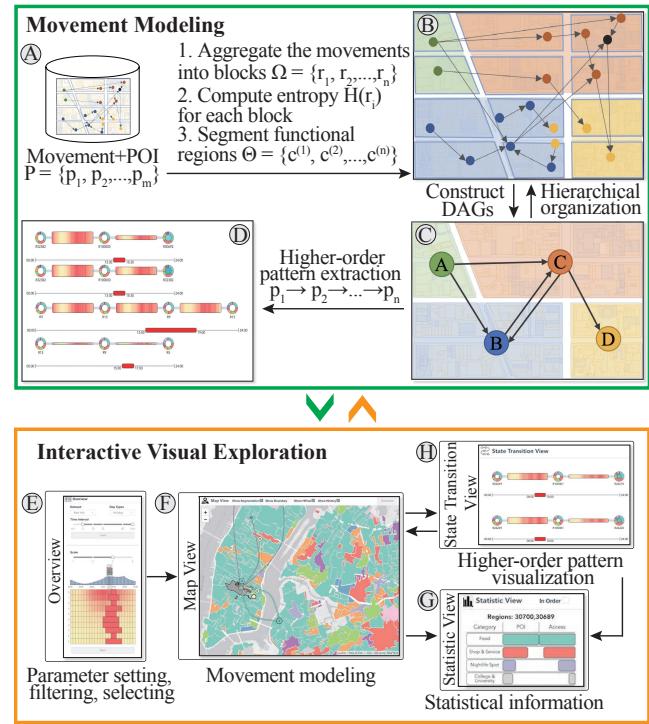


Fig. 2 The overview of *HoLens*. *HoLens* consists of two stages: movement modeling (A, B, C, and D) and interactive visual exploration (E, F, G, and H).

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 in 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, the regions can be organized at different scales. Afterward, a DAG is constructed with each region as a vertex and the flow between the regions as the edge (Fig. 2(C)). In accordance with the DAG, the higher-order patterns are extracted, with consideration to temporal variability (Fig. 2(D)).

For the interactive visual exploration phase, an interactive visual analytics interface (Fig. 3) is developed. It is a web-based multi-view interface that aims to provide interactive visual analytics for higher-order pattern exploration. Data collection and preprocessing are run 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 the global higher-order patterns in the temporal dimension. In the map view (Fig. 2(F)), users can visualize the results of movement aggregation and the spatial distribution of the 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 the map view and the state transition view.

4 Model

This section first introduces the construction of a data structure (Sect.4.1). Next, a novel method for data aggregation (Sect.4.2.1) and hierarchical movement self-organization (Sect.4.2.2) is introduced. Finally, the method for higher-order pattern extraction in this research (Sect.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 is also utilized. Although each movement data record contains a timestamp that can reflect the time point of one moving object existing in a certain position, it cannot provide the period each moving object stays in the same position.

Thus, in this research, a shortest path-based method is applied to estimate the period of one moving object’s stay in a certain position. We first form 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 the two consecutive points with large intervals have little associations, and then the sub-trajectory with just a single point is filtered out. Afterward, the travel time of two adjacent movement points is computed in accordance with the traffic condition and the optimal path by using the mainstream commercial map application (*i.e.*, the API from Google Maps^②). Finally, the moving object’s stay time at each position is deduced by subtracting the travel time between two adjacent points.

Table 1 Summary of Major Variables

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

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 here 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 [57], which employed Connected Component Labeling (CCL) on a map divided by the road network. Moreover, Entropy can be used to identify higher-order patterns, as higher-order patterns are patterns that involve diverse POIs visited. For example, a higher-order pattern in human mobility data might be a sequence of visits to different types of POIs,

^①<https://www.openstreetmap.org/>

^②<https://www.google.com/maps/>

such as a home, a workplace, and a restaurant. By combining entropy, we can easily achieve this. The scale of the region can be modified manually by selecting different types of roads (e.g., freeway, primary, secondary). Here, these regions serve as the 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 Points of Interest (POIs) located within specific regions. To achieve this, we begin by mapping 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 has a diverse set of POIs, while a region with low entropy has a more homogeneous set of data points. Subsequently, we aggregate 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 i^{th} 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 $g = Rtree(\Omega)$, where $g : p \rightarrow r$ is the function of the R tree that returns the located region of p . Thus, the projected process for region r_i can be written as

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

where Θ is the set of POI.

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

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

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 j^{th} POI category in r_i , so that it satisfies $\sum_j D_{i,j} = 1$. The entropy H serves as the degree of confusion within the region since we aim to ignore the human mobilities that have many consistent attributions. For example, such higher-order dependencies: $Tr : p_1 \rightarrow p_2 \rightarrow \dots \rightarrow p_n$, containing multiple repetitive POI types, are undesirable. The process of this step is shown in Algorithm 1.

4.2.2 Hierarchical Movement Organization

We introduce an innovative algorithm for hierarchical self-organization of movement data. This algorithm facilitates

Algorithm 1 Region Entropy Computation

```

Input: spatial region set  $\Omega = \{r_1, r_2, \dots, r_n\}$ , set of
       GPS points  $P = \{p_1, p_2, \dots, p_m\}$ .
Output:  $\Theta_r = \{c^{(1)}, c^{(2)}, \dots, c^{(n)}\}$ 
1  $g = Rtree(\Omega)$  // ST query
2 for  $r_i \in \Omega$  do
3    $O_i = \{c^{(j)} \mid g(p_j) = r_i\}, \forall p_j \in P,$ 
4    $H(r_i) = - \sum_{a_i \in D_i} P(a_i) \log_e P(a_i)$ 
5 end for
6 for  $r_i \in \Omega$  do
7    $\Theta \leftarrow \Theta + BFS(r_i, V)$ 
8 end for
9 return  $\Theta$ 

```

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, reducing spatial entropy as mentioned in Sect.4.2.1. Technologically, we concatenate those regions that present lower entropy after aggregating. Given the set $\Omega = \{r_1, r_2, \dots, r_n\}$, 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

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

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

As in R.2, *HoLens* aims to provide the multi-level region scale for systemically understanding the higher-order movement pattern. The first-order merged procedure has been introduced above; thus, for the next-level aggregation, only those 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 the high-level region styles always come from the 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 the higher-order patterns (as illustrated in Fig. 2(C) and Fig. 2(D)). In this study, we consider the temporal variability in the higher-order patterns. Thus, we first modify the transi-

^①<https://api.foursquare.com/v2/venues/categories/>

Algorithm 2 BFS

Input: spatial region set $\Omega = \{r_1, r_2, \dots, r_n\}$, entropy H , visited set $V, \Theta_r = \{\}$.

