REGULAR PAPER

Xiaoying Shi · Fanshun Lv · Dewen Seng · Baixi Xing · Bin Chen

Visual exploration of mobility dynamics based on multi-source mobility datasets and POI information

Abstract Learning more about human mobility is crucial for official decision makers and urban planners. Mobility datasets characterize human daily travel behaviors. Most current researches only studied human dynamics from one kind of mobility dataset. However, people may use different means of transportation to different places for different purposes. Besides, the spatial distributions of different types of point of interests (POIs) reflect the land-use types. How to jointly analyze the multi-source mobility datasets and POI information is a great challenge. In this paper, we adopt multi-source datasets, including taxi dataset, public bicycle system dataset and POI dataset, and propose a visual analytics methodology to explore human mobility dynamics insightfully. Two region–feature–time tensors are constructed first, and a tensor decomposition method is employed to classify the mobility patterns automatically. Then, a new POI– mobility glyph is designed to visualize multi-source datasets in a compact manner. Several interactive visual views are also designed to visualize the spatiotemporal patterns from global, regional and locational perspectives. Case studies based on real-world datasets demonstrate the effectiveness of our method, which supports the visual reasoning of trip purposes and mixed urban functions.

Keywords Multi-source datasets · Visual analysis · Human dynamics · Tensor decomposition

1 Introduction

An understanding of human mobility is crucial for official decision makers and urban planners. With the development of information technology, the large-scale and high-quality individual mobility datasets can be obtained from various sources, such as traffic sensors and public transportation usage records. These datasets contain records reporting people's visited places and the corresponding timestamps over a period of time,

X. Shi · F. Lv · D. Seng (✉) · B. Xing · B. Chen
School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, China
E-mail: sengdw@hdu.edu.cn

X. Shi
E-mail: shixiaoying@hdu.edu.cn

F. Lv
E-mail: 172050041@hdu.edu.cn

B. Xing
E-mail: xingbx@hdu.edu.cn

B. Chen
E-mail: chenbin@hdu.edu.cn

X. Shi · F. Lv · D. Seng · B. Xing · B. Chen
Key Laboratory of Complex Systems Modeling and Simulation, Ministry of Education, Hangzhou, China

reflecting human daily travel behaviors. Many visual analysis methods have been proposed to process various mobility datasets, including taxi data (Ferreira et al. 2013; Zhou et al. 2018), public bicycle system (PBS) data (Shi et al. 2018; Yan et al. 2018) and bus data (Pei et al. 2018).

In addition to mobility datasets, the land-use types of regions also play an important role in the analysis of human movements. People usually go to different places for different kinds of activities, such as studying, working and shopping. The POI information for a place reflects the latent region functions (Gao et al. 2017; Yao et al. 2017). Location-based social networks such as Foursquare offer a web service providing detailed geo-located information of different POIs. The POI categories are called venue categories in Foursquare, i.e., food, residence, nightlife spot, shop and service. Therefore, the check-in records collected from Foursquare can be used to construct a POI dataset, containing the venue categories of various locations, which help to understand the trip purposes.

Many previous methods focused on exploring patterns from one kind of mobility dataset (Ferreira et al. 2013; Pei et al. 2018; Shi et al. 2018; Yan et al. 2018; Zhou et al. 2018), which were unable to investigate the usage variances for different travel modes. People usually use different means of transportation to different places for different purposes. Information associated with multi-source datasets provides unprecedented insight into human behavior and mobility patterns. How to utilize multi-source datasets to perceive real human travel conditions and discover the important underlying patterns in different datasets poses a substantial challenge: (1) Different mobility datasets have different forms. How to extract unified features from the large-scale datasets and characterize the underlying patterns? (2) The mobility datasets and POI dataset are heterogeneous. How to analyze and visualize the relationship between human mobility and POI distribution in a compact form?

In this paper, we propose a visual analytics methodology that addresses the above problems. Our method uses three real-world datasets: taxi dataset, public bicycle system dataset and POI dataset. Tensors are constructed to depict the mobility datasets in a unified form, and a tensor decomposition method is adopted to classify the underlying spatiotemporal patterns automatically. A visual analytics system with multiple coordinated views is designed for comprehensive exploration of mobility dynamics from different perspectives. The contributions of this research are as follows:

1. We model the mobility datasets as the tensors and employ tensor decomposition to classify the underlying movement patterns automatically.
2. We propose a novel visual analytics system to support the exploration of human dynamics based on multi-source datasets. We also design a POI–mobility glyph to demonstrate the heterogeneous data in a compact manner.
3. We collect three real-world datasets and conduct case studies to explore the human trip purposes and mixed urban functions.

The remainder of this paper is organized as follows. Section 2 introduces related work. Section 3 describes the datasets and the system pipeline. Section 4 explains our method in detail. Section 5 demonstrates the feasibility of the proposed method via case studies and expert interview. We conclude our paper with an outline of future work in Sect. 6.

2 Literature review

Various kinds of mobility datasets were adopted to analyze human activity patterns. For taxi dataset, Ferreira et al. (2013) proposed a visually query model to retrieve taxi trips. Zhou et al. (2018) employed a matrix factorization method to classify urban functional areas and visualized the mobility patterns of urban areas with different functions. For PBS dataset, Shi et al. (2018) proposed an interactive visual analytics system to explore the spatial–temporal changes, the relationships between flow pairs and the influence of multi-dimensional factors. Yan et al. (2018) employed tensor factorization to extract user activity patterns and visualized the patterns from different dimensions. Pei et al. (2018) presented a visual analysis system to analyze the bus station congestion patterns and the importance of bus stations based on bus dataset.

The above methods all focused on analyzing a single-source mobility data. In order to explore the relationship between human movement and POI information, Zeng et al. (2017) introduced a compact visual representation 'POI–mobility signature' to effectively visualize such relationship. They only analyzed one kind of mobility dataset with POI information. Shi et al. (2019) adopted taxi dataset and PBS dataset to explore the evolutionary patterns of urban activity areas, which used the two datasets separately for analysis.

To analyze the heterogeneous urban data, Chen et al. (2018a) designed a visual analysis framework to support interactive visual queries and exploration over cross-domain data. Their work focused on the design of a visual query model. They also proposed a progressive detect-and-filter scheme for visual reasoning with egocentric connections in terms of time, space and social networks (Chen et al. 2018b). This work focused on studying the egocentric relationships of an individual. Xie et al. (2018) proposed a visual analytics approach to classify the heterogeneous graph-based datasets using a hypergraph model. Their method focused on the graph-based datasets. Different from the above methods, we aim to explore and visualize human dynamics based on heterogeneous urban datasets.

Tensors can be considered as a generalization of matrices and are useful for representing third- or higher-order data. Tensor decomposition is able to extract latent features in dataset, as well as reduce data dimensionality, which has been used in anomaly analysis (Cao et al. 2018), missing traffic data completion (Ran et al. 2016) and traffic prediction (Han and Moutarde 2016). Liu et al. (2019) proposed a tensor-based algorithm to support the automated partitioning and multi-dimensional pattern extraction on spatiotemporal data. In this paper, we adopt tensor decomposition to extract the underlying human mobility patterns.

3 Data description and system overview

3.1 Data description

We adopt three real-world datasets collected from New York City (NYC). The PBS dataset[1] and taxi dataset[2] are collected in September 2014, including 1,509,444 and 10,469,897 trips, respectively. As the location-based social media services become increasingly popular, more and more people use such services to record information about POIs. In our paper, we collected the check-in data from Foursquare to construct the POI dataset,[3] which contains 227,428 check-ins from April 2012 to February 2013 (Yang et al. 2015).

