J Vis
https://doi.org/10.1007/s12650-020-00707-1

REGULAR PAPER

Zhuochen Jin · Nan Cao · Yang Shi · Wenchao Wu · Yingcai Wu

EcoLens: visual analysis of ecological regions in urban contexts using traffic data

Received: 30 June 2020 / Revised: 17 August 2020 / Accepted: 19 August 2020

Abstract The increasing availability of spatiotemporal data provides unprecedented opportunities for understanding the structure of an urban area in terms of people's activity pattern and how they form the latent regions over time. However, existing solutions are limited in their capacity of capturing the evolutionary patterns of dynamic latent regions within urban context. In this work, we introduce an interactive visual analysis approach, EcoLens, that allows analysts to progressively explore and analyze the complex dynamic segmentation patterns of a city using traffic data. We propose an extended nonnegative matrix factorization-based algorithm smoothed over both spatial and temporal dimensions to capture the spatiotemporal dynamics of the city. The algorithm also ensures the orthogonality of its result to facilitate the interpretation of different patterns. A suite of visualizations is designed to illustrate the dynamics of city segmentation and the corresponding interactions are added to support the exploration of the segmentation patterns over time. We evaluate the effectiveness of our system via case studies using a real-world dataset and a qualitative interview with the domain expert.

Keywords Visual analysis · Urban segmentation · Matrix factorization · Traffic data

1 Introduction

The rapid development of urbanization during the past decades has significantly improved people's life. Tremendous efforts have been put on reasonably and optimally segmenting the city into functional regions

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12650-020-00707-1) contains supplementary material, which is available to authorized users.

Z. Jin · N. Cao · Y. Shi (✉)
iDVx Lab, Tongji University, Shanghai, China
E-mail: yangshi.idvx@gmail.com

Z. Jin
E-mail: zcjin.idvx@gmail.com

N. Cao
E-mail: nan.cao@gmail.com

W. Wu
Siemens Ltd., Beijing, China
E-mail: wenchao.wu@siemens.com

Y. Wu
Zhejiang University, Hangzhou, China
E-mail: ycwu@zju.edu.cn

Published online: 16 October 2020

to serve various needs for its citizens and best utilize the limited city resource. However, functional regions are typically defined based on a static boundary. This segmentation strategy rarely reflects an individual's day-to-day experience of the space in which they live and visit for a variety of purposes. Also, the use of these static boundaries has limited not only the city's ability to assess the dynamic processes that shape its urban areas, but also its opportunity to improve the city management with smart-city services (Su et al. 2011; Bakıcı et al. 2013). Therefore, a better understanding of the structure of an urban area in terms of people's activity patterns and how these patterns form the latent regions over time can provide profound insights for effective applications in urban planning and business intelligence.

The increasing availability of human mobility data generated within an urban context opens up unprecedented opportunities to better understand an urban area. Prior research studied mobility patterns in the urban context (Zheng et al. 2016a), and most of them focus on identifying predefined events or features in data (e.g., Wu et al. 2016; Zheng et al. 2016b). However, these approaches were limited in capturing the dynamic formation of regions in an urban area. There have been some attempts in developing automatic algorithms (Wang et al. 2014; Yuan et al. 2015) to extract latent regions in an urban area. However, analysts found it difficult to understand and interpret the result or combine their domain knowledge with real-world applications. Hence, we propose an analysis technique combined with human supervision to explore the *ecological regions* (i.e., *dynamic latent regions*) in urban context that reflect the city dwellers' dynamic moving patterns and capture how they share similar moving behavior during a short period of time. Visual analysis provides an effective way to involve human knowledge in a data exploration process by applying their perceptual abilities to the target dataset and leveraging their domain knowledge to guide the exploration. The state-of-the-art approach, MobiSeg (Wu et al. 2017), enables interactive exploration of people's movement to segment an urban area into regions while neglecting the continuity of people's patterns in either spatial or temporal domain. Thus, its approach cannot be directly employed to illustrate the evolutionary patterns of dynamic latent regions in the urban context.

To address the above issues, this work presents an interactive visual analysis approach, *EcoLens*, which allows analysts to progressively explore and analyze the complex evolutionary patterns of latent regions. To this end, we proposed a novel nonnegative matrix factorization-based (NMF-based) algorithm for dynamic latent region detection based on people's mobility patterns, which takes the temporal and spatial smoothness into consideration. Using the NMF-based algorithm, a set of visualizations was designed to illustrate the extracted regions within spatial and temporal context. To evaluate the effectiveness of the proposed algorithm, we conducted a comparative analysis. We demonstrated the performance of EcoLens through case studies using a large-scale real-world dataset, which consists of over 450,000 taxi trips collected in Manhattan from July 2014 to December 2014. We also reported the qualitative feedback from an expert in the field of urban planning regarding the usefulness of EcoLens. The results indicated that our system is capable of identifying the mobility patterns to form latent regions, uncovering the dynamics of latent regions, and interpreting the mobility patterns of regions.

The major contributions of this paper are summarized as follows:

- *System* We designed a novel visual analytic system, EcoLens, to help explore the dynamic segmentation patterns of a city using large-scale traffic datasets.
- *Algorithm* We introduced a novel NMF-based algorithm smoothed over both spatial and temporal dimensions to capture the spatiotemporal dynamics of the city. Orthogonality of the latent patterns is guaranteed to facilitate the interpretation of different patterns.
- *Evaluation* We evaluated the effectiveness of the proposed algorithm and the EcoLens system via comparative analysis, case studies, and a domain expert interview.

2 Related work

Our work builds on prior research work on urban segmentation and visualization of mobility patterns.

2.1 Urban segmentation

Urban segmentation has been studied extensively for years in the fields of urban planning and geographic information system (GIS). Remote sensing data (e.g., images) are frequently used (Deng et al. 2009; Seto and Fragkias 2005). These techniques recognized different regions by calculating visual differences based on satellite images. The results are usually limited by the low resolution and the missing of context details especially in regions with complicated geographic conditions. To address this issue, more and more research attention has been put on identifying functional urban regions based on people's daily activity patterns (Wang et al. 2014; Wu et al. 2017; Yuan et al. 2015). As early as the 1970s, Goddard (1970) analytically differentiated functional regions in central London based on taxi flows. Following this work, Yuan et al. (2015) recently employs latent Dirichlet allocation (LDA), a generative statistical model that was originally designed for text analysis, to identify the latent functional regions in a city based on people's mobility patterns. Kraft and Marada (2017) applied the local minimum and maximum values of transport intensities to delimit functional regions. Demšar et al. (2017) applied principal components analysis (PCA) to taxi flows for obtaining functional regions. Zhang et al. (2017) analyzed the patterns of the urban roads based on taxi GPS data. Wakamiya et al. (2015) applied nonnegative matrix factorization (NMF) to analyze urban area characterization based on Twitter data. When compared to these techniques which produce static segmentation results, our work focuses on revealing the regional dynamics. The algorithms and visualization designs are thus introduced.

Despite the aforementioned analysis driven methods, MobiSeg (Wu et al. 2017) is the first visual analysis system designed for interactive region segmentation in the urban context. As the most relevant work, MobiSeg also employed NMF for urban segmentation and introduced the visual interface to facilitate results interpretation and interactive latent region analysis, comparison, and exploration. When compared to MobiSeg, our work focuses on analyzing and revealing temporal patterns of the dynamic transition of the latent regions. To this end, we introduce a dynamic NMF algorithm that analyzes and smooths the transition over both temporal and spatial domains, which produces more continuous and interpretable results when compared to the results without smoothing. In addition, to facilitate the tracking of regional transition patterns and the change of the regional functions, we also introduce a pattern tracking algorithm based on a Sankey diagram design. Most importantly, we conduct case studies on real data which reveal interesting findings that can hardly be detected in MobiSeg.