Output: $\Theta_r = \{c^{(i)}, c^{(j)}, \dots\}$

```

1 if  $r_i \notin V$  then
2   for  $r_j \in r_i.Neighbors$  do
3     if  $\frac{1}{2}(H(r_i) + H(r_j)) \geq \alpha H(r_i + r_j)$  then
4        $r_i \leftarrow merge(r_i, r_j)$ 
5        $r_i.Neighbors \leftarrow$ 
6          $r_j.Neighbors \cup r_i.Neighbors$ 
7        $V \leftarrow V + r_j$ 
8        $\Theta_r \leftarrow \Theta_r + c^{(j)}$ 
9     end if
10   end for
11   for  $r_j \in r_i.Neighbors$  do
12     BFS( $r_i, V$ )
13   end for
14 return  $\Theta_r$ 
```

tion probability for the first-order Markov model by adding a temporal constraint (**R.3**):

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

where $i_{\Delta T}^k$ indicates the k th vertex in the constructed DAG, and ΔT denotes the selected period. The edge weight $\omega(i_{\Delta T}^k \rightarrow i^{k+1})$ indicates the intensity of the connection in a certain period ΔT , which in our scenario can be assigned as the sum of pairwise connections ($i^k \rightarrow i^{k+1}$) 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 is not fixed; thus, we refer to the method of Xu et al. [2] and modify the method by adding temporal constraints.

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

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, $[D|C, B, A]$. I represents the sequence C, B, A , and $i_{\Delta T}^k | I$ indicates that D with the higher-order dependency $[D|C, B, A]$ exists in ΔT .

We take advantage of the Kullback-Leibler divergence (KLD) [62] to measure recursively 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.

5 Visualizations

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

5.1 Overview

The overview aims to help analysts configure the parameters and control the analytical process from the temporal dimension. The analysts can select the dataset, pre-configure the parameters, control the temporal process, and obtain the temporal distribution of the higher-order patterns. The overview consists of a configuration panel (Fig. 3(A₁)), a global controlled temporal distribution matrix (Fig. 3(A₂)), and a legend board (Fig. 3(A₅)).

The configuration panel (Fig. 3(A₁)) supports dataset selection, analytical day types selection (weekday/weekend), and time interval filter. The global controlled temporal distribution matrix consists of a controlled timeline (Fig. 3(A₃)) and a temporal distribution matrix (Fig. 3(A₄)) (**R.4**). The controlled timeline supports controlling the period and visualizing the flow from the temporal dimension (**R.3**). It (Fig. 3(A₃)) is composed of a time axis, a single-directional histogram, a bi-directional histogram, and a sliding window. The single-directional histogram reflects the tendency of the active movements distributed in one day. The sliding window aims to enhance the interaction when analyzing the detailed higher-order patterns. When the analysts select an AOI in the map view Fig. 3(B₁), the bi-directional histogram Fig. 3(B₃) 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 in this period, and the height of the bi-directional histogram reflects the volume of the flow in detail.

The temporal distribution matrix (Fig. 3(A₄)) 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 the 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 the analysts select one 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)) is designed to support exploring the higher-order patterns in a spatial dimension. Analysts can

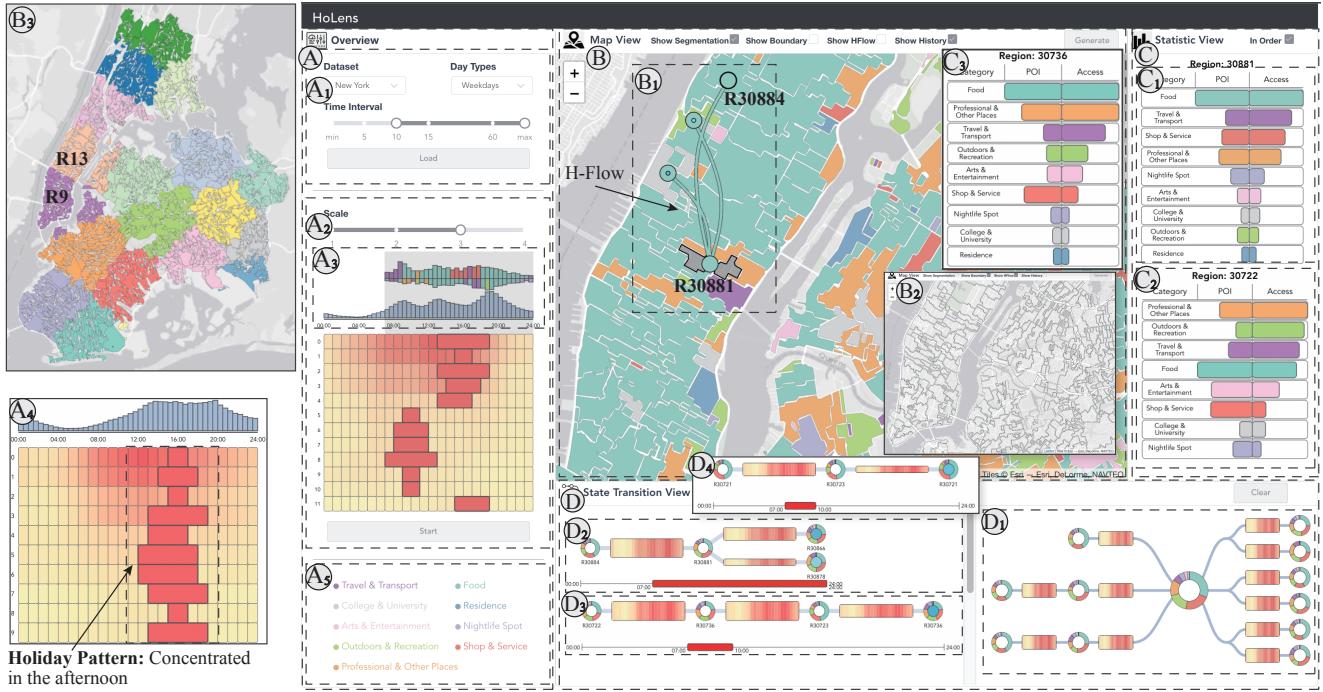


Fig. 3 *HoLens* supports visually exploring the urban higher-order patterns, which consists of four views: (A) the overview provides a configuration panel and a global representation to the end-users, (B) the map view displays the result of movement aggregation and supports visualizing the higher-order 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 of the higher-order patterns.

interact with other views (overview and state transition view) through the map view, and the result in spatial dimension can be displayed in the map view.

Region segmentation. The map view supports segmenting the urban areas, providing an intuitive representation of the movement aggregation on the map. As mentioned in Sect. 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(A₃)) in the overview) at each segmented region, enabling the analysts to understand efficiently what people in the segmented region tend to do.