PBS dataset The public bicycle system provides last mile connectivity to other public transport (Fishman 2016). Users can rent bikes from nearby stations and return them to other stations in the city. PBS dataset includes two kinds of records: station record and trip record. Station records contain attributes of stations, such as station name, longitude and latitude. Each trip record represents a flow from an originating station to a destination station, which is defined as:

$$\text{TR}_{\text{PBS}} = (leaseStation, leaseTime, returnStation, returnTime) \tag{1}$$

Taxi dataset Taxicabs come in two varieties in NYC: yellow and green. In our paper, we adopt both yellow and green taxi trip records. Each trip record captures the pick-up/drop-off timestamps and corresponding GPS coordinates, which is defined as:

$$\text{TR}_{\text{TAXI}} = (pLong, pLat, pTime, dLong, dLat, dTime) \tag{2}$$

where *pLong*/*pLat* refer to the longitude/latitude coordinates of the pick-up place and *dLong*/*dLat* refer to the longitude/latitude coordinates of the drop-off place. *pTime* and *dTime* refer to the pick-up and drop-off timestamps of one trip.

POI dataset The POI dataset is collected based on the check-in records from Foursquare. Each check-in record is associated with a venue category, location name, GPS coordinates (latitude and longitude) and the timestamp. This dataset contains ten different venue categories, which provide insights for people visiting purposes of locations implicitly. Table 1 shows the statistics of the total check-in counts for all categories.

3.2 Task analysis

Aiming at a comprehensive and efficient analysis of human mobility dynamics based on multi-source datasets, we discussed with the domain experts and characterized a list of analytical tasks as below.

R.1 Automated classification of the mobility patterns based on different datasets Different mobility datasets have different data forms. We need to construct a unified data representation and classify the latent mobility patterns automatically from each mobility dataset.

Table 1 The statistics of check-ins from Foursquare records

Venue category	Check-ins	Venue category	Check-ins
Food	23,416	Professional and other places	12,350
Shop and service	10,038	Nightlife spot	9830
Outdoors and recreation	9633	Travel and transport	8501
Arts and entertainment	4963	Residence	1789
College and university	1783	Event	0

R.2 Correlation of the multiple mobility datasets and POI information The mobility datasets and POI dataset are heterogeneous. We need to design a meaningful glyph, which not only supports the jointly analysis of classified mobility patterns based on different travel modes, but also supports visualizing mobility datasets and POI information simultaneously.

R.3 Progressive and interactive visual exploration of multi-source datasets The designed visual analytics system should support a top-down, divide-and-conquer analytics workflow and provide a set of convenient user interactions, to help analysts gain deeper insights of mobility patterns from different perspectives.

3.3 System overview

To achieve the above tasks, a visual analysis system is designed. Figure 1 shows the system pipeline. Multi-source datasets are collected first. Then, three phases are performed: data modeling, tensor decomposition and visual exploration.

During the data modeling phase, the city is first divided into structured blocks in terms of road networks via map segmentation method. Then, the time ranges are divided into time intervals. At last, the taxi, PBS and POI data are organized according to the temporal partition and map segmentation results.

During the tensor decomposition phase, two region–feature–time tensors are first built up for PBS and taxi data, respectively. Then, a nonnegative tensor decomposition method is applied to identify the latent mobility patterns automatically (R.1).

During the visual exploration phase, interactive visual views are designed for progressive pattern recognition from different perspectives (R.3). The global view is given first, showing the overall spatiotemporal patterns based on tensor decomposition. When selecting a specific region, the region-related information is presented. A well-designed POI–mobility glyph shows the POI proportion and mobility temporal variation in a compact manner (R.2). The flow map chart demonstrates the detailed trips meeting several pre-defined constraints. When selecting a location on the flow map chart, the related POI information for a specific location is further shown.

4 Methods

4.1 Data modeling

Map segmentation Regions are used as the basic units to study human dynamics. The regular map segmentation methods divided the city into regular grids (Tang et al. 2018) or hexagons (Zhou et al. 2018), which ignored the road condition of the city. Here, we adopt an irregular segmentation method (Yuan et al. 2015) to divide the city into N disjoint regions in terms of road networks. The method adopts the raster-based model to represent the road network and utilizes morphological image processing technique to segment the map. In order to support multi-source data analysis, we only choose the regions having both taxi and cycling trips and obtain 96 regions finally. These regions carry more semantic meanings than uniform grids or hexagons.

Temporal partition Since the user behavior is usually stochastic when the time interval is small, we split the entire time period into time slices (e.g., hours). Without loss of generality, we call an hour as a time interval.

Data preprocessing To process the mobility datasets uniformly, we abstract the original trip records TR_{PBS} and TR_{TAXI} into a similar form: $TR_{Abs} = (D_{Loc}, D_t, A_{Loc}, A_t)$. D_{Loc}/A_{Loc} are the departure/arrival locations, which refer to *leaseStation/returnStation* for PBS data and pick-up/drop-off places for taxi data,
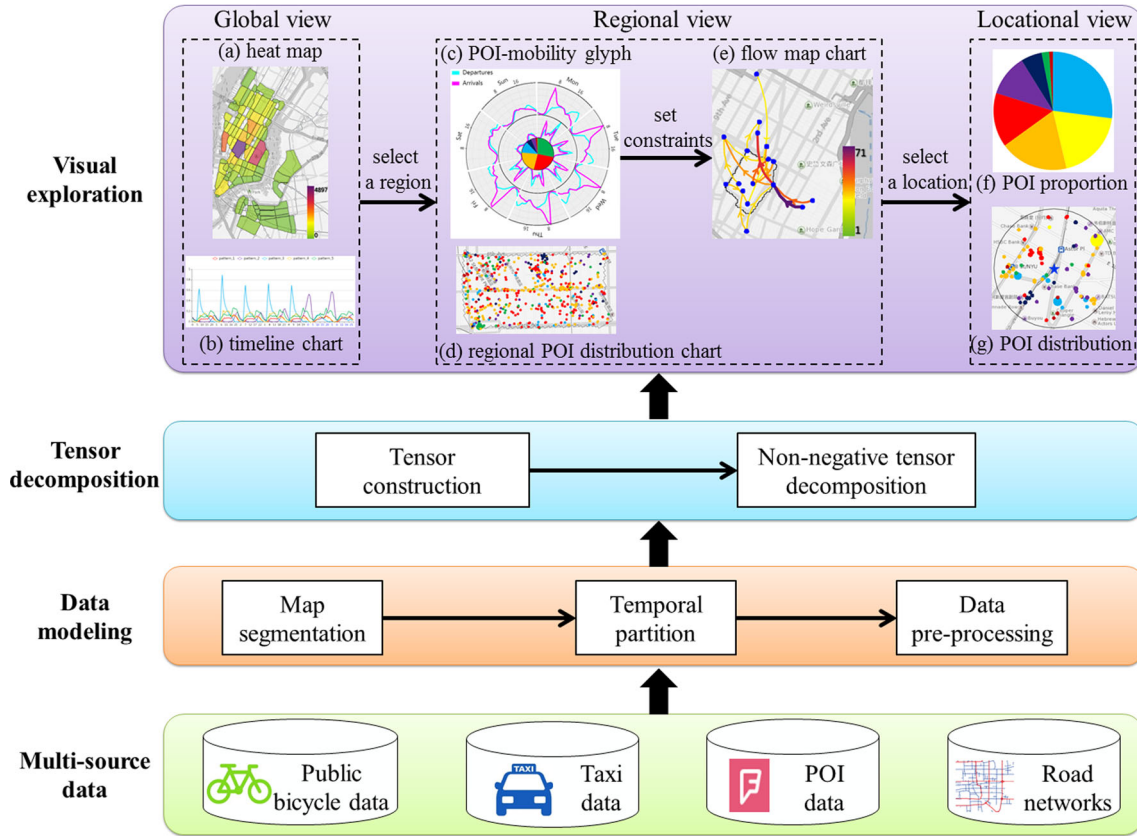
Fig. 1 The system pipeline

respectively. D_t/A_t refer to the departure/arrival timestamps. After that, we aggregate trips according to the segmented regions on an hourly basis and obtain $\mathrm{TR_{Sum}}$:

$$\mathrm{TR_{Sum}} = (D_{\mathrm{Date}}, D_{\mathrm{Hour}}, D_{\mathrm{LOC}}, A_{\mathrm{LOC}}, \mathrm{num}) \tag{3}$$

For instance, $\mathrm{TR_{Sum}}$ = ('2014-09-02.' '8,' '14,' '89,' 3) means three people start from location '14' and go to location '89' from 8 to 9 a.m. on September 2, 2014.