2.2 Visualization of mobility patterns

Visual analysis of latent urban regions falls into the general topic of analyzing mobility patterns. Efforts have been devoted to developing visualization methods to meet the needs of analyzing and understanding mobility patterns within the urban context [see Zheng et al. (2016a) for a comprehensive survey]. Our goal in this work is to develop a visual analysis approach to present the evolution of mobility patterns, thus helping analysts explore and understand dynamic latent regions within the urban context. When presenting evolving mobility patterns, time and movement are two fundamental components of telling a full story and can help structure the information. Therefore, in this part, among the vast amount of visualization techniques, we focus on the visualization of time and movement which are the most relevant to our work.

There are various ways of mapping time to visual variables (Aigner et al. 2011; Zhao et al. 2019). Within an urban context, the axis-based design is one of the most popular methods (Wu et al. 2014; Zheng et al. 2016b) due to its simplicity and interpretability. Besides, temporal information can also be conveyed through a dynamic representation, resulting in visualizations that change over time automatically (i.e., animation), which is a popular design choice for visualizing the dynamics of a city (Kloeckl et al. 2011; Rosling 2009). However, as demonstrated by Robertson et al. (2008), the animation techniques are generally not effective for analysis tasks due to the limitation of human short-term memory. We employ both designs in our work; a Sankey diagram is used to provide an overview of the regional transition trend while an animated map view to illustrate the dynamic change of regions in spatial context.

When visualizing movements, there are three major types of techniques, including direct depiction, summarization, and pattern extraction techniques (Andrienko et al. 2013). Direct depiction techniques (Tominski et al. 2012) present paths of movement directly. Summarization techniques (Wu et al. 2016; Andrienko et al. 2017) conduct statistical calculations of movement and present the result based on divided spatial or temporal intervals. Pattern extraction techniques (Zheng et al. 2016b) enable an interactive discovery and analysis of various movement patterns. In this paper, we integrate different types of techniques and enhance them with new features. With EcoLens, analysts could observe the evolution of an area frame by frame and explore its corresponding mobility patterns interactively.

3 System overview

We designed EcoLens for revealing the dynamics of latent regions in which people share similar temporal mobility patterns. Following a user-centric design process, we worked closely with a domain expert who is a researcher at the institute of urban planning and design in China. Regular meetings with the expert were scheduled and lasted for about six months to help us understand the requirements and refine the prototype. During the meetings, we focused on discussing what kinds of dynamic patterns need to be captured and how to reveal and interpret them within the urban context. As a result, we found that people's mobility pattern is the most critical feature to capture as it is directly relevant to their daily behaviors and thus implies the functionality of a region. For example, during the weekdays, people tend to travel from home to the office in the morning. This mobility pattern suggests potential residential areas and business areas. More specifically, the desired system should satisfy the following requirements:

R1 *Identifying the mobility patterns occurred in different areas to form latent regions for investigation* The system should be able to differentiate mobility patterns occurred in various urban areas based on people's collective daily moving behaviors that are extracted from a large dataset. Areas with similar patterns should be further grouped into latent regions to help imply the corresponding regional functionality within the urban context. To support the efficient identification and interpretation of the latent regions, the visualization should present the segmentation result of the latent regions with geographic information, the mobility features of the patterns, and the evolution of latent regions within the temporal context.

R2 *Capturing the dynamics of the latent regions over time to uncover regular or irregular regional transition patterns* The system should be able to reveal the dynamic change of the spatial mobility patterns and the corresponding change of the latent regions. It is necessary to provide a high-level summarization of the overall change of the city over time. This will help analysts identify the regular changes due to people's regular daily behaviors and thus help them identify those irregular ones due to certain events.

R3 *Facilitating the regional pattern comparison, exploration, inspection, and interpretation* The system should also be able to intuitively illustrate the aforementioned patterns and the corresponding changes of the latent regions in the rich urban context so that analysts can easily explore, compare, and understand the patterns and their changes to make a proper conclusion as well as a correct decision. The visualization should be able to reveal the raw traffic data inside each latent region and provide the corresponding statistical information to help with the interpretation, validation, and comparison of the region segmentation results.

According to the design requirements, we develop EcoLens, an interactive visualization system for analyzing the evolution patterns of latent regions within the urban context. The system consists of three primary modules, including the *prepossessing module*, *analysis module*, and *visualization module* (Fig. 1). The *prepossessing module* is designed to clean the raw data (i.e., the taxi-trips in our case) and transform them into the matrix time series with desired features. This whole prepossessing step runs in parallel on a Spark cluster, and the processed data are stored in MongoDB[1] for later querying. This module is designed to extract the collective moving behaviors of people from raw data (**R1**). The *analysis module* derives the latent regions (**R1**) and the evolution patterns (**R2**) overtime based on the preprocessed data via nonnegative matrix factorization. The *visualization module* presents the analysis results via multiple coordinated views. These views reveal the evolution of the latent regions and facilitate the interpretation of the corresponding pattern within each region. Various interactions are provided to support flexible data exploration and result calibration (**R3**). In the next, we will describe the details of the analysis and visualization modules in the following sections.

4 Context preserving dynamic region segmentation

In this section, we introduce the algorithm used in the analysis module that is developed for revealing the regional dynamics of the traffic evolution patterns inside a focal urban area. The proposed algorithm leverages the nonnegative matrix factorization (NMF) Lee and Seung (1999) and smooths the change over both the spatial and the temporal dimensions to facilitate interpretation. The NMF algorithm is used to

[1] https://www.mongodb.com/.

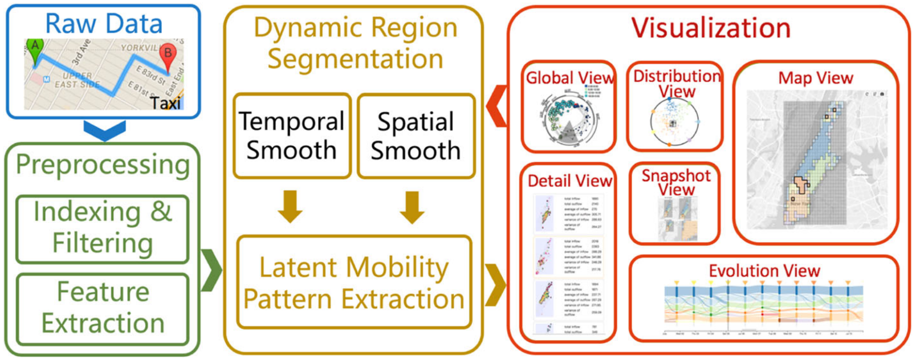

Fig. 1 The system pipeline of EcoLens. Three primary modules, including preprocessing module, analysis module, and visualization module, support the analysis of evolution patterns of latent regions within the urban context

decompose a sparse matrix into the product of two nonnegative matrices. In the context of urban, decomposing the original feature matrix of transportation can obtain two matrices that capture the semantics of latent traffic patterns and the spatial distribution of these latent patterns, respectively. In addition, unlike SVD which may produce negative values in the analysis results, our method guarantees nonnegative values, which are meaningful and interpretable. We first describe the data and the corresponding features used in our prototype system, followed by the algorithm details as well as the design rationales.