Comparison & Multi-region Exploration. The map view supports comparison (**R.4**) and multi-region exploration, helping analysts compare the movement aggregation at different periods and explore higher-order patterns based on the user demand. The comparison function in the map view is different 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 Sect. 4.2.1, the method enables adaptive aggregation, and the aggregation

result changes over time. Therefore, a grid design (Fig. 8(D) to (G)) is used to compare the historical selection. When analysts adjust the sliding window (Fig. 3(A₃)). The movement aggregation result changes in the map view, but the previously selected region remains (Fig. 8(D) to (G)) for comparison in spatial dimensions. When analyzing the higher-order pattern, analysts sometimes need to analyze a continuous group of regions. Therefore, users can click one or a group of regions to analyze, and *HoLens* may see these regions as a whole to extract the higher-order patterns.

Higher-order Pattern Visualization. Inspired by Kelpfusion [63], a novel designed flow visualization (Fig. 3(B₁)) named **Higher-order Flow** (H-Flow) is proposed to visualize the 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(A₁)). For two adjacent orders in the higher-order dependency, they are linked by a curve (Fig. 4(A₁)). First, we rotate the midpoint p at an angle θ , and then connect the two nodes and the rotated point (p') sequentially controlled by the Bezier curve function [64]. We regulate a deflection in the same direction (clockwise/counterclockwise) to represent the direction. As [1], we provide a representation of the entropy rates [2, 18] of each

extracted higher-order pattern, indicating the 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(A₁)) in the last state of the pattern, that is, the color saturation, to indicate the certainty of the whole 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 distinguished by different colors (Fig. 4(A₂)).

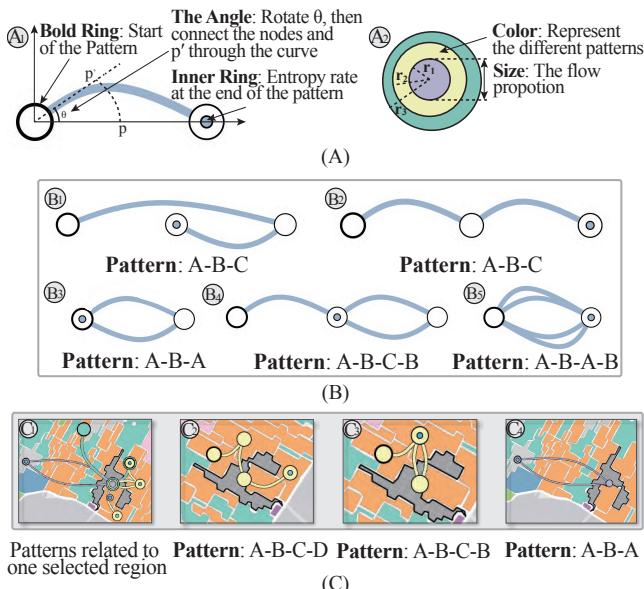


Fig. 4 Illustration of H-Flow. (A) The design scheme of H-Flow, which aims to represent the higher-order dependencies in (B) the higher-order pattern from the (C) spatial dimension.

Details of the H-Flow (Fig. 4(B)): This research focuses on exploring second-/third-order patterns (more discussion in Sect. 7.2). Therefore, according to the real cases (Sect. 6.1) and the feedback from the experts' interview (Sect. 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., pattern $A - B - C$, pattern $A - B - C - D$. 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 can be seen in Fig. 4(B₁, B₂ and B₃). 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 different flows emitted from the same node with the same direction, the second rotation at an angle of $2/3\pi$ is made. Thus, it can distinguish the different flows in the higher-order patterns, e.g., Fig. 4(B₄ and B₅). Furthermore, the corresponding higher-order patterns are generated in the

state transition view (Fig. 3(D)) in the form of the higher-order state sequence chart (Fig. 3(D₂)) when clicking the node of the 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(B₂)).

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 tornado diagrams (Fig. 3(C₁)) to represent the proportion of the 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 are 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 the statistical information from different views to provide an intuitive visual comparison. The upper (Fig. 3(C₁)) part displays the statistical information of the selected region in the map view. By contrast, the lower (Fig. 3(C₂)) 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 the higher-order patterns (R.4) and support comparison between different higher-order patterns (R.5). In this view, we design a higher-order state sequence chart (Fig. 5) that can not only represent the higher-order state but also reveal the flows' change between two states.

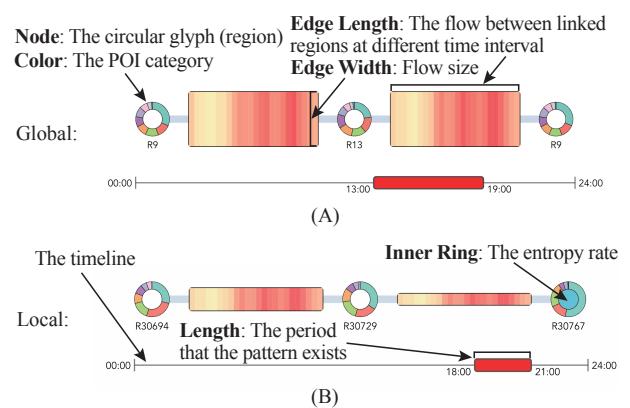


Fig. 5 Illustration of the higher-order state sequence chart. The higher-order state sequence chart includes two types: (A) the global and (B) the local higher-order state sequence chart.

The higher-order state sequence chart is composed of nodes and edges. The node is a circular glyph (Fig. 5(A₁)), and

each node represents one region. The sector of the circular glyph is the distribution of the flow for each category of POI. The edge (Fig. 5(A₁)) between two adjacent nodes represents the flow between two regions. In comparison with the conventional Sankey diagram[40], we first use a rectangle (Fig. 5(A₁)) located at the center between two adjacent nodes instead of the edge of the Sankey diagram to represent the flow. The width of the rectangle indicates the overall flow between two adjacent nodes. In accordance with Zeng et al. [65], a heatmap is applied to represent the flow change between 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(A₂)) under the higher-order state sequence chart to represent the temporal dimension that the current higher-order pattern occurs. The design of the two kinds of patterns 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(A₂)) to show the entropy rate of the local higher-order pattern, which is consistent with the design of H-Flow (Fig. 3(B₁)).

Specifically, H-Flow shows all 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(D₁)) is generated in the right of the state transition view when the collection of the 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. The higher-order sequence chart can be generated in two ways. The first one is to click 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 visualize multiple higher-order patterns simultaneously to compare their flow distribution, the POI composition of the regions participating in the higher-order pattern, the formation time of the higher-order pattern, and the temporal distribution of the flow in two adjacent regions.