For POI data, we find that the category 'food' has the largest total check-in counts (Table 1), and the corresponding check-in locations spread over the city. Since the locations related to 'food' may exist in the recreational area, working area or residential area, this category has no significant effect on the judgment of regional functions. Therefore, we remove this category. Besides, the check-in counts of category 'event' is zero in Table 1; thus, we obtain eight categories. Finally, we map each check-in location to a specific region according to its longitude and latitude and count the total number of check-ins for each POI category in each region.

4.2 Tensor construction and decomposition

4.2.1 Tensor construction

In order to analyze the spatial, temporal and mobility pattern in a unified manner, a region–feature–time tensor is constructed to model the multi-dimensional relationships. A tensor $\mathcal{X} \in \mathcal{R}^{N \times F \times T}$ with three dimensions denotes N regions, F features and T time intervals, respectively.

Region dimension The first dimension denotes N regions obtained after the map segmentation.

Temporal dimension To explore the periodical human activity regularity within a regularly week cycle, the temporal dimension T corresponds to 168 h in 1 week. We project all records of $\mathrm{TR_{Sum}}$ into 1 week. For example, all records of $\mathrm{TR_{Sum}}$ whose D_{Date} satisfies Monday and D_{Hour} = '8' are aggregated.

Feature dimension This dimension characterizes a region's incoming and outgoing flow numbers. Since we have N regions, the feature dimension $F = 2N$. The first N features capture the numbers of trips leaving from a specific region to every other region, and the next N features indicate the numbers of trips entering that region from every other region.

An entry An entry $\mathcal{X}(i, j, k)$ denotes the value of jth feature in ith region during the kth time interval. To be more specific, for region i, $\mathcal{X}(i, j, k)$ denotes the number of trips from region i to region j in kth hour, and $\mathcal{X}(i, N + j, k)$ denotes the number of trips from region j to region i in kth hour.

Based on the above definition, we build up two tensors, one is for taxi data, and the other is for PBS data.

4.2.2 Nonnegative tensor decomposition

After tensor construction, we apply nonnegative tensor decomposition on each tensor to extract the latent mobility patterns. Figure 2 illustrates the tensor decomposition process. Nonnegative tensor decomposition seeks to approximate a given tensor $\mathcal{X}$ as the sum of R rank-one tensors with nonnegative components. The PARAFAC-like decomposition (Kolda and Bader 2009) is adopted:

$$\hat{\mathcal{X}} = \sum_{r=1}^{R} a_r \circ b_r \circ c_r = [[A, B, C]] \quad \text{s.t.} \quad a_r, b_r, c_r \geq 0 \tag{4}$$

$$\min \left\| \mathcal{X} - \hat{\mathcal{X}} \right\|_2^2 \tag{5}$$

where $a_r \in \mathcal{R}^N$, $b_r \in \mathcal{R}^F$ and $c_r \in \mathcal{R}^T$ for $r = 1, 2, \ldots, R$. R is the decomposition rank representing the number of desired latent patterns, which can be assigned by analyst. $\|\cdot\|_2^2$ denotes the Frobenius norm of a tensor. PARAFAC decomposition can be represented in its matrix form $[\![A, B, C]\!]$, where the columns of matrices A, B and C are a_r, b_r and c_r. In particular, the shapes of A, B and C are $N \times R$, $F \times R$ and $T \times R$, respectively. As the symbol '$\circ$' represents outer product of vectors, each element of the tensor $\mathcal{X}$ can be written as:

$$\mathcal{X}(i, j, k) \approx \sum_{r=1}^{R} A_{ir} B_{jr} C_{kr} \tag{6}$$

where the entry A_{ir} indicates the *spatial importance* of region i in the rth pattern and the entry C_{kr} represents the *temporal importance* of kth time interval in the rth pattern. The entry B_{jr} indicates the strength of jth feature in the rth pattern. Therefore, the outer product of the vectors $a_r \circ b_r \circ c_r$ captures the structural relations among region, mobility feature and time.

After tensor decomposition, the dimensionality of the data is reduced from $N \times F \times T$ to $R \times (N + F + T)$. The tensor decomposition results generate an ensemble of interpretable spatiotemporal patterns, which can be further visualized for pattern discovery.

4.3 Visual designs

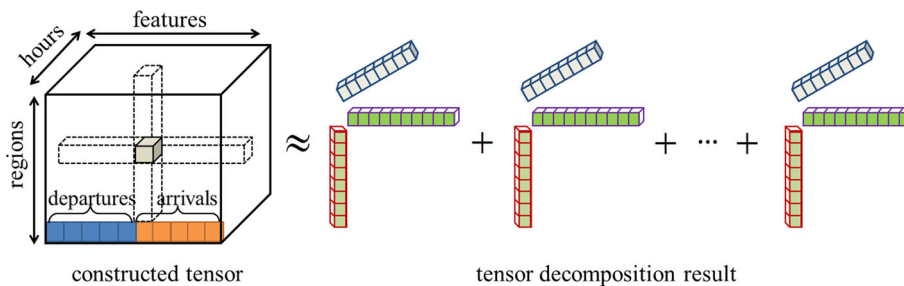To facilitate the analyzing and understanding of the latent patterns, multiple coordinate visual views are designed.

Fig. 2 The illustration of tensor decomposition process

4.3.1 Global view

The global view presents a general overview of temporal pattern and spatial distribution for mobility datasets based on tensor decomposition results, including a timeline chart and R heat maps.

The timeline chart (Fig. 1b) shows the temporal variation tendency of each pattern derived from the factor matrix C. Each column of the matrix C represents one pattern. For the rth pattern, we use a polygonal line to show the *temporal importance* based on tensor decomposition values $C_{kr}(k = 1, \ldots, 168)$. Therefore, the x-axis corresponds to 168 h in 1 week, and the y-axis is the value of C_{kr}. The line number is consistent with the decomposition rank R. Because the lines may intertwine with each other, analysts can further uncheck several lines, to take a closer look at the remaining patterns.

The heat map (Fig. 1a) shows *spatial importance* in the geographic environment derived from the factor matrix A. Each column of the matrix A is corresponding to one heat map. In a specific heat map, each region is rendered with a color related to the decomposition value A_{ir}, reflecting the importance of region i for the rth pattern. A gradient color scheme (purple–red–yellow–green) is used to encode the importance. We use this color scheme in other views for consistency. A purple region indicates a higher level of regional importance in the current pattern.

By looking at the global view, analysts can not only find the important regions for both travel modes, but also compare the differences in the time domain.

4.3.2 Regional view

When selecting a region on the heat map, the regional view depicts detailed information for the selected region, including three components: POI–mobility glyph, flow map chart and regional POI distribution chart.

POI–mobility glyph Inspired by the work of Zeng et al. (2017), we design a POI–mobility glyph to show the POI distribution and multi-source human mobility in a unified manner (Fig. 5a). The POI proportion for a region is calculated first:

$$p_j = \frac{\text{checkins}_j}{\sum_{i=1}^{cn} \text{checkins}_i} \tag{7}$$

where p_j denotes the density of the jth POI category. checkins_j denotes the check-in amount of the jth category in the current region. cn is the category number, and $cn = 8$ in our case. All p_j are sorted in descending order. After that, all sorted p_j are drawn in the center pie chart starting from the y-axis in clockwise order. The sector size is proportional to the category's density. We use a unique color to represent one POI category. The POI legend is shown below.