4.1 Feature extraction

Our prototype system employs the public New York City taxi trip dataset[2] to capture the change of regional mobility patterns. To this end, we divided the Manhattan area, our focal investigation region, into N grids ($N = 300$ with the granularity of 0.005 longitude $\times$ 0.005 latitude in our implementation) and counted the number of incoming and outgoing trips in each grid as the grid's features for later analysis. Each grid i is described by a $2N$-dimensional feature vector with the field p in the vector indicates the number of trips from the ith grid to the pth grid. Therefore, the first N fields in the vector indicate the number of outgoing trips from grid i to the rest $N-1$ grids and the last N fields indicate the number of incoming trips from other grids to the ith grid. In this way, during a given time interval t ($t = 2$ hours in our implementation), a $2N \times N$ feature matrix $\mathbf{X}_t$ can be obtained, which captures the mobility patterns during t. The matrices from different intervals thus formed a feature matrix time series. This matrix series characterizes mobility patterns in each region over time and is used for the later analysis.

4.2 Dynamic region segmentation

We propose a context preserving algorithm based on nonnegative matrix factorization (NMF) to analyze the regional dynamics of the taxi trips captured in the aforementioned feature matrix time series. The algorithm optimizes and balances among four carefully designed terms that are formally defined as follows:

$$W_t, H_t, M_t = \operatorname{argmin}(||X_t - W_t H_t^T||^2 \tag{1}$$

$$+ \alpha||W_{t-1} - W_t M_t^T||^2 \tag{2}$$

$$+ \beta||H_{t-1} - H_t M_t^T||^2 \tag{3}$$

$$\begin{aligned} &+ \lambda \sum_{i,j} A_{ij}||h_{ti} - h_{tj}||^2) \\ \text{s.t.}\quad & W_t^T W_t = I, t = 1, 2, \ldots, n, W_t \succeq 0, H_t \succeq 0, M_t \succeq 0 \end{aligned} \tag{4}$$

Pattern extraction The first term is proposed to extract latent mobility patterns from the raw data. It employs the nonnegative matrix factorization (NMF) to decompose a feature matrix $(\mathbf{X}_t)_{2N\times N}$ into the product of two nonnegative matrices $(H_t)_{N\times K}$ and $(W_t)_{2N\times K}$ that, respectively, captures the spatial distribution of the latent

[2] http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml.

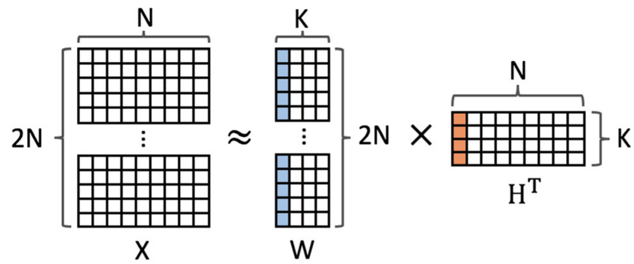

Fig. 2 Matrix X is decomposed into the product of matrix H and W. The column in H^T highlighted in red indicates the likelihood of the K mobility patterns occurred in the 1st regions. The column in W highlighted in blue shows the probability of the 1st mobility pattern having a certain feature

patterns as well as the pattern semantics. Specifically, $(H)_{N\times K}$ indicates the likelihood of each of the K patterns occurred in each of the N regions. $(W)_{2N\times K}$ shows the probability of a latent pattern having a certain feature. As shown in Fig. 2, the red column in matrix H shows the likelihood of the patterns occurred in the first regions and the blue column in matrix W shows the probability of the first pattern having a certain feature. Here, K is the number of desired latent patterns to be found during the analysis. It is a hyperparameter usually given by the analyst before the analysis. Note that the number of latent patterns, K, is usually the prior knowledge provided by the analysts. However, in many real applications, the ground truth of K is unknown. To address this issue, in our implementation, we employ Mean-Shift, a nonparametric clustering algorithm (Cheng 1995), to compute the clusters at each timestamp and then use the numbers of clusters as the values of K for our analysis. Although this approach is heuristic, it provides meaningful results.

Temporal smoothness The second and the third terms are the regularization terms that ensure the temporal smoothness of the analysis result. They, respectively, preserve the similarity of W_t and H_t across different time to eliminate the dramatic sudden change due to the noisy data that may break the overall the transition trend of the regional mobility patterns. α, β controls the degree of smoothness. Considering the number of patterns K may vary from time to time, a transition matrix M_t is introduced to connect patterns at different time intervals together. A element $M_{t,i,j}$ in M_t implies the probability of a previous pattern i at time $t-1$ transiting to a current pattern j at time t. Thus, $W_tM_t^T$ and $H_tM_t^T$ should be close to the previous $W_t - 1$ and $H_t - 1$ at time $t-1$.

Spatial smoothness The fourth term is the spatial smooth regularization term that ensures a region will share similar mobility patterns with its neighborhood. This design is due to the common understanding and observation that nearby regions will show similar mobility patterns as the functional area (e.g., CBD area) may locate across multiple adjacent regions. Here, an adjacency matrix A is introduced with each element $A_{ij} \in \{1, 0\}$ indicates whether or not two regions i and j are adjacent to each other. The pattern differences (i.e., $||h_{t,i} - h_{t,j}||^2$) between those adjacent ones are minimized and the degree of the minimization is controlled by λ.

The above optimization problem can be solved based on block coordinate descent (Kim et al. 2014). The outputs of the algorithm, including the latent patterns W_t, the pattern distribution in different regions H_t, and the pattern transition probability M_t, are captured at each timestamp t. Based on H_t which indicates the likelihood of each of the patterns occurred in each of the regions, the regions with the same maximum-likelihood pattern form a latent region via a clustering analysis. All these produced analysis results are used for building meaningful visualization views that will be introduced in the next section.

5 Visualization

In this section, we describe the visualization designs that illustrate the analysis results of the spatiotemporal dynamics of the city. The interface of EcoLens consists of seven views (Fig. 3): (1) the *global view* that shows the overview of the segmentation results in the temporal context; (2) the *map view* that combines the segmentation results with geographic information, as well as a flow glyph design that illustrates the raw traffic trip information of the regions; (3) the *pattern view* and (4) the *grid view* that present the mobility feature of each latent region, as well as statistics of the raw information to help validate the mobility pattern; (5) the *distribution view* that displays the probability distribution of different mobility patterns over regions;

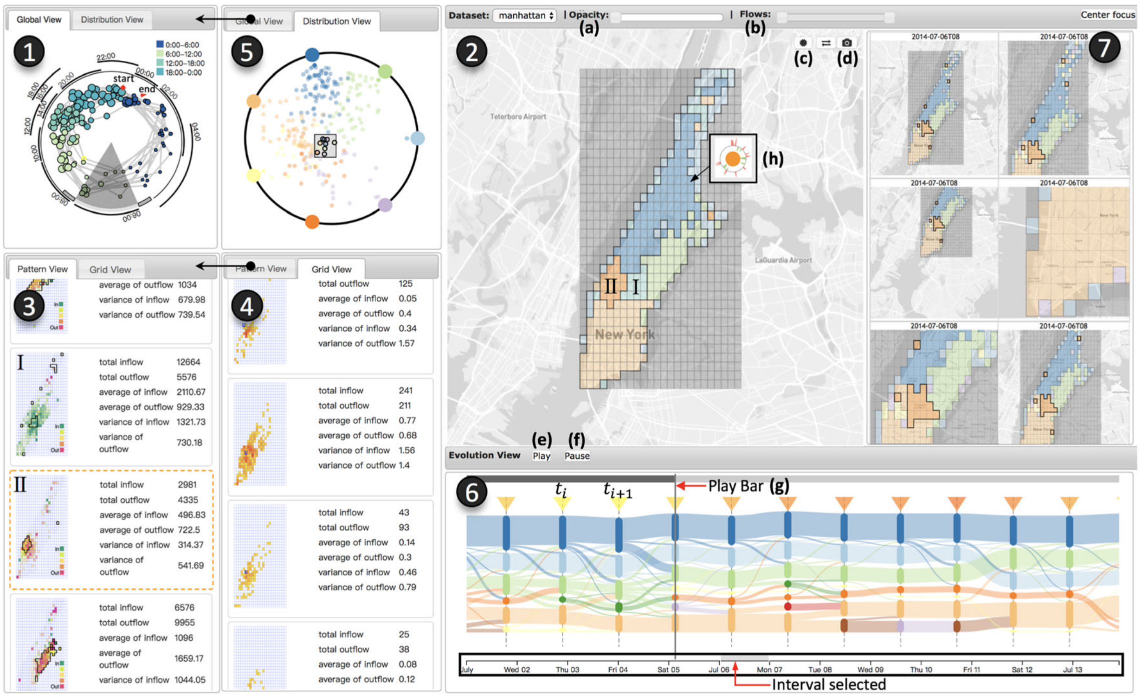

Fig. 3 The user interface of EcoLens consists of seven major views: (1) global view, (2) map view, (3) pattern view, (4) grid view, (5) distribution view, (6) evolution view, and (7) snapshot view