5.5 Alternative Designs

We modify and iterate the design by closely communicating with the domain experts. Fig. 6(A) is the first attempt for

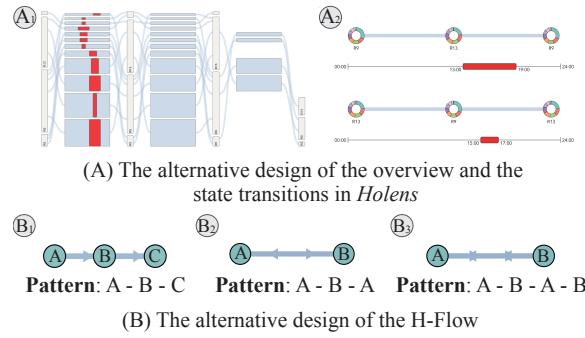


Fig. 6 The alternative designs. (A) is the overview and the global higher-order state sequence chart to visualize the higher-order patterns. (B) is the initial design of the H-Flow.

visualizing the global higher-order pattern. We referred to the design of the conventional Sankey diagram [40] and modified it to satisfy our demand. We aggregate the same regions at the same order level, visualized as one node. A rectangle located at the center between two adjacent nodes is used to represent the flow. At the first edge, the inside red rectangle represents the period of the pattern. Fig. 6(A₂) 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 due to the intersection of links. Thus, we choose the current design in Fig. 7(A) as the final design.

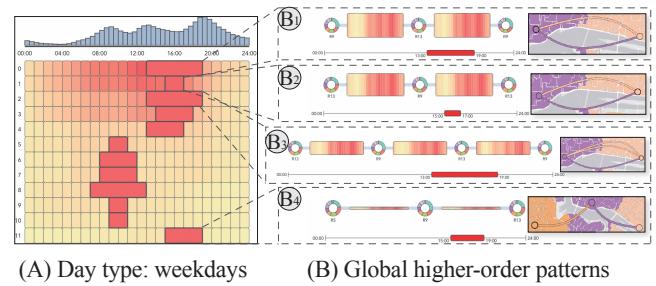


Fig. 7 From the exploration of global higher-order patterns in (A) the global controlled temporal distribution matrix, we summarize two modes: the the annular higher- order pattern (e.g., B₁, B₂, and B₃) and the linear higher-order pattern (e.g., B₄).

The initial design (Fig. 6(B)) for the H-Flow is to link the adjacent nodes directly through the straight line and place an arrow one-third from the destination to indicate flow direction. However, during the interview with the domain experts, we found that the higher-order pattern sometimes may probably 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 the destination of the higher-order pattern are the same. Therefore, when visualizing the patterns such as A - B - A,

the two directions of the H-Flow will coincide, and too many arrows may bring visual clutter. To this end, the curve is used instead of a straight line to represent the flow between adjacent nodes (Fig. 4). The details of the final version design of the H-Flow are detailed in Sect. 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, they can use a bi-directional slider to filter the period when people stay in the POI and modify the scale of the 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 demand. *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 the circular glyph of the higher-order state sequence chart in the state transition view, the corresponding region in the map view is highlighted automatically, which facilitates the user in understanding the spatial information.

6 Evaluation

This section conducts the case studies (Sect. 6.1, and Sect. 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 978054 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. Then, he 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 are concentrated in the morning and the evening from (Fig. 7(A)). In addition, except for these two periods, he also found from (Fig. 3(A₃)) that there are many active people at noon. Moreover, he found that the higher-order patterns are 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 the weekdays). The result of the movement aggregation is shown in Fig. 8(A₁). He then arbitrarily selected one region (Fig. 8(A₂)) for further exploration. The movement modeling and region segmentation results are shown in Fig. 8(A). By glancing at the statistic view (Fig. 8(A₃)), **E.A** found that the distribution of POI shows that the dominant POI in this region is *Food*, whereas the results by considering the 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 category. Next, he moved the sliding window to another period (12:00-15:00). The movement aggregation and region segmentation results are shown in (Fig. 8(B₁)). He found that the result for segmented regions changed, and the regions he previously selected (Fig. 8(A₂)) are no longer the same. The category of this newly selected region is *Food* (Fig. 8(B₃)) 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(C₁)). The result in the related region (Fig. 8(C₂)) is categorized as *Nightlife Spot* (Fig. 8(C₃)). **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 the movement aggregation through 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** kept the period constant and set another scale (scale 2). The results of the new scale are shown 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 due to the change of 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 other scales under the same period.