The temporal variations of mobility are arranged in a radial layout through two circular rings. The outer ring represents the taxi mobility, while the inner ring represents the PBS mobility. The rings are divided into seven sectors, in agreement with 7 days of 1 week. Each ring has two curves, representing the aggregated flow numbers of departure and arrival in 1 week for this region. When hovering on the curves, the detailed information will be displayed.

Flow map chart By looking at the POI–mobility glyph, analysts can compare the peak hours of both travel modes easily. They can further set several constraints and investigate the detailed OD trips in the region through flow map chart (Fig. 1e). The flow map chart supports multiple constraint settings: travel mode constraint (PBS or taxi), temporal constraint and magnitude constraint. Analysts can set various temporal constraints; for example, select the time range, choose the date type (workday or weekend), and assign the scope of hours. For instance, they can observe the trips on workday morning from 8 to 9 a.m., or they can view the trips on Friday and Saturday evening from 23 p.m. to 2 a.m. of the next day. Besides, directly showing all trips will cause visual clutter. Magnitude constraint is supported for filter out unimportant trips. Analysts can input a threshold, the flows whose magnitudes less than that threshold are ignored.

Based on the pre-defined constraints, the flow map chart draws all satisfied trips starting or ending at the current region. The blue dot on the map stands for a location. For PBS data, it corresponds to a station in the region. However, taxi trips have no fixed pick-up and drop-off places. They are very loosely distributed in the region. We further divide the region into 400 m * 400 m grids and aggregate taxi trips according to the grids. Therefore, the blue dot represents the grid center for taxi data. Arcs with arrows represent the flows between locations. The flow number is double-encoded by arc thickness and color. We employ the same

gradient color mapping approach as in the heat map. A thicker and darker purple arc indicates a larger number of trips.

Regional POI distribution chart The regional POI distribution chart illustrates all check-in places in the region (Fig. 1d). A dot on the map represents one check-in place. Its color is consistent with the POI legend, and the dot size is proportional to the check-in amount in this place.

4.3.3 Locational view

When clicking the blue dot on the flow map chart, the related locational view is updated. The locational view mainly demonstrates the POI information around a bicycle station or in a taxi grid. It includes two charts: *POI proportion chart* (Fig. 1f) and *POI distribution chart* (Fig. 1g).

For PBS data, a specific station is selected when clicking a blue dot. The related POI information contains all POIs falling into a surrounding circle centered on the selected station. Analysts can adjust the circle's radius. The POI proportion chart shows the density of the jth POI category in the circle. Meanwhile, the POI distribution chart draws all POIs in the circle, and the selected station is represented by a five-pointed star. One check-in place is represented by a dot, whose color is consistent with the POI legend, and the dot size is proportional to the check-in amount in this place. When clicking the dot, the name and check-in counts of this POI will be shown.

For taxi data, the blue dot is no longer stands for a specific location. Instead, it represents the grid center and contains multiple pick-up/drop-off places. Because the drop-off place is always close to the destination for taxi trips, the POI proportion chart shows the density of the jth POI category in the grid, and the POI distribution chart draws all POIs in the grid directly.

5 Result analysis

In this section, we present two case studies to show the effectiveness of our method and discuss how to select the decomposition rank R.

5.1 Case study 1: The overview analysis of spatiotemporal patterns for both travel modes

The first case study compares the overall spatiotemporal patterns for both datasets. Figure 3 illustrates the global view for taxi data ($R = 5$). As seen from the timeline chart (Fig. 3a), we can see the temporal variations of five patterns. Pattern 3 has obvious morning peaks on workdays. Pattern 2 has late evening
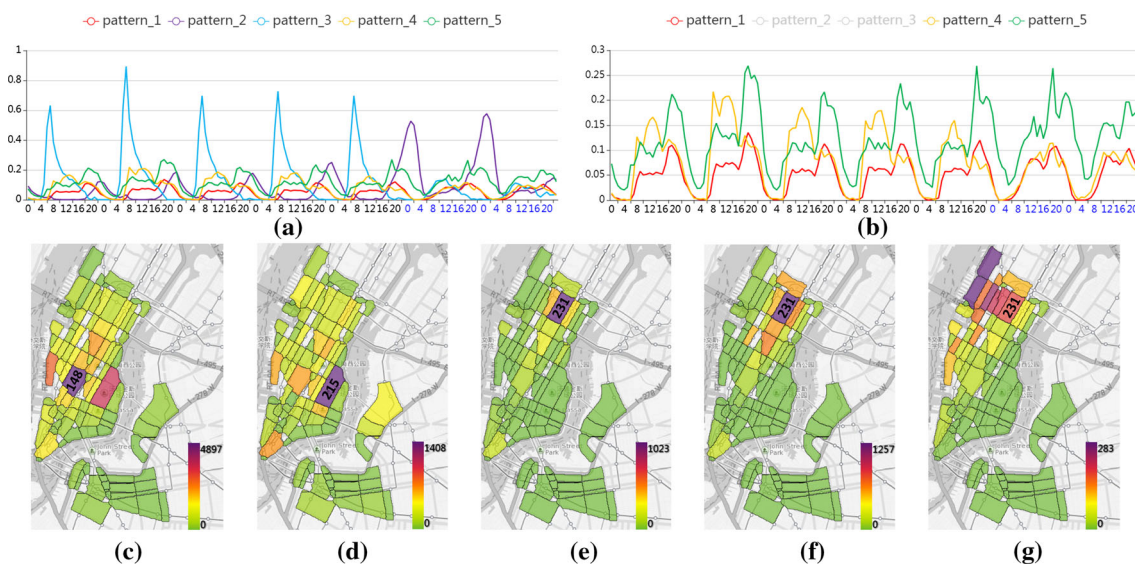
Fig. 3 The global view for taxi data when $R = 5$. **a** The timeline chart; **b** the timeline chart with refined three patterns; **c** the heat map for pattern 1; **d** the heat map for pattern 2; **e** the heat map for pattern 3; **f** the heat map for pattern 4; and **g** the heat map for pattern 5
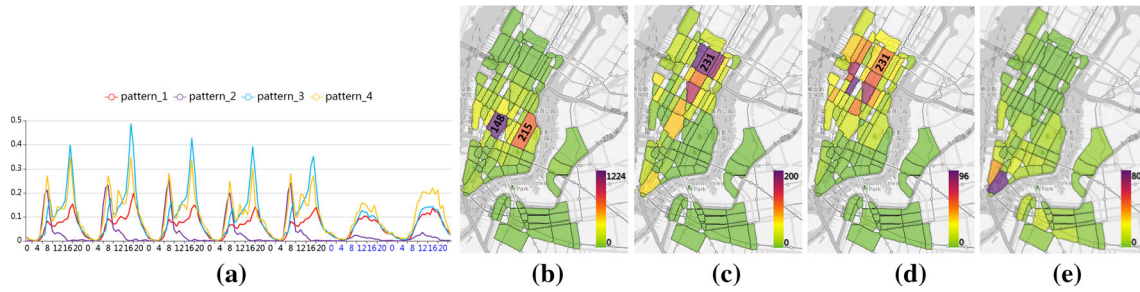
Fig. 4 The global view for PBS data when $R = 4$. **a** The timeline chart; **b** the heat map for pattern 1; **c** the heat map for pattern 2; **d** the heat map for pattern 3; and **e** the heat map for pattern 4

peaks during the whole week, and the peaks on Friday and Saturday evenings are particularly significant. In order to take a closer look at the remaining three patterns, we uncheck pattern 2 and pattern 3 and obtain Fig. 3b. Pattern 4 has higher flow numbers during 8 a.m. to 3 p.m. on workdays and during 11 a.m. to 10 p.m. on weekends. Pattern 1 and pattern 5 have evening peaks around 7 p.m., and the flow magnitude of pattern 5 is bigger than that of pattern 1. Figure 3c–g illustrates the corresponding heat maps for five patterns, from which we can easily find the most important regions for each pattern.