(6) the *evolution view* that shows the dynamics of the latent regions in the temporal context, and (7) the *snapshot view* that allows users to take a snapshot of the segmentation results shown in the map view for later retrieval and further analysis. These views are interactively linked to illustrate the dynamics of latent regions generated based on the aforementioned segmentation results.

We employ four color encoding schemes in our design. The first color scheme ranging from light green to dark blue shows the categorization of segmentation results in the global view (Fig. 3(1)). The second color scheme is used to encode the segmentation results in the map view (Fig. 3(2)), distribution view (Fig. 3(5)), and evolution view (Fig. 3(6)). Regions with the same color at a given time range indicates they share similar mobility patterns over spaces. Regions with the same color across different time range indicate these regions have a similar pattern changing trend. The third color scheme is used to represent the difference between the incoming and outgoing traffic flow in the pattern view (Fig. 3(3)) and the glyph in the map view. The colors ranging from green to yellow, and to red indicate the larger, equal, and smaller incoming flow when compared to the outgoing flow. We use the fourth color scheme ranging from yellow to red to indicate the amount of the flow in the grid view (Fig. 3(4)) and the triangular glyph in the evolution view (Fig. 3(6)).

5.1 Global view

The global view (Fig. 3(1)) provides an overview of the segmentation results in time series whose design is inspired by van den Elzen et al. (2016). It illustrates the distribution of the overall segmentation results of a focal area in a feature space. In particular, the segmentation results at different time intervals are summarized and shown as points in the view with the point size indicating the amount of traffic within the corresponding time period and the color indicating the parts of the day (i.e., dawn, morning, afternoon, and night). The characteristic of a point, v_t, is captured by an $N \times 2N$-dimensional feature vector which is the vectorization results of the corresponding feature matrix X_t introduced in Sect. 4 as shown in Fig. 4. Here, N indicates the number of grids and $2N$ is the number of features of each grid. With the above feature vector, we illustrate the distribution of the overall segmentation results (i.e., points in the view) in the feature space via principal component analysis (PCA) Dunteman (1989). These points are further connected by a timeline from the earliest time to the latest time in the data with the start and end points, respectively, marked with a red rectangle and red arrow.

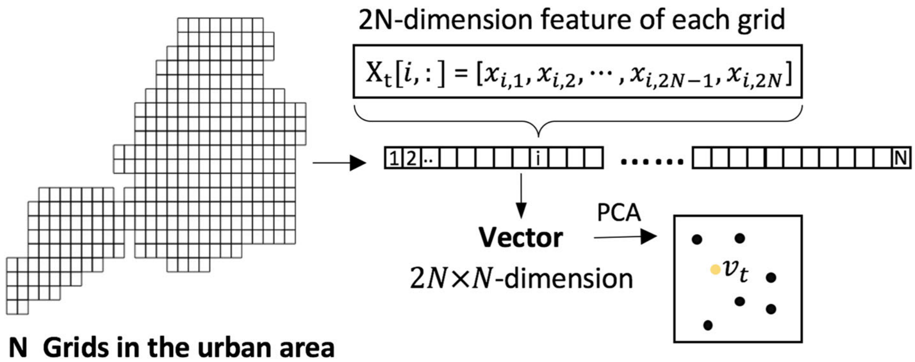

Fig. 4 The characteristic of point v_t is captured by an $N \times 2N$-dimensional feature vector, where N indicates the number of grids in a region and $2N$ is the number of features of each grid

Figure 3(1) illustrates the visualization results of the aforementioned NYC taxi trip data collected from the Manhattan area, which forms a periodical circular pattern with each loop in the circle indicates a day that is segmented into 12 time periods. Within each period, a segmentation is calculated and thus are visualized as a point in the view. A circular brush tool is also designed in the view to facilitate the selection of a time range and the corresponding points. The selected points will be expended into details and shown in other views for exploration and comparison.

5.2 Map view

We overlay the segmentation analysis results on a map to illustrate its spatial context (Fig. 3(2)). In particular, the equal-sized grids, in which the mobility features are calculated, are visualized in the background. Each grid i is colored by their primary mobility patterns (i.e., the largest field in the vector $H_t[i,:]$). The grids share similar mobility patterns are grouped together into latent regions, and the boundary of the latent regions is further highlighted by a thicker line.

To summarize and illustrate the raw traffic flows inside each grid, we introduce a novel *flow-glyph* design in the map view as shown in Fig. 5. The design of this glyph aims to encode and illustrate three types of the following information regarding a focal grid: (1) the traffics inside the grid; (2) the exchange of the traffics between the focal grid and other grids; (3) the statistic of the total amount of traffics related to the grid. The glyph, as shown in Fig. 5a, follows a circular design that consists of two major components: (1) the center circle with the size indicating the total amount of relevant traffic flows in the grid, and (2) the outer ring that summarizes the traffics to or from 72 different directions (5 degrees a direction) with the focal grid in the center. The number of the flows is visualized as bars (Fig. 5b) and enhanced by a colorful peak with red indicates outflow and green indicates inflow (Fig. 5c). Intuitively, the outflows are visualized outside the ring, whereas the inflows are visualized inside the ring.

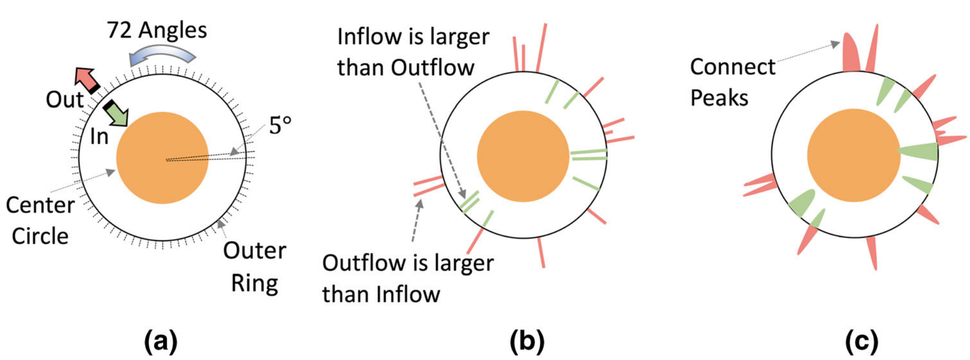

Fig. 5 Flow-glyph design. **a** The center circle and outer ring, respectively, show the amount and direction of traffic flows in the grid. **b** These flows are visualized as bars and **c** enhanced by a colorful peak with red indicates outflow and green indicates inflow

5.3 Evolution view

The evolution view illustrates the temporal transition trend of the latent regions, as shown in Fig. 3(6). It employs a Sankey diagram design in which x-axis indicates the time and the vertical nodes at each timestamp indicate the latent regions generated at that time. The transitions of the latent regions across different timestamps are shown by the strips whose thickness indicates the number of raw grids merged into or split from a latent region. Here, the colors of the strips provide a visual hint, from which an analyzer can trace the change across different timestamps.