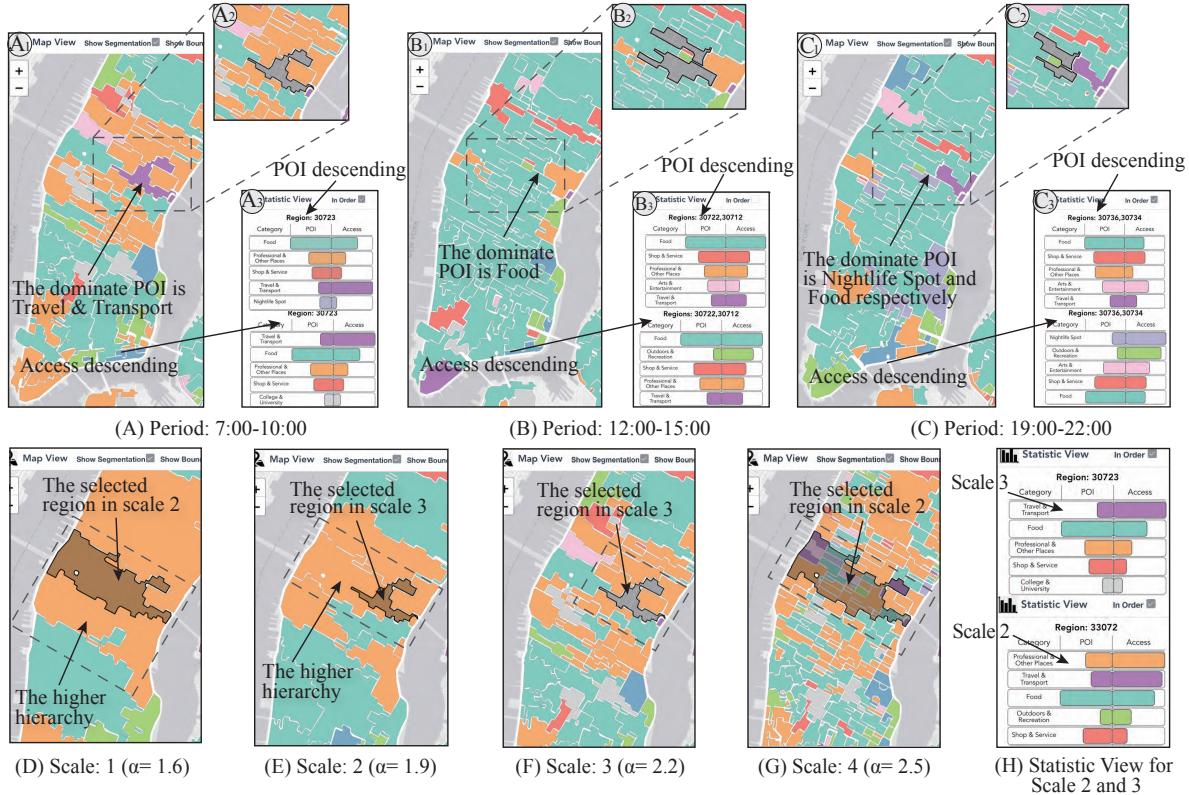


Fig. 8 The movement modeling algorithm supports adaptively organizing the movement data with temporal variability (e.g., (A) 7:00-10:00, (B) 12:00-15:00, and (C) 19:00-22:00). Moreover, the algorithm also supports hierarchical multi-scale modeling (e.g., (D) scale:1, (E) scale: 2, (F) scale: 3, and (G)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

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 can have an overview of what kind of activity is 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, they can pay more attention to fire control and public security at night in such regions, according to the gained information, so as 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 the higher-order pattern analysis (requirement R.3, R.4, and R.5), which aims to show how analysts explore and analyze the higher-order patterns (global/local) to make decisions.

Global Higher-order Pattern Exploration. E.A first se-

lected holidays as the exploration day type, and it shows that the global higher-order patterns are all concentrated in the afternoon and evening in the temporal dimension (Fig. 3(A₄)). On the contrary, when E.A selected the weekdays, the higher-order patterns are distributed in the morning and the afternoon (Fig. 3(A₂)). 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 are associated with R9 and R13 (Fig. 3(B₃)), where these two regions are the downtown and midtown of Manhattan. In the state transition view (Fig. 7(B₁)), the statistical result shows that the dominant activities for humans in R9 are shopping and having meals (the proportion of arc). Moreover, 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 he explored all the higher-order patterns in holidays and weekdays, it shows that although most global higher-order patterns are concentrated in Manhattan, the temporal dimensions of these patterns are different. In addition, the details of the function in these regions vary 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** also summarized 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 is relatively simple.



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.

Local Higher-order Pattern Exploration. According to the global higher-order patterns, **E.A** selected the research area in Manhattan. He set the scale 3, as a 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 midtown in 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 the H-Flow (Fig. 3(B_1)), **E.A** found that all corresponding regions are 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 both start at **R30884** and pass through **R30881**, but at the end, two different destinations appear. 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 the endpoints are different is that people rarely check in at home. The 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 the commuting time in terms of public transportation scheduling between the two regions, such as increasing public transportation shifts between the two regions in 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 of *Travel & Transport* (Fig. 9(B)). Most of the surroundings (Fig. 9(B)) exhibit the function of *Professional & other Places*, which means a business function. In addition, the regions associated with this area are close. Combined with the in-flow from the timeline (Fig. 9(A)), the transportation POI dominates in 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, through the exploration of higher-order patterns, **E.A** found some related areas, such as **R30884** and **R30881**, where people are more willing to go after arriving at this place. 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 case of disasters.

6.3 Expert Interview

We interviewed three independent domain experts to evaluate and provide suggestions for this research. The first expert (**E.A**) is a research assistant professor at University S, whose research lies on urban computing. The second expert (**E.B**) is an assistant professor at University T, specializing in urban visualization and intelligent traffic. The third expert (**E.C**) is 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 MSc 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 city focused on in the case study.

First we briefly introduce our research and show them a demo video of the *HoLens*. We then collect the feedback and summarize it from the aspect of feasibility, usability, and effectiveness. All the experts provided valuable feedback and suggestions according to their backgrounds. The summary of the interview is detailed below.

Feasibility & Usability. The three experts agreed that the aggregation method in *HoLens* is a good attempt. The visual analytics on higher-order patterns based on the aggregation method enlightens urban planning. This study is beneficial for their 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 the higher-order pattern at different scales in the 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 the higher-order patterns has 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 is 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. The project is to provide strategies for taxi companies to reduce the no-load rate of taxis. They applied the algorithm based on the Monte-Carlo search tree. However, the 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 the 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 the 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 the 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 just 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 need to better distinguish between the origin and destination of the 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 bold the origin node better to 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 the spatial proximity. In addition, the higher-order pattern exploration pipeline can help analysts predict the potential activity scope of the infectious source and infer the region where the infectious source may have passed in the past. Moreover, all the experts said that the computational efficiency of the modeling algorithm could be improved.

7 Discussion

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

7.1 Parameter Selection

In this research, two parameters determine the performance and the visual effect, that is, the aggregating threshold α and the aggregation range β . α controls the threshold of the entropy when aggregating the regions, which is the scale of the aggregation. Intuitively, the smaller α , the easier the aggregation. β is to control the aggregation process under 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.

Fig. 10 is the different combinations of the α and β . We select one region in R13 as the AOI, and the exploration period is 18:00-20:00. We select 3-5, 3-7, 3-9, and 5-9 as the candidates of β . As shown in Fig. 10, the different combinations have different visual effects. In Fig. 10(A, E, 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,K) ($\beta = 3-9$), it can be summarized that if the upper limit of β is too small, the 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 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 significantly differ.

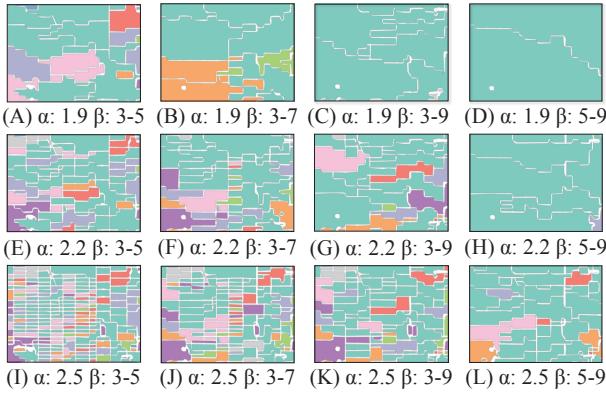


Fig. 10 Parameter selection: (A) to (D) are the aggregating range β with different range with the aggregating threshold $\alpha = 1.9$; (E) to (H) are the aggregating range β with different range with the aggregating threshold $\alpha = 2.2$; (I) to (L) are the aggregating range β with different range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

The spatial range ensures that the aggregated region will not become extreme due to the setting of α , that is, the area of the aggregated regions varies. This variation may mislead the decision-making when the size of the segmented region varies greatly 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 focus on exploring midtown and downtown Manhattan, where data is most abundant. In addition, the blocks in this region are 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 the regular blocks in Manhattan and the irregular blocks in other areas simultaneously is difficult. Moreover, when the region is

aggregated at different scales, a larger scale is aggregated on the basis of a 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 the best one, which just 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 for the future, but it is not involved in this work.