Figure 4 shows the global view for PBS data ($R = 4$). Pattern 4 has morning and evening peaks on workdays. Pattern 2 has distinct morning peaks, while pattern 1 and pattern 3 have obvious evening peaks with different flow magnitudes. Figure 4b–e illustrates the corresponding heat maps for four patterns.

5.2 Case study 2: The analysis of a specific region

The global view demonstrates the summarized spatiotemporal patterns for both datasets, from which we can easily find that the same region exhibits different patterns with different kinds of importance in different datasets. The goal of the second case study is to investigate the relationships among different travel modes and POI information for regions. In the following, we select several representative regions for explanation.

Region 231 We first investigate region 231, a very important region for both travel modes in the heat maps (Figs. 3e–g, 4c, d). As seen from the POI–mobility glyph (Fig. 5a), we find that the top two POI categories are 'Travel and Transport' and 'Professional and Other Places.' The peak hours for both travel modes appear on workdays. For PBS data, the morning arrival peaks appear around 7–8 a.m., and the evening departure peaks appear around 4–5 p.m. For taxi data, the morning arrival peaks also appear around 7–8 a.m. The evening departure peaks appear around 8–10 p.m., whose peak values are less than taxi morning peaks and 4 h later than PBS evening peaks.
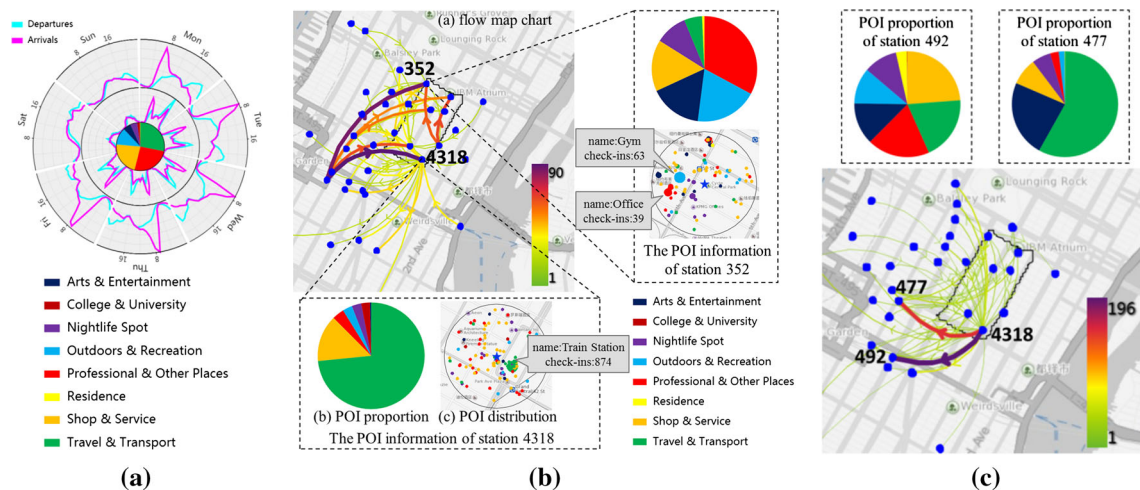
Fig. 5 The analysis of region 231 on workdays by bicycle. **a** The POI–mobility glyph; **b** the flow map chart and corresponding POI information for morning peaks (7–8 a.m.); and **c** the flow map chart and corresponding POI information for evening peaks (4–5 p.m.)
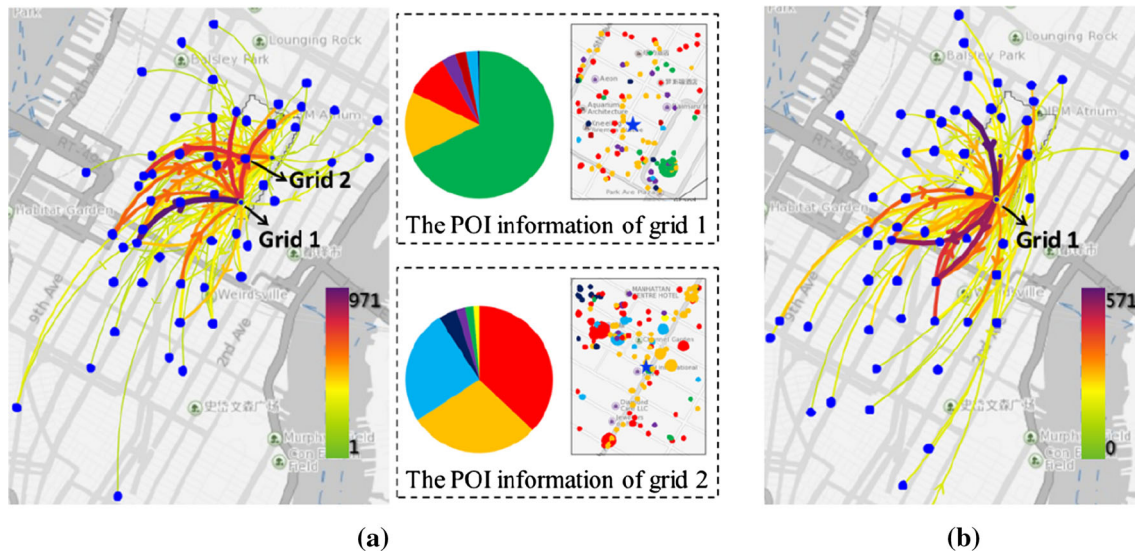
Fig. 6 The analysis of region 231 on workdays by taxi. **a** The flow map chart and corresponding POI information for morning peaks (7–8 a.m.); and **b** the flow map chart for evening peaks (8–10 p.m.)

We further observe the flow map chart of PBS data and filter out unimportant flows. During morning rush hours (Fig. 5b), the popular destinations are station 4318 and 352. As seen from corresponding POI information, there is a train station (Grand Central Terminal) with largest check-ins near station 4318, which is the busiest transportation hub in NYC. The popular check-in places are gyms and offices near station 352. During evening rush hours (Fig. 5c), the popular original station is 4318 near the train station, and the popular destination stations are 492 and 477, which are near bus stations and shops.

The taxi trips with larger usage numbers are more complex. Figure 6a, b illustrates the flow conditions of morning and evening peaks by taxi, respectively. The flows spread over the region. During the morning peaks, two grids are very popular. Grid 1 contains the train station 'Grand Central Terminal,' and Grid 2 contains many offices and shops. During the evening peaks, large flows all concentrate on the train station.

In general, the morning arrival peaks and evening departure peaks are very obvious for both travel modes in this region. The morning travel destinations are near the train station and workplaces, and the evening trips are mostly starting from the train station. As seen from the map in Fig. 6, the taxi trips have longer distance. From the above observation, we consider that this region mainly serves as work places and transportation hub. The travel purposes by bicycle and taxi are similar, both for commuting. People may choose different travel modes for the same purpose, and they prefer to choose taxi when the distance is longer.

Region 215 Then, we observe region 215, which has obvious evening peaks by bicycle (Fig. 4b) and late evening peaks by taxi (Fig. 3d). As seen from the POI–mobility glyph (Fig. 7), this region has the largest POI proportion of 'Nightlife Spot.' For cycling trips, it has obvious departure peaks (7–8 a.m.) and arrival peaks (5–6 p.m.) on workdays. For taxi trips, the morning peaks are nonsignificant. From Monday to Thursday, the evening arrival peaks start from 8 p.m., and the evening departure peaks appear 3 h later. On Friday and Saturday evenings, the arrival peaks start from 11 p.m., and the departure peaks also appear 3 h later. The peak values are obviously larger than those on workdays.