The number of latent regions may vary dramatically over time, thus making the assignment of a proper color to a node or a strip is a difficult problem. The goal is to find the best color matching so that the colors of the succeeding nodes and strips can best inherit from the colors of the previous nodes so that users can easily follow and track the transition trend of the latent regions. We convert the problem into a maximum-weighted bipartite matching problem by constructing a bipartite graph. In particular, the regions at two adjacent stages are regarded as the nodes in the graph. We connect a latent region at a previous stage to a latent region at the succeeding stage if these two regions have overlapped underlying grids. The weight of the connection is given by the number of overlapped grids. In this way, a maximum-weighted matching between the regions at two adjacent stages will help find the best inherent colors for the succeeding nodes from their most relevant nodes at the previous stage. The problem can be optimally solved based on the Kuhn–Munkres algorithm (Kuhn 1955). It is worth mentioning that in our implementation based on the NYC taxi trip data, we initially segment the Manhattan into functional regions and assign each region a color based on the administrative divisions of the city. These initial colors are then used for the color assignments and matching during the rest of regional transition processes. In Fig. 3(6), the blue vertices are the latent regions at the previous stage and the green vertices are the latent regions at the current stage. A blue vertex and a green vertex are connected if they have overlapped grids and the weight of the connection is given by the number of the overlapped grids. Here, the solution of the maximum weighted matching are links highlighted in orange.

5.4 Other views

The EcoLens also uses other supportive views to illustrate information details from different perspectives.

Pattern view The pattern view (Fig. 3(3)) shows a list of mobility patterns for each latent region. As described in Sect. 4, the mobility pattern of a latent region can be presented by a 2N-dimensional feature vector (a column vector in the matrix W_t). The first N fields in the vector indicate the occurrence probability of the outgoing trips from this latent region to all the N grids in the urban area, while the last N fields indicate the occurrence probability of the incoming trips from all the grids to this latent region. To visualize this feature vector intuitively, we employ a red-to-yellow-to-green color gradient on the heatmap of each latent region. If the probability that people in a grid enter into the latent region (highlighted via black strokes in the heatmap) is higher than the probability that the people leave from the latent region to this grid, the grid in the heatmap will be colored in green. Otherwise, the grid will be colored in red. If both the probability of outing and incoming trips in a grid are equal to zero, the grid will be filled with no color. For example, in the heatmap of the highlighted item (Fig. 3(3-II)), most of the grids on the east side of the latent region are colored in red. This pattern suggests that people in the latent region II are likely to enter into the east side of this latent region. In the map view, we can observe that the east of latent region II is latent region I. We further inspect the mobility pattern of latent region I (Fig. 3(3-I)) and find that the grids around latent region I are colored in green, indicating people in those grids are likely to enter into this latent region. Therefore, we could infer that people in the latent region II intend to travel to latent region I.

Grid view In the grid view (Fig. 3(4)), the grid view further reveals the detailed mobility patterns of grids that compose a certain latent region selected in the pattern view. Each item in the grid view consists of statistic information and a heatmap. In the heatmap, the color encodes the total amount of flow between the grid and others. The darker the color, the larger the amount.

Distribution view The distribution view (Fig. 3(5)) reveals relationship among mobility patterns and the regions by visualizing the details of the matrix H_t. Each row vector of the matrix H_t indicates the probability distribution of the mobility patterns over a certain region. A circle is divided into K (the number of the mobility patterns derived from Sect. 4) sections equally, so as to build a barycentric coordinate. The categories of the mobility patterns are illustrated as colored nodes along the boundary of the circle. The regions are encoded as scattered points in the barycentric coordinate. Each pattern and its corresponding points of the regions are assigned with a specific color. The coordinate of the point representing region i is calculated as follows:

$$rx_i = \sum_j^k H_{t,i,j} \times cx_j, \quad ry_i = \sum_j^k H_{t,i,j} \times cy_j \tag{5}$$

where rx_i and ry_i are x-coordinate and y-coordinate of region i, cx_j and cy_j are x-coordinate and y-coordinate of mobility pattern j, and $H_{t,i,j}$ is the element at row i and column j of matrix H_t.

6 Evaluation

We evaluated the effectiveness of the proposed algorithm and the EcoLens system via comparative analysis, case studies, and a domain expert interview. Our evaluation is based on a dataset consisting of over 450,000 taxi trips from July 2014 to December 2014 collected from over 50,000 Yellow Cabs in the Manhattan area. We segmented the city into a web of grids with a granularity of 0.005 lng $\times$ 0.005 lat and calculates the regional features every two-hour to ensures a reasonable computation time and precise preservation of the traffic patterns.

6.1 Algorithm validation

We validated the effectiveness of the algorithm by estimating the constraints and regularization terms introduced in our dynamic city segmentation algorithm.

6.1.1 Verification of the temporal smoothness

To verify the effectiveness of the temporal smooth regularization terms (Eq. 4(2, 3)), we compare the analysis results generated without/with the term in the algorithm as, respectively, shown in Fig. 6a, b. Generally, the transitions of the latent regions change dramatically in Fig. 6a when compared to the case shown in Fig. 6b. In particular, the highlighted green strip (i.e., a latent region) in Fig. 6a splits into branches during the period from 12:00 to 14:00, which are later merged into another latent region shown as the blue strip in the next stage during the period of 14:00–16:00. The corresponding map view (Fig. 7) provides more insights into the changes of these latent regions that helped with the validation of the results. In particular, blue and green regions, respectively, correspond to the above blue and green strips, illustrating the areas with two different latent patterns. These two areas changed dramatically in Fig. 7a: a subarea, highlighted by the red box, originally in the green region merged into the blue region. This subarea, according to the map, is the East Harlem, where schools, residential areas, and parks are located and the traffic patterns are seldom changed during the non-traffic hours like the period from 12:00 to 16:00. A further investigation of the raw data verified our guessing; the change is due to a small number of random taxi trips which are the data noise that affects the analysis result. A slight smooth over temporal dimension addressed this problem (Fig. 7b), which verified the usefulness of the temporal smooth regularization term.

Using the temporal smoothness can result in missing some outliers caused by emergency events in the urban. However, the change of the flow data caused by emergencies is more dramatic than that due to random trips. We use the parameter α to control the degree of temporal smoothness. When analyzing outliers, the influence of spatial smoothness can be reduced by decreasing the value of α.

6.1.2 Verification of the spatial smoothness

To verify the effects of the regularization term for spatial smoothing (Eq. 4(4)), we compare the analysis results produced without/with spatial smoothness as shown in Fig. 8a, b, respectively. In this example, the region highlighted in the red circle is a part of Stuyvesant Town, a small residential area, in which people suppose to behave similarly. Therefore, the discontinuity of the regional clusters shown in Fig. 8a is most likely due to the noise of the input data instead of different mobility patterns. This problem has been eliminated by adding the spatial smoothness regularization term into our algorithm as shown in Fig. 8b.