7.2 Order Number Determination

In this research, the “higher-order” is a multi-step propagation. However, the number of the order is a complicated problem that needs long pathways to distinguish the real effects due to the lack of data [4]. Previous studies have had many different viewpoints *e.g.*, Rosvall et al. [18] mentioned that 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 decide whether the extraction of the higher-order patterns stops. However, these criteria depend on our threshold setting. We select these

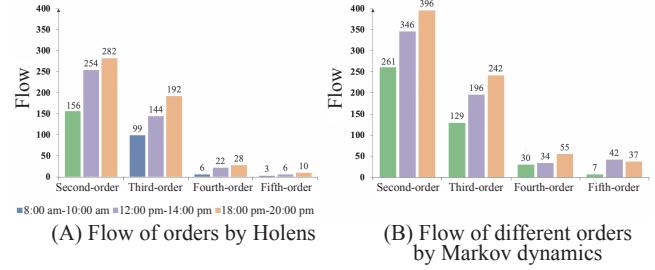


Fig. 11 The results of average flow statistics of higher-order patterns of different numbers of orders by (A) our proposed method and by (B) Markov dynamics.

periods (7:00-10:00, 12:00-14:00, and 18:00-20:00) and several arbitrary regions in R13 to explore the number of orders. We count the number of flows in each pattern to observe the statistical effect on different orders. Fig. 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 increases to four in both methods, a significant decline occurs in the flow in the higher-order pattern. That is, as the number of orders increases, each higher-order pattern becomes unique. Therefore, the flow with the same trajectory in a certain period decreases. To meet the real situation, HoLens also considers the flow volume. If the higher-order pattern contains few trajectories, it still does not consider a higher-order pattern. And according to Fig. 11, the order numbers are limited to 3 at most, which means we only focus

on the second-and third-order in *HoLens*. Nevertheless, during this research, an automatic algorithm for deciding the number of orders is still an urgent demand in future exploration.

7.3 Data Perspective

This work uses the check-in data from [66, 67]. This dataset labels what the people are doing at that time in the checked timestamp. It is a digital footprint for the people. In contrast to the taxi trajectory data, which the sensor may sample at regular temporal intervals, the human movement check-in data are sampled by humans; thus, the temporal interval for the data is random, but each sampling point has real meaning. In this research, the analysts did not conduct a one-day or real-time analysis of the traffic management and urban planning tasks. Only two day types need to be considered: weekdays and weekends. Thus, when we overlay these data into two different day types, the dataset is relatively dense. However, these data form sequences of points but not a series of continuous stay periods. Therefore, we deduce the stay for each person at one location by the shortest path (network distance) between two consecutive check-in points. We ignore their commuting time if the two consecutive points are in the same building. Humans have various ways to commute between two check-in points, *i.e.*, by vehicle or walking, and some may use public transportation. Although one cannot acquire the accurate commuting time when one person moves inside one 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 is no check-in data from the dataset in these regions, which is the limitation of the data itself and will be further improved with the improvement of data quality. Furthermore, the label of the dataset may affect the analysis of the higher-order pattern in the urban area. The users tend to check in at public places, *e.g.*, restaurants, metro stations, gyms, shops, etc. However, seldom 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 will be solved with the improvement of data richness and data quality.

7.4 Multi-variant Visualization

HoLens uses a color scheme to represent the POI categories. The dataset contains nine different POIs; thus, it can select nine from the color library to represent different POI

categories. However, if there are too many types of POI, the current representation will become unreasonable. Given that too many colors are displayed on the map, the interface is chaotic. In addition, increasing colors leads to visual proximity between different colors, thus affecting analysts' decision-making. For such multi-variant visualization problems, several researchers (*e.g.*, [68–71]) have focused on this. In DFSeer [71], the authors used colors to represent different categories of ML models. However, this work chooses 5 of the 22 models to visualize. Therefore, it only needs to deal with the selected five models and gives each selected model a color, which does not need to be consistent. This is not suitable for 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 Conclusion

In this study, we proposed *HoLens*, a comprehensive visual analytics approach for analyzing higher-order patterns in the urban area, 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 visual effect of the conventional movement aggregation methods in satisfying the requirement of adaptively aggregating the urban movement data by considering the spatial proximity and the urban contextual information. Moreover, *HoLens* supports hierarchical organizing, indicating that the analysts can analyze the higher-order pattern at different scales. The visual design of *HoLens* satisfies the requirement 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 the different combinations of the parameters and the alternative designs, and we choose an optimal selection with good performance on intuitive visualization. Finally, two real-world case studies and the feedback from an interview with domain experts demonstrate 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 (Grant No. ZDSYS20210623092007023) in part by National Natural Science Foundation of China (No. 62172398) and Guangdong Basic and Applied Basic Research Foundation (2021A1515011700).

Declarations

There is no conflict of interest/competing Interest in this research.

