J Vis (2019) 22:1193–1208
https://doi.org/10.1007/s12650-019-00600-6

REGULAR PAPER

Guodao Sun · Baofeng Chang · Lin Zhu · Hao Wu · Kai Zheng · Ronghua Liang

TZVis: visual analysis of bicycle data for traffic zone division

Received: 7 July 2019 / Accepted: 19 August 2019 / Published online: 10 September 2019

Abstract Traffic zone is a geographical unit which is proposed to reduce the complexity of traffic control and management systems in modern urban road networks. Traffic zone could be generated by different division approaches (e.g., rule-based approaches, data-driven approaches), which may lead to different diverse division results and characteristics. The traditional division approaches usually rely on human experience or static constraints (e.g., political boundaries or natural environment) while neglecting the traveling pattern among different regions. Data-driven approaches are also utilized in recent years; however, domain knowledge and rules are not well integrated and represented in such systems. This paper proposes a new approach that both combines patterns that is hidden in the data and user-defined rules. It not only fully employs the traffic hot spots and traveling patterns among different regions, but also considers the realistic constraints such as road network. We also design TZVis, a visual analysis system that fully integrates with the division approach. The system allows users to generate multiple traffic zone division schemes under different requirements, analyze the schemes in multiple views and select an optimal traffic zone scheme. It provides a map overview to show the traffic zone division result, a matrix view to show the traffic relationship of zones, a parallel coordinates plot for displaying the specific traffic records, a LineUp view to display all the division results and a radar chart to show the features of each zones. We further test our division approach and demonstrate the usefulness of our system based on the shared bicycle dataset.

Keywords Traffic zone division · Shared bicycle data · Road network constraint · Visual analysis · Interactive visualization

---

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12650-019-00600-6) contains supplementary material, which is available to authorized users.

G. Sun (✉) · B. Chang · L. Zhu · H. Wu · R. Liang (✉)
Zhejiang University of Technology, Hangzhou, China
E-mail: godoor.sun@gmail.com

R. Liang
E-mail: rhliang@zjut.edu.cn

B. Chang
E-mail: baofeng.chang@foxmail.com

L. Zhu
E-mail: zhulin3695@gmail.com

H. Wu
E-mail: haowu919@gmail.com

K. Zheng
Hangzhou Comprehensive Transportations