Using the spatial smoothness can also result in missing outliers in the flow data (e.g., the flow data of different places in a functional area can be different). However, we mainly focus on discovering latent

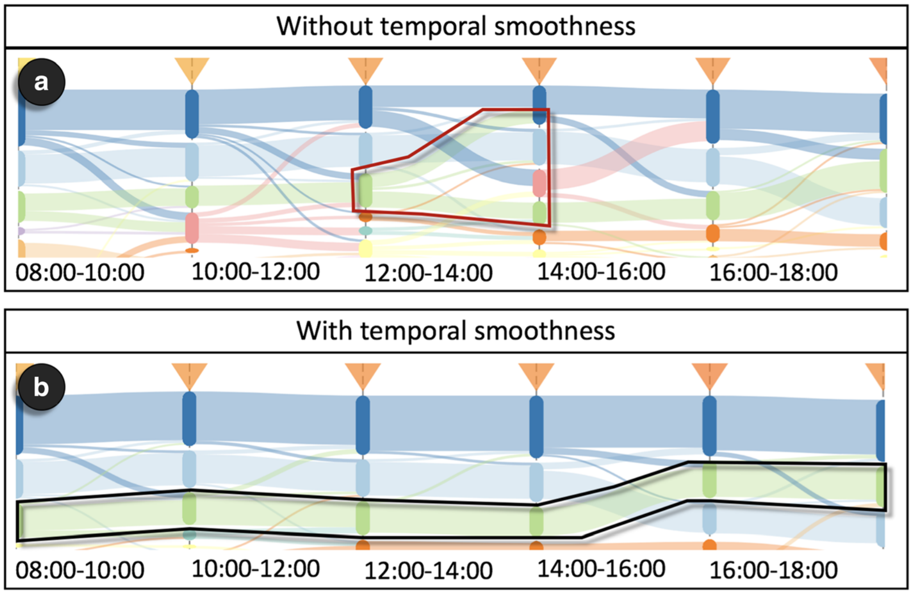

Fig. 6 The analysis results in the evolution view generated **a** without and **b** with the temporal smooth regularization terms

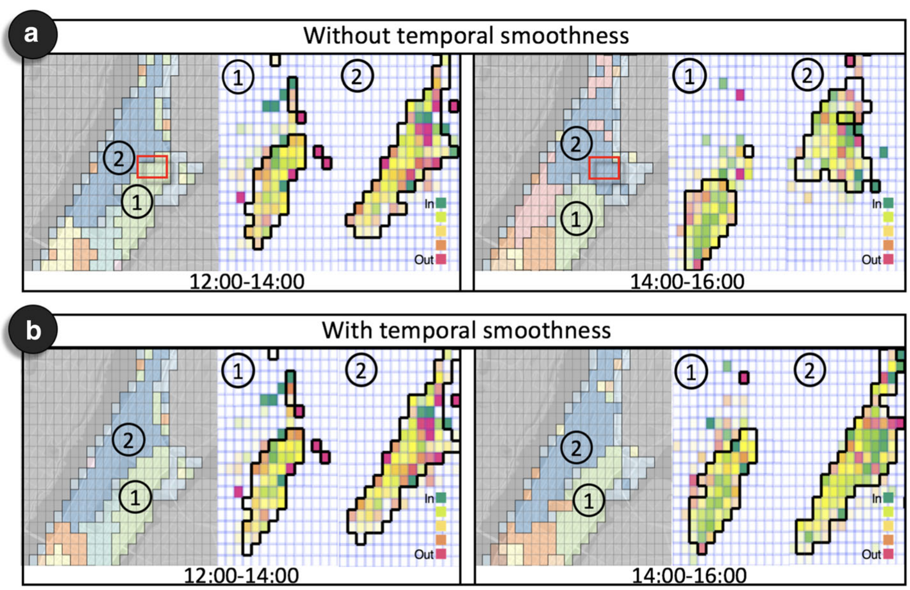

Fig. 7 The analysis results in the map view and pattern view generated **a** without and **b** with the temporal smooth regularization terms

regions, which leads to the hope that the nearby regions in a functional area will share similar mobility patterns. We use the parameter β to control the degree of spatial smoothness. The functional areas can be further divided by decreasing the value of β.

6.1.3 Verification of the orthogonality

To estimate the effects of the orthogonality constraint, we also compare the analysis results produced without/with the constraint as shown in Fig. 9a, b, respectively. Obviously, the patterns shown in Fig. 9b are more differentiable than the one shown in Fig. 9a. This finding is further verified by the corresponding correlation matrices shown in Fig. 9(I, II). In these matrices, each column or row indicates a latent pattern. A cell at the *i*th row and the *j*th column indicates the correlation value of the *i*th and the *j*th patterns, which is

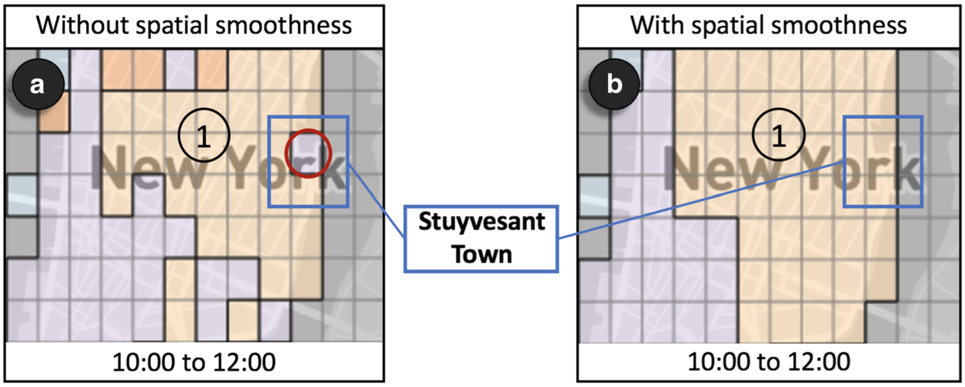

Fig. 8 The analysis results in the map view generated **a** without and **b** with the spatial smooth regularization terms

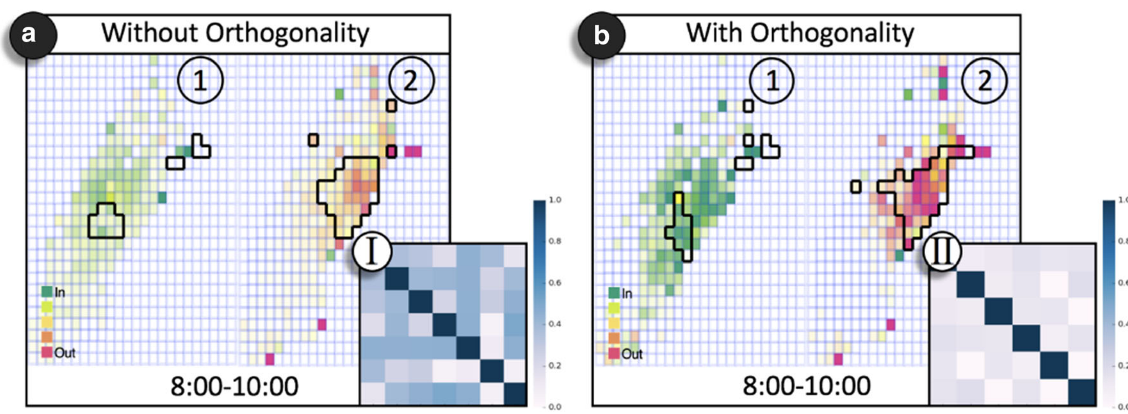

Fig. 9 The analysis results in the pattern view generated **a** without and **b** with orthogonality constraint. (I) and (II) show the corresponding correlation matrices of **a** and **b**, respectively

proportional to the color saturation. Therefore, Fig. 9(II) illustrates the patterns produced by following the orthogonality constraint are less relevant to each other when compared to the case shown in Fig. 9(I).

6.2 Case study

To further evaluate the usability and usefulness of the EcoLens system, we provide two case studies demonstrating its capability in analyzing the change of the mobility patterns in Manhattan, NYC.

6.2.1 Evolution exploration

We analyzed the daily evolution of the mobility patterns by comparing the results produced at four different time intervals on July 9th, 2014, respectively, in the morning, the afternoon, the evening, and at night. The visualization results are captured in Fig. 10.