References

- [1] Tao J, Xu J, Wang C, Chawla NV. HoNVis: Visualizing and exploring higher-order networks. In *2017 IEEE Pacific Visualization Symposium (PacificVis)*, 2017, 1–10.
- [2] Xu J, Wickramarathne TL, Chawla NV. Representing higher-order dependencies in networks. *Science advances*, 2016, 2(5): e1600028.
- [3] Markov JJSK Andrei Andreevich, PST staff T. Theory of algorithms. 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*, 2012, 609–618.
- [5] Shannon CE. A mathematical theory of communication. *The Bell system technical journal*, 1948, 27(3): 379–423.
- [6] Grundy E, Jones MW, Laramee RS, Wilson RP, Shepard EL. Visualisation of sensor data from animal movement. In *Computer Graphics Forum*, volume 28, 2009, 815–822.
- [7] Nekovee M, Moreno Y, Bianconi G, Marsili M. Theory of rumour spreading in complex social networks. *Physica A: Statistical Mechanics and its Applications*, 2007, 374(1): 457–470.
- [8] Kareiva P, Shigesada N. Analyzing insect movement as a correlated random walk. *Oecologia*, 1983, 56(2): 234–238.
- [9] Andrienko N, Andrienko G. State transition graphs for semantic analysis of movement behaviours. *Information Visualization*, 2018, 17(1): 41–65.
- [10] Andrienko N, Andrienko G, Stange H, Liebig T, Hecker D. Visual analytics for understanding spatial situations from episodic movement data. *KI-Künstliche Intelligenz*, 2012, 26: 241–251.
- [11] Gaffney S, Smyth P. Trajectory clustering with mixtures of regression models. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, 1999, 63–72.
- [12] Gaffney SJ, Robertson AW, Smyth P, Camargo SJ, Ghil M. Probabilistic clustering of extratropical cyclones using regression mixture models. *Climate dynamics*, 2007, 29(4): 423–440.
- [13] Lee JG, Han J, Whang KY. Trajectory clustering: a partition-and-group framework. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, 2007, 593–604.
- [14] Andrienko G, Andrienko N. Spatio-temporal aggregation for visual analysis of movements. In *2008 IEEE symposium on visual analytics science and technology*, 2008, 51–58.
- [15] Adrienko N, Adrienko G. Spatial generalization and aggregation of massive movement data. *IEEE Transactions on Visualization and Computer Graphics*, 2010, 17(2): 205–219.
- [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*, 2019, 25(1): 43–53.
- [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*, 2009, 15(6): 969–976, doi:10.1109/TVCG.2009.181.
- [18] Rosvall M, Esquivel AV, Lancichinetti A, West JD, Lambiotte R. Memory in network flows and its effects on spreading dynamics and community detection. *Nature communications*, 2014, 5(1): 1–13.
- [19] Zeng W, Fu CW, Müller Arisona S, Schubiger S, Burkhard R, Ma KL. Visualizing the Relationship between Human Mobility and Points-of-Interest. *IEEE Transactions on Intelligent Transportation Systems*, 2017, 18(8): 2271–2284.
- [20] Dodge S, Weibel R, Lautenschütz AK. Towards a taxonomy of movement patterns. *Information Visualization*, 2008, 7(3): 240–252, doi:10.1057/palgrave.ivs.9500182.
- [21] Slingsby A, van Loon E. Exploratory Visual Analysis for Animal Movement Ecology. *Computer Graphics Forum*, 2016, 35(3): 471–480.
- [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*, 2015, 22(1): 270–279.
- [23] Chen W, Guo F, Wang FY. A Survey of Traffic Data Visualization. *IEEE Transactions on Intelligent Transportation Systems*, 2015, 16(6): 2970 – 2984, doi:10.1109/TITS.2015.2436897.
- [24] Andrienko G, Andrienko N, Dykes J, Fabrikant SI, Wachowicz M. Geovisualization of Dynamics, Movement and Change: Key Issues and Developing Approaches in Visualization Research. *Information Visualization*, 2008, 7: 173–180.
- [25] Kapler T, Wright W. GeoTime Information Visualization. In *IEEE Symposium on Information Visualization*, 2004, 25–32.
- [26] Andrienko N, Andrienko G. Spatial generalization and aggregation of massive movement data. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(2): 205–219.
- [27] Guo D, Zhu X. Origin-Destination Flow Data Smoothing and Mapping. *IEEE Transactions on Visualization and Computer Graphics*, 2014, 20(12): 2043–2052.
- [28] Scheepens R, Willems N, van de Wetering H, Andrienko G, Andrienko N, van Wijk JJ. Composite Density Maps for Multivariate Trajectories. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2518–2527.

- [29] Feng Z, Li H, Zeng W, Yang SH, Qu H. Topology Density Map for Urban Data Visualization and Analysis. *IEEE Transactions on Visualization and Computer Graphics*, 2021, 27(2): 828–838.
- [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*, 2007, 330–339.
- [31] Laube P, Purves RS. An approach to evaluating motion pattern detection techniques in spatio-temporal data. *Comput. Environ. Urban Syst.*, 2006, 30(3): 347–374, doi:10.1016/j.compenvurbys.2005.09.001.
- [32] Zeng W, Fu CW, Arisona SM, Schubiger S, Burkhard R, Ma KL. A visual analytics design for studying rhythm patterns from human daily movement data. *Visual Informatics*, 2017, 1(2): 81–91.
- [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*, 2023, 9(1): 3–39.
- [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*, 2016, 22(1): 160–169, doi:10.1109/TVCG.2015.2467771.
- [35] Zeng W, Shen Q, Jiang Y, Telea A. Route-Aware Edge Bundling for Visualizing Origin-Destination Trails in Urban Traffic. *Comput. Graph. Forum*, 2019, 38(3): 581–593.
- [36] Pretorius A. Visualization of state transition graphs. 2008.
- [37] Lu WL, Wang YS, Lin WC. Chess Evolution Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 2014, 20(5): 702–713, doi:10.1109/TVCG.2014.2299803.
- [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*, 1996, 221–227.
- [39] Krstajic M, Bertini E, Keim DA. CloudLines: Compact Display of Event Episodes in Multiple Time-Series. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2432–2439, doi:10.1109/TVCG.2011.179.
- [40] Schmidt M. The Sankey diagram in energy and material flow management. *Journal of Industrial Ecology*, 2008, 12(1): 82–94, doi:10.1111/j.1530-9290.2008.00004.x.
- [41] van Ham F, van de Wetering H, van Wijk JJ. Interactive visualization of state transition systems. *IEEE Transactions on Visualization and Computer Graphics*, 2002, 8(4): 319–329, doi:10.1109/TVCG.2002.1044518.
- [42] Wongsuphasawat K, Guerra Gómez JA, Plaisant C, Wang TD, Taieb-Maimon M, Shneiderman B. LifeFlow: visualizing an overview of event sequences. In *Proceedings of the SIGCHI conference on human factors in computing systems*, 2011, 1747–1756.
- [43] Monroe M, Lan R, Lee H, Plaisant C, Shneiderman B. Temporal Event Sequence Simplification. *IEEE Transactions on Visualization and Computer Graphics*, 2013, 19(12): 2227–2236, doi:10.1109/TVCG.2013.200.
- [44] Shen Z, Wei J, Sundaresan N, Ma KL. Visual analysis of massive web session data. In *IEEE symposium on large data analysis and visualization (LDAV)*, 2012, 65–72.
- [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*, 2017, 36(3): 527–538, doi:10.1111/cgf.13208.
- [46] Zeng W, Fu CW, Müller Arisona S, Qu H. Visualizing Interchange Patterns in Massive Movement Data. *Computer Graphics Forum*, 2013, 32(3pt3): 271–280, doi:10.1111/cgf.12114.
- [47] Slingsby A, van Loon E. Exploratory visual analysis for animal movement ecology. In *Computer Graphics Forum*, volume 35, 2016, 471–480.
- [48] Ware C, Arsenault R, Plumlee M, Wiley D. Visualizing the underwater behavior of humpback whales. *IEEE Computer Graphics and Applications*, 2006, 26(4): 14–18.
- [49] Zhang J, Guo H, Yuan X. Efficient unsteady flow visualization with high-order access dependencies. In *2016 IEEE Pacific Visualization Symposium (PacificVis)*, 2016, 80–87.
- [50] Lu Y, Steptoe M, Burke S, Wang H, Tsai J, Davulcu H, Montgomery DC, Corman SR, Maciejewski R. Exploring Evolving Media Discourse Through Event Cueing. *IEEE Transactions on Visualization and Computer Graphics*, 2016, 22(1): 220–229, doi:10.1109/TVCG.2015.2467991.
- [51] Riehmann P, Hanfler M, Froehlich B. Interactive sankey diagrams. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, 2005, 233–240.
- [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*, 2012, 18(12): 2659–2668, doi:10.1109/TVCG.2012.2225.
- [53] Gotz D, Stavropoulos H. DecisionFlow: Visual Analytics for High-Dimensional Temporal Event Sequence Data. *IEEE Transactions on Visualization and Computer Graphics*, 2014, 20(12): 1783–1792, doi:10.1109/TVCG.2014.2346682.
- [54] Perer A, Gotz D. Data-driven exploration of care plans for patients. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, 2013, 439–444.
- [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*, 2015, 259–268.
- [56] Perer A, Sun J. Matrixflow: temporal network visual analytics to track symptom evolution during disease progression. In *AMIA annual symposium proceedings*, volume 2012, 2012, 716.
- [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*, 2012, 186–194.