Next, we disclose the regional cycling condition. As seen from Fig. 8, the main trips have opposite directions during morning peaks and evening peaks. During morning rush hours (Fig. 8a), the popular origins are stations 445, 432 and 317. The largest POI proportion of these stations is the category 'Nightlife Spot.' However, the mobility pattern indicates that these trips may be related to residential areas, since people leave in the morning and arrive in the late afternoon. We find that the check-in amount of category 'Residence' is relatively few in Table 1, and thus, we infer that people are unwilling to check in near home for the reason of privacy. Considering the mobility pattern, we speculate that this region contains places of residence, and the bicycle trips are related to daily commuting. Besides, during evening peaks, station 504 near shops attracts a lot of traffic (Fig. 8b). People may go there for shopping after work.
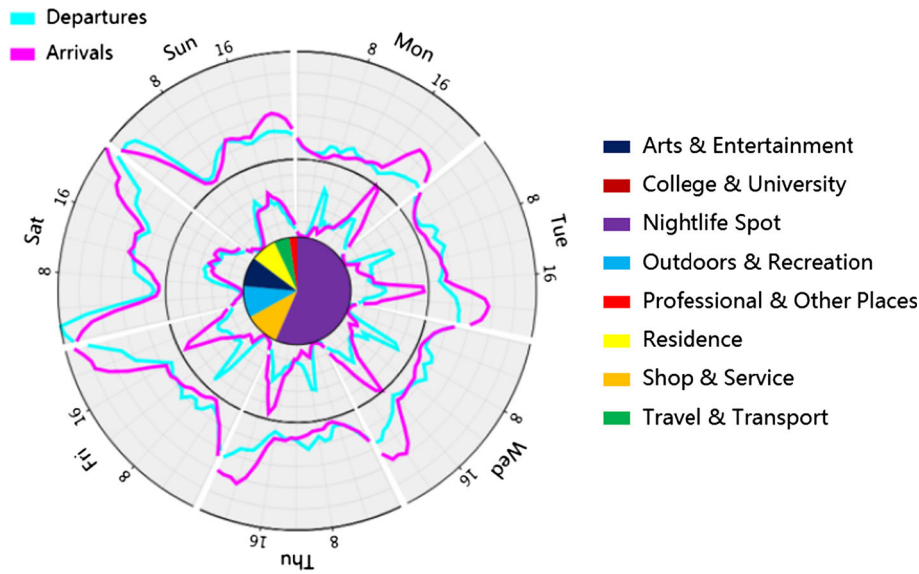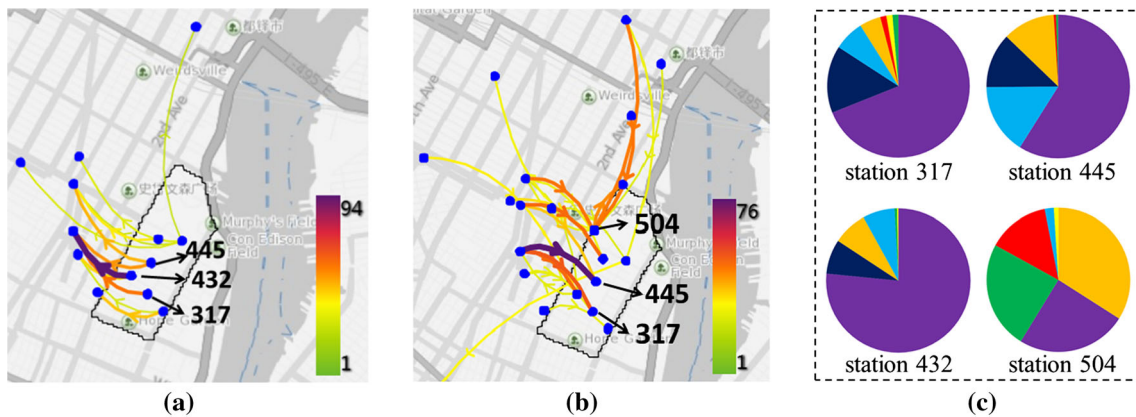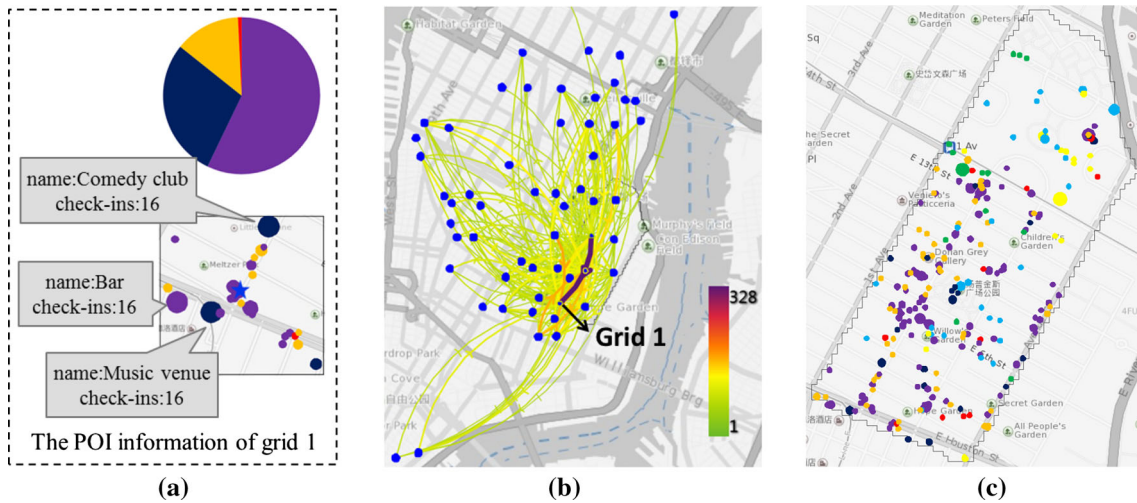
Fig. 7 The POI–mobility glyph of region 215

Fig. 8 The analysis of region 215 on workdays by bicycle. **a** The flow map chart for morning peaks (7–8 a.m.); **b** the flow map chart for evening peaks (5–6 p.m.); and **c** the POI proportion charts for specific stations

After that, we look at the taxi trips on Friday and Saturday evenings. Since the evening peaks last from 11 p.m. to 2 a.m. of the next day, we set the temporal constraint. As seen from the flow map chart (Fig. 9b), the flows spread over the region and come from different places in the city. Since this region contains many nightlife places (Fig. 9c), people go to this region to enjoy life in the night. The most popular destination (Grid 1) contains many bars and clubs (Fig. 9a).

In general, we conclude that people may choose different transportation means for different travel purposes. They tend to ride bikes for commuting and enjoy nightlife by taxi. The analysis results also indicate that this region has mixed functions revealed by different mobility datasets, e.g., residential areas and nightlife areas.

Region 148 Lastly, we explore region 148, which exhibits similar mobility patterns for different traffic modes (Figs. 3c, 4b). The POI–mobility glyph (Fig. 10) shows that the departure amounts are equal to the arrival amounts for both modes. For cycling trips, there are morning peaks (8–9 a.m.) and evening peaks (5–6 p.m.) on workdays. For taxi trips, the morning peaks appear around 9 a.m. on workdays. The evening peaks appear from 7 to 10 p.m. from Monday to Thursday. On Friday and Saturday evenings, the arriving peaks appear from 8 p.m., and the departure peaks appear later.

We further observe the flow map chart for PBS data. During morning peaks on workdays (Fig. 11a), some people ride bikes to this region, and others leave this region. The most popular station is 293 with

Fig. 9 The analysis of region 215 on Friday and Saturday evenings by taxi. **a** The locational view for grid 1; **b** the flow map chart; and **c** the regional POI distribution chart
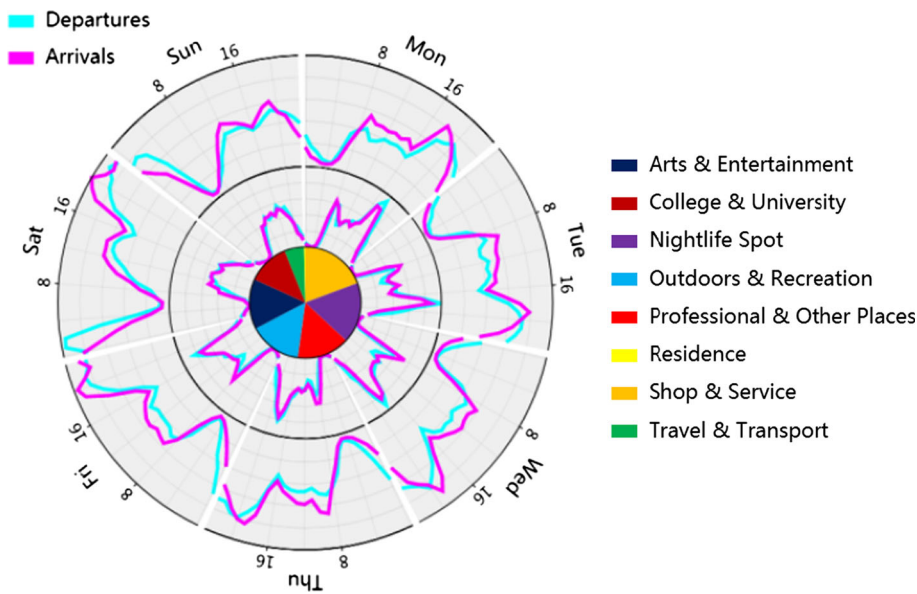

**Fig. 10** The POI–mobility glyph of region 148

various kinds of POIs nearby, like 'Outdoors and recreation' and 'Shop and Service' relating to work, as well as 'Residence.' Therefore, this station contains obvious departure trips and arrival trips simultaneously. During the evening peaks (Fig. 11b), some trips are opposite with those in the morning, and the most popular original stations also contain various kinds of POI categories.

For taxi trips, Fig. 12a illustrates the flow condition during workday morning. The most popular grids contain many work places, such as shops, universities and offices. Figure 12b depicts the flow condition on Friday and Saturday evenings. The most popular grids contain many nightlife spots.

In general, we consider that this region has multiple functions, including job-housing places and nightlife areas. When a region contains both residential areas and work places, the departures and arrivals tend to be balanced. These regions can also be regarded as job-housing balance regions.
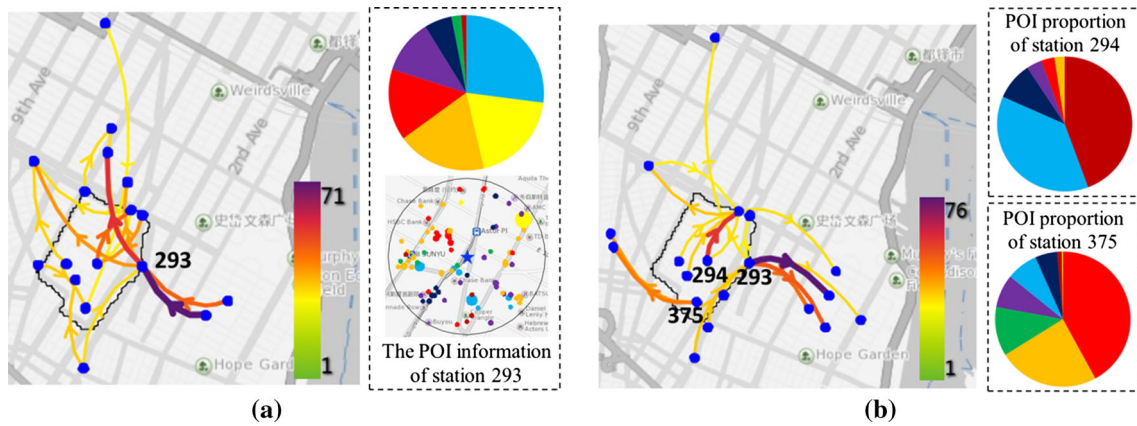
Fig. 11 The analysis of region 148 on workdays by bicycle. **a** The flow map chart and corresponding POI information for morning peaks (8–9 a.m.); and **b** the flow map chart and corresponding POI proportion charts for evening peaks (5–6 p.m.)
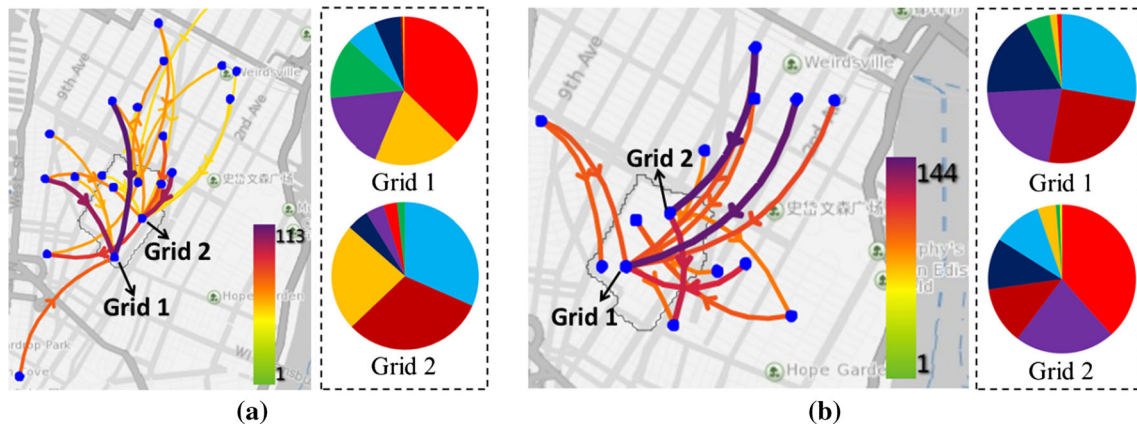

Fig. 12 The analysis of region 148 by taxi. **a** The flow map chart and corresponding POI proportion charts for morning peaks on workdays (9 a.m.); and **b** the flow map chart and corresponding POI proportion charts for Friday and Saturday evening peaks (8–11 p.m.)

5.3 Expert interview

We invited two experts from the traffic management department and conducted one-on-one interviews with them. We collected their feedback and summarized as follows:

Visual designs and interactions According to the experts' feedback, most of the visual components meet our design goals successfully. For instance, the experts felt that the timeline chart, the heat map, the flow map chart and the POI-related charts were very easy to comprehend. The experts also appreciated the design of POI–mobility glyph and commented that 'The POI-mobility glyph is informative and easy to understand. It indeed helps us to identify the usage peak hours for different transportation means in each region, as well as know the POI distribution of each region.' One expert also mentioned that 'The visual analysis system is well designed and the interfaces are quite user-friendly. It is really simple to interaction with multiple views and discover interesting phenomenon progressively.'