As shown in Fig. 10a, most of the grids in the latent region (1) and (3), the residential areas, are in red. This suggests that these areas have a greater outflow in the morning. In comparison, the regions marked as (2) and (4) have a greater inflow (shown in green) at the same time, where are the CBD (i.e., Central Business District) areas in the town. This reveals the mobility pattern of morning traffic hours within Manhattan. At noon and in the early afternoon (12:00–14:00), as shown in Fig. 10b, the boundary of the latent regions (1,2,3) largely remains the same as that in the morning. However, the traffic patterns are dramatically changed as illustrated in the aside heatmap. The yellow color indicates the amount of incoming and outgoing traffic flows in these regions are similar, which implies people travel around for lunch inside the nearby regions. We also observed a new region (5), the financial area in the city, split from region (4). The aside heatmap revealed the reason for this change as the traffics is seldom across these two regions in

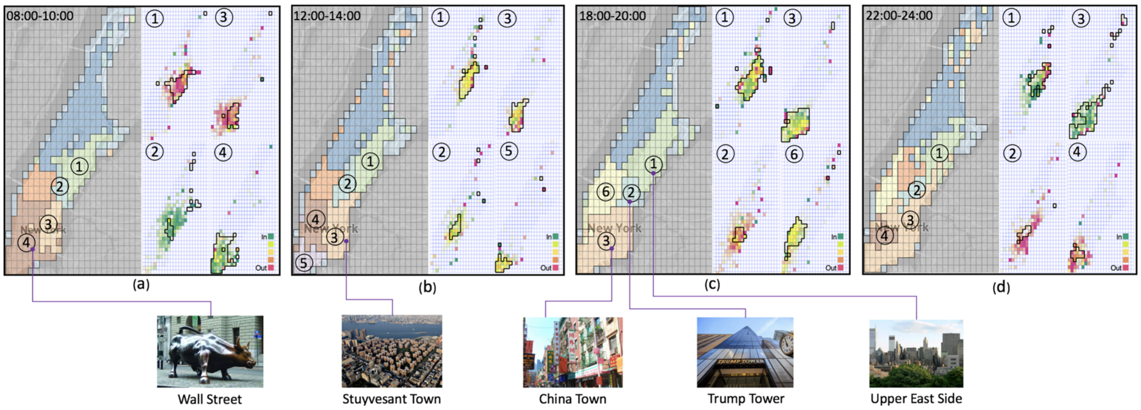

Fig. 10 The EcoLens system summarizes the daily evolution of the mobility patterns in Manhattan, NYC. The system employs the *map view* (left) and *region view* (right) to, respectively, show the result of urban segmentation and the mobility feature of each latent region. Four mobility patterns at different time intervals are captured, including **a** the morning from 8:00 to 10:00, **b** the early afternoon from 12:00 to 14:00, **c** the evening from 18:00 to 20:00, and **d** the night from 22:00 to 24:00

this period of time, thus making them separated from each other. We believe people work in this area such as the stock dealer will be too busy to leave this area at noon.

Later in the evening (18:00–20:00) as shown in Fig. 10c, the regions (3, 4, 5) are merged together, forming an area labeled as (3) where people are moving around and the incoming and outgoing traffic flow are balanced. This is due to some people leaving for home and some of them coming for dinner and fun, since the area is also served as the entertainment district in the urban area where Broadway and many bars and restaurants are located. Similar patterns also occur in region (1) and (6). However, the traffic pattern shown in region (2) changes dramatically when compared to that of an early stage shown in Fig. 10b. It suggests that people start to leave this area and the traffic spread all over the nearby regions. Fig. 10d illustrates the traffic pattern from the CBD to the residential areas in the city, which is just opposite to the patterns shown in Fig. 10a.

6.2.2 Anomaly detection

EcoLens was also used to detect anomalous situations. Fig. 11 illustrates an overview of the mobility patterns at different timestamp calculated based on people's daily moving behavior. It generally reveals a strong periodical pattern through the connected points shown in a circular form, with each circle indicates the period of a regular weekday. Among all the points connected by the timeline, a point representing the time interval 22:00–24:00 on July 4th, highlighted in the red circle, is considered to be an outlier as it is laid out away from other points in the surrounding context that captures the history of the same period of time. A detailed investigation of this outlier is illustrated in Fig. 12. We believe this abnormal pattern is due to the change of traffics around the Brooklyn Bridge on July 4th (the Independent day) during the period of 22:00–24:00. Usually, little volume of traffic goes through the bridge at the late night (Fig. 12a). However, on the Independent day, the fireworks aside the river attract a great number of people which dramatically increases the amount of traffic flow (Fig. 12b), thus resulting in a different traffic pattern and captured by the global view of EcoLens system.

6.3 Expert interview

We collected user feedback and comments on the EcoLens system through an in-depth interview with a domain expert from the institute of urban planning and design in China. We first showed a tutorial that explains the goal, visual encoding, followed by a demonstration of an example illustrating the evolution process of an urban area to help the expert get familiar with the system. The expert was asked to explore the capabilities of our system and analyze the dynamic patterns in the urban area with his domain knowledge.

System Generally, the system impressed the expert. He commented that the results were "*reasonable*," and the design of the visualization views was "*intuitive*" and "*comprehensive*." "*We used to analyze the urban areas based on the statistical information overlaid on top of a map. This system provides us a new approach to exploring the dynamics of the urban data.*," which was considered to be "*novel*" and "*useful*."

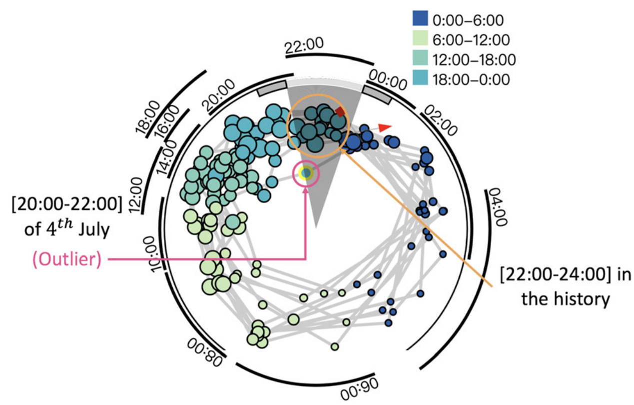

Fig. 11 In the global view, a point representing the time interval 22:00–24:00 on July 4th (highlighted in the red circle) is considered to be an outlier

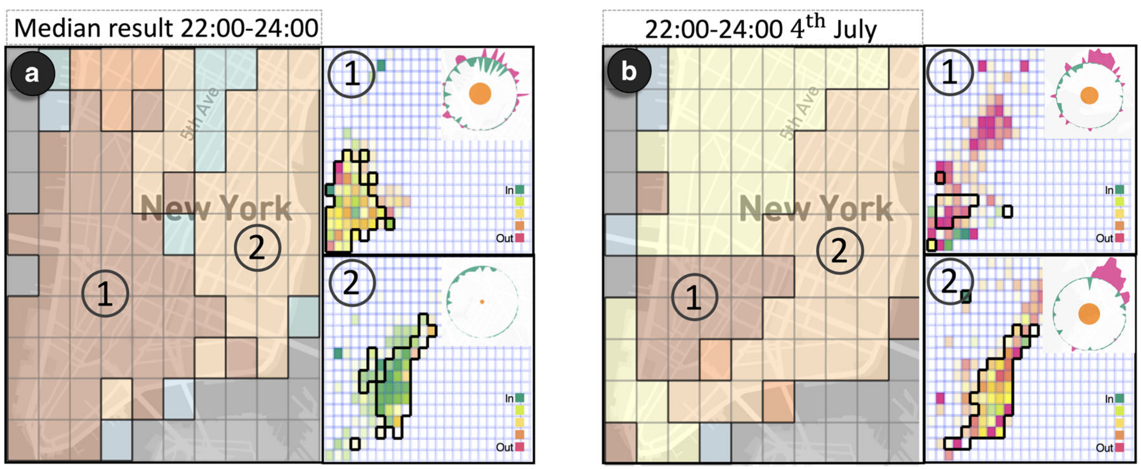

Fig. 12 By comparing to **a** the normal pattern, **b** the anomalies traffic flows near the Brooklyn Bridge on July 4th, 2014 (the Independent day) is revealed