- [58] Beckmann N, Kriegel HP, Schneider R, Seeger B. The R*-tree: An efficient and robust access method for points and rectangles. In *Proceedings of the 1990 ACM SIGMOD international conference on Management of data*, 1990, 322–331.
- [59] Schwarz G. Estimating the dimension of a model. *The annals of statistics*, 1978: 461–464.
- [60] Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 1974, 19(6): 716–723.
- [61] Van der Heyden MJ, Diks CG, Hoekstra BP, DeGoede J. Testing the order of discrete Markov chains using surrogate data. *Physica D: Nonlinear Phenomena*, 1998, 117(1-4): 299–313.
- [62] Kullback S, Leibler RA. On information and sufficiency. *The annals of mathematical statistics*, 1951, 22(1): 79–86.
- [63] Meulemans W, Riche NH, Speckmann B, Alper B, Dwyer T. Kelpfusion: A hybrid set visualization technique. *IEEE Transactions on Visualization and Computer Graphics*, 2013, 19(11): 1846–1858.
- [64] Baydas S, Karakas B. Defining a curve as a Bezier curve. *Journal of Taibah University for Science*, 2019, 13(1): 522–528.
- [65] Zeng W, Fu CW, Müller Arisona S, Erath A, Qu H. Visualizing waypoints-constrained origin-destination patterns for massive transportation data. In *Computer Graphics Forum*, volume 35, 2016, 95–107.
- [66] Yang D, Qu B, Yang J, Cudré-Mauroux P. Revisiting user mobility and social relationships in lbsns: a hypergraph embedding approach. In *The world wide web conference*, 2019, 2147–2157.
- [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*, 2020.
- [68] Cao N, Gotz D, Sun J, Qu H. Dicon: Interactive visual analysis of multidimensional clusters. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2581–2590.
- [69] Yang J, Hubball D, Ward MO, Rundensteiner EA, Ribarsky W. Value and relation display: Interactive visual exploration of large data sets with hundreds of dimensions. *IEEE Transactions on Visualization and Computer Graphics*, 2007, 13(3): 494–507.
- [70] Post FJ, van Walsum T, Post FH, Silver D. Iconic techniques for feature visualization. In *Proceedings Visualization'95*, 1995, 288–295.
- [71] Sun D, Feng Z, Chen Y, Wang Y, Zeng J, Yuan M, Pong TC, Qu H. DFSeer: A visual analytics approach to facilitate model selection for Demand forecasting. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, 1–13.



Zezheng Feng received a B.E. degree from Northeastern University (NEU), China, in 2017, and an M.S. degree (with distinction) from Loughborough University, UK, in 2018. He is currently a Ph.D. candidate in the Department of Computer Science and Engineering (CSE) at the Hong Kong University of Science and Technology (HKUST) and sponsored by a joint Ph.D. program between HKUST and Southern University of Science and Technology (SUSTech). 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 a B.E. degree in computer science and technology from the Southern University of Science and Technology, China, in 2022. He is working toward an M.S. degree from 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 a B.E. degree from the Nanjing University of Posts and Telecommunications, China, in 2019. His research interests are broadly in machine learning, urban computing, explainable AI, data mining, and data visualization.

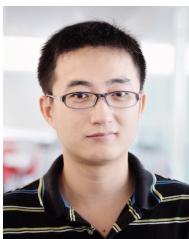


Jianing Hao received a B.E. degree in computer science and technology from Shandong University, China, in 2022. She is currently a Ph.D. student at The Hong Kong University of Science and Technology (Guangzhou). Her research interests include Human-AI collaboration and visual analytics.

Author biography



Shuang-Hua Yang received the B.S. degree in instrument and automation and the M.S. degree in process control from the China University of Petroleum (Huadong), Beijing, China, in 1983 and 1986, respectively, and the 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 Next Generation of Industrial Internet at the Southern University of Science and Technology, China, and also the Head of Department of Computer Science at the University of Reading, UK. His research interests include cyber-physical systems, the Internet of Things, wireless network-based monitoring and control, and safety-critical systems. He is a fellow of IET and InstMC, U.K. He is also an Associate Editor of IET Cyber-Physical Systems: Theory and Applications.



Wei Zeng is an assistant professor in the Hong Kong University of Science and Technology (Guangzhou). He received the PhD degree in computer science from Nanyang Technological University, and worked as a senior research at Future Cities Laboratory, ETH Zurich, and 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 more information, please visit <https://zeng-wei.com/>



Huamin Qu is a professor in the Department of Computer Science and Engineering (CSE) at the Hong Kong University of Science and Technology (HKUST) and also the director of the interdisciplinary program office (IPO) of HKUST. He obtained a BS in Mathematics from Xi'an Jiaotong University, China, an MS and a PhD in Computer Science from the Stony Brook University. His main research interests are in visualization and human-computer interaction, with focuses on urban informatics, social network analysis, E-learning, text visualization, and explainable artificial intelligence (XAI). For more information, please visit <http://huamin.org/>.