Usability and applicability The experts both agreed that the proposed system supports the in-depth exploration and understanding of mobility dynamics for pattern recognition and traffic management. One expert commended that 'The POI-mobility glyph is very useful for observing regional temporal patterns in real applications. By interacting with the flow map chart and locational view, I can also find the popular trips and destinations in one region, and infer the trip purposes. Thus, we can conduct targeted traffic flow management near popular destinations in different time periods.' The second expert said that 'It is difficult to capture valuable information from multi-source datasets by ourselves. This system really helps the domain experts with limited computer knowledge to explore the underlying patterns from big data.'
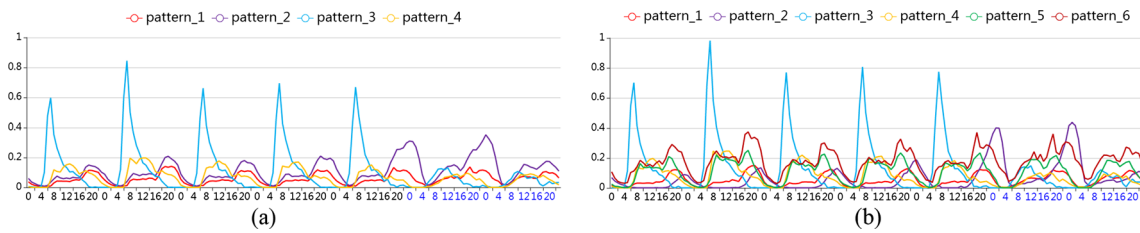
Fig. 13 The timeline charts for taxi data when $R = 4$ and $R = 6$. **a** The timeline chart for taxi data when $R = 4$; and **b** the timeline chart for taxi data when $R = 6$
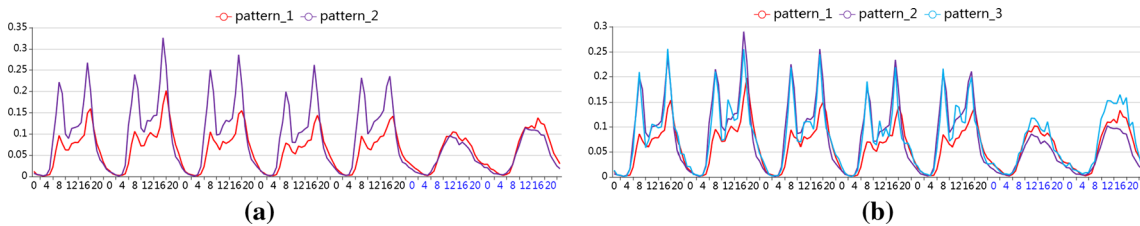
Fig. 14 The timeline charts for PBS data when $R = 2$ and $R = 3$. **a** The timeline chart for PBS data when $R = 2$; and **b** the timeline chart for PBS data when $R = 3$

Suggestions for system improvements The two experts also provided some valuable advices for system improvements. The first expert mentioned that 'Although the tensor decomposition method can classify the underlying spatiotemporal patterns automatically, how to classify the regions is still a problem. In the current system, the users need to observe the heat maps and select the important regions for further analysis. If the system can classify the regions based on their usage patterns, we can manage regions with different attributes more conveniently.' The second expert mentioned that 'The system is able to reveal general usage patterns for different transportation means in different regions. If it can provide new functions such as abnormal pattern detection, and help the users in interpreting such anomalies, we can plan better traffic controls to prevent future accidents.'

5.4 Discussion

The decomposition rank R controls the quality and quantity of features characterizing the derived mobility patterns. Increasing the rank number R improves the reconstruction accuracy. However, revealing the underlying spatiotemporal patterns is more important than reconstructing the original data. Therefore, there is a trade-off between the quantity and quality of features.

To determine the appropriate R, we compare the temporal variation of each pattern when R is different. For taxi data, we compare the timeline charts with R from 4 to 6. When $R = 4$ (Fig. 13a), patterns 2–4 are consistent with those of patterns when $R = 5$ (Fig. 3a). Pattern 1 ($R = 4$) is further subdivided into pattern 1 and 5 ($R = 5$), which both have evening peaks but with different peak numbers. When $R = 6$ (Fig. 13b), patterns 2–4 are also consistent with those of patterns when $R = 5$. Pattern 5 ($R = 5$) is similar to pattern 6 ($R = 6$), and pattern 1 ($R = 5$) is further subdivided. In general, when R is smaller, some similar patterns are merged. When R becomes larger, some patterns have a more detailed division. In our experiment, we set $R = 5$, containing the most dominating patterns.

The cycling patterns are not diverse as taxi patterns. Therefore, we compare the timeline charts with R from 2 to 4 for PBS data. When $R = 2$ (Fig. 14a), pattern 2 has obvious morning and evening peaks on workdays, and pattern 1 has higher evening peaks. When $R = 3$ (Fig. 14b), a new pattern 3 with noon peaks appears. When $R = 4$ (Fig. 4a), the patterns are further to be subdivided. Pattern 4 has double peaks. Pattern 2 has obvious morning peaks. Patterns 1 and 3 correspond to evening peaks with different flow numbers. In our experiment, we set $R = 4$.

6 Conclusion and future work

In this paper, we present an approach for the visual exploration of human mobility dynamics based on multi-source datasets. The tensor-based analysis method takes the rich content information into account in a unified manner, and a visual analysis system is designed to support the data exploration from various perspectives. Our approach is able to extract information that would be hidden using a single data subset. We demonstrate the efficiency of our approach via case studies. By using the system, we can infer the trip purposes, find the mixed functions for a region and gain a more insightful understanding of human mobility. In the future, because the flow map chart may look clutter when the trip number is large, we plan to employ force directed edge bundling to reduce clutter. Besides, as the domain experts are interested in abnormal events, we will try to design visual analytics system for abnormal pattern detection and interpretation based on multi-source datasets.

References

Cao N, Lin C, Zhu Q et al (2018) Voila: visual anomaly detection and monitoring with streaming spatiotemporal data. IEEE Trans Vis Comput Graph 24(1):23–33

Chen W, Huang Z, Wu F et al (2018a) VAUD: a visual analysis approach for exploring spatio-temporal urban data. IEEE Trans Vis Comput Graph 24(9):2636–2648

Chen W, Xia J, Wang X et al (2018b) Relationlines: visual reasoning of egocentric relations from heterogeneous urban data. ACM Trans Intell Syst Technol 10(1):2

Ferreira N, Poco J, Vo HT et al (2013) Visual exploration of big spatio-temporal urban data: a study of New York city taxi trips. IEEE Trans Vis Comput Graph 19(12):2149–2158

Fishman E (2016) Bikeshare: a review of recent literature. Transp Rev 36(1):92–113

Gao S, Janowicz K, Couclelis H (2017) Extracting urban functional regions from points of interest and human activities on location-based social networks. Trans GIS 21(3):446–467

Han Y, Moutarde F (2016) Analysis of large-scale traffic dynamics in an urban transportation network using non-negative tensor factorization. Int J Intell Transp Syst Res 14(1):36–49

Kolda TG, Bader BW (2009) Tensor decompositions and applications. SIAM Rev 51(3):455–500

Liu D, Xu P, Ren L (2019) TPFlow: progressive partition and multidimensional pattern extraction for large-scale spatio-temporal data analysis. IEEE Trans Vis Comput Graph 25(1):1–11

Pei W, Wu Y, Wang S et al (2018) BVis: urban traffic visual analysis based on bus sparse trajectories. J Vis 21(5):873–883

Ran B, Tan H, Wu Y et al (2016) Tensor based missing traffic data completion with spatial–temporal correlation. Phys A 446:54–63

Shi X, Yu Z, Chen J et al (2018) The visual analysis of flow pattern for public bicycle system. J Vis Lang Comput 45:51–60

Shi X, Lv F, Seng D et al (2019) Exploring the evolutionary patterns of urban activity areas based on origin-destination data. IEEE Access 7:20416–20431

Tang Y, Sheng F, Zhang H et al (2018) Visual analysis of traffic data based on topic modeling. J Vis 21(4):661–680

Xie C, Zhong W, Xu W et al (2018) Visual analytics of heterogeneous data using hypergraph learning. ACM Trans Intell Syst Technol 10(1):4

Yan Y, Tao Y, Xu J et al (2018) Visual analytics of bike-sharing data based on tensor factorization. J Vis 21(3):495–509

Yang D, Zhang D, Zheng VW et al (2015) Modeling user activity preference by leveraging user spatial temporal characteristics in LBSNs. IEEE Trans Syst Man Cybern Syst 45(1):129–142

Yao Y, Li X, Liu X et al (2017) Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. Int J Geogr Inf Sci 31(4):825–848

Yuan NJ, Zheng Y, Xie X et al (2015) Discovering urban functional zones using latent activity trajectories. IEEE Trans Knowl Data Eng 27(3):712–725

Zeng W, Fu CW, Arisona SM et al (2017) Visualizing the relationship between human mobility and points of interest. IEEE Trans Intell Transp Syst 18(8):2271–2284

Zhou Z, Yu J, Guo Z et al (2018) Visual exploration of urban functions via spatio-temporal taxi OD data. J Vis Lang Comput 48:169–177