Visualization The expert believed that most of the visualizations well supported the design requirements. For example, when analyzing the evolution of the urban area, the expert compared the results with the urban planning map and pointed out that “the areas gather together in different time intervals showing different patterns. These patterns are meaningful given the functionality (e.g., residential or CBD areas) of the underlying urban areas.” He also believed that the overview is useful as it provided a “clear periodical pattern” and “revealed a few outliers that fail to align with others.” The expert also commented on the glyph design and believed it was a “good and novel” approach for summarizing and illustrating raw trip data. He also believed the snapshot view was “particular useful” for analysis tasks. He said “the system is useful and practical in terms of supporting the exploration of city dynamics.”

Application The experts further mentioned the applications of this system in the field of smart-city services. “ *With the help of this system, we can know more about the dynamics of the city. The information is useful in various applications, such as smart traffic light control for energy saving*.” He also suggested that “you can use different types of mobility data, such as subway records and mobile phone locations, to detect more meaningful features for analyzing. [...] Currently, the analysis in each time period is static. Adding the time dimension to the data matrix will make the result more powerful.”

7 Conclusion

This paper presents a visual analysis system, EcoLens, for analyzing the dynamic latent regions that shape the urban area. EcoLens was designed according to real-world requirements, such as interpreting mobility pattern and urban evolution. An NMF-based algorithm was introduced to reveal the regional dynamics of the mobility evolution patterns inside a focal urban area. We evaluated the performance and effectiveness of EcoLens using a taxi-trip dataset of Manhattan Island through case studies and an interview with a domain expert. Our study results indicated that our system is capable of identifying the mobility patterns to form latent regions, uncovering the dynamics of latent regions, and interpreting the mobility patterns of regions. In the future work, we plan to use more data resources in the analysis and process the growing scale of data for real-time analysis.

Acknowledgements This work was supported in part by the Open Project Program of the State Key Lab of CAD&CG (Grant No. A2013), Zhejiang University, the National Natural Science Foundation of China (Grant No. 61802283), the Fundamental Research Funds for the Central Universities in China, and the Natural Science Foundation of Shanghai, China (Grant No. 20ZR1461500).

References

Aigner W, Miksch S, Schumann H, Tominski C (2011) Visualization of time-oriented data. Springer, Berlin

Andrienko G, Andrienko N, Bak P, Keim D, Wrobel S (2013) Visual analytics of movement. Springer, Berlin

Andrienko G, Andrienko N, Fuchs G, Wood J (2017) Revealing patterns and trends of mass mobility through spatial and temporal abstraction of origin-destination movement data. IEEE Trans Visual Comput Graphics 23(9):2120–2136

Bakıcı T, Almirall E, Wareham J (2013) A smart city initiative: the case of barcelona. Journal of the Knowledge Economy 4(2):135–148

Cheng Y (1995) Mean shift, mode seeking, and clustering. IEEE Trans Pattern Anal Mach Intell 17(8):790–799

Demšar U, Reades J, Manley E, Batty M (2017) Revisiting the past: replicating fifty-year-old flow analysis using contemporary taxi flow data. Ann Am Assoc Geogr: 1–18

Deng JS, Wang K, Hong Y, Qi JG (2009) Spatio-temporal dynamics and evolution of land use change and landscape pattern in response to rapid urbanization. Landsc Urban Plan 92(3):187–198

Dunteman GH (1989) Principal components analysis, vol 69. Sage, Thousand Oaks

Goddard JB (1970) Functional regions within the city centre: a study by factor analysis of taxi flows in Central London. Trans Inst Br Geogr 49:161–182

Kim J, He Y, Park H (2014) Algorithms for nonnegative matrix and tensor factorizations: a unified view based on block coordinate descent framework. J Glob Optim 58(2):285–319

Kloeckl K, Senn O, Di Lorenzo G, Ratti C (2011) Live singapore!-an urban platform for real-time data to program the city. Comput Urban Plan Urban Manag CUPUM 4

Kraft S, Marada M (2017) Delimitation of functional transport regions: understanding the transport flows patterns at the micro-regional level. Geografiska Annaler Ser B Hum Geogr 99(1):79–93

Kuhn HW (1955) The Hungarian method for the assignment problem. Nav Res Logist (NRL) 2(1–2):83–97

Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401(6755):788–791

Robertson G, Fernandez R, Fisher D, Lee B, Stasko J (2008) Effectiveness of animation in trend visualization. IEEE Trans Vis Comput Graph 14(6):1325–1332

Rosling H (2009) Gapminder, vol 91. GapMinder Foundation, Stockholm

Seto KC, Fragkias M (2005) Quantifying spatiotemporal patterns of urban land-use change in four cities of China with time series landscape metrics. Landsc Ecol 20(7):871–888

Su K, Li J, Fu H (2011) Smart city and the applications. In: International conference on electronics, communications and control. IEEE, pp 1028–1031

Tominski C, Schumann H, Andrienko G, Andrienko N (2012) Stacking-based visualization of trajectory attribute data. IEEE Trans Vis Comput Graph 18(12):2565–2574

van den Elzen S, Holten D, Blaas J, van Wijk JJ (2016) Reducing snapshots to points: a visual analytics approach to dynamic network exploration. IEEE Trans Vis Comput Graph 22(1):1–10

Wakamiya S, Lee R, Kawai Y, Sumiya K (2015) Twitter-based urban area characterization by non-negative matrix factorization. In: *Proceedings of the international conference on big data applications and services*. ACM, pp 128–135

Wang J, Gao F, Cui P, Li C, Xiong Z (2014) Discovering urban spatio-temporal structure from time-evolving traffic networks. In: *Asia-pacific web conference*. Springer, pp 93–104

Wu W, Zheng Y, Qu H, Chen W, Gröller E, Ni LM (2014) Boundaryseer: visual analysis of 2d boundary changes. In: *IEEE conference on visual analytics science and technology*, pp 143–152

Wu W, Xu J, Zeng H, Zheng Y, Qu H, Ni B, Yuan M, Ni LM (2016) Telcovis: visual exploration of co-occurrence in urban human mobility based on telco data. IEEE Trans Vis Comput Graph 22(1):935–944

Wu W, Zheng Y, Cao N, Zeng H, Ni B, Qu H, Ni LM (2017) Mobiseg: interactive region segmentation using heterogeneous mobility data. In: *IEEE pacific visualization symposium*

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

Zhang K, Sun D, Shen S, Zhu Y (2017) Analyzing spatiotemporal congestion pattern on urban roads based on taxi gps data. J Transp Land Use 10(1):675–694

Zhao Y, Luo X, Lin X, Wang H, Kui X, Zhou F, Wang J, Chen Y, Chen W (2019) Visual analytics for electromagnetic situation awareness in radio monitoring and management. IEEE Trans Vis Comput Graph 26(1):590–600

Zheng Y, Wu W, Chen Y, Qu H, Ni LM (2016a) Visual analytics in urban computing: an overview. IEEE Trans Big Data 2(3):276–296

Zheng Y, Wu W, Zeng H, Cao N, Qu H, Yuan M, Zeng J, Ni LM (2016b) Telcoflow: visual exploration of collective behaviors based on telco data. In: *IEEE international conference on big data*, pp 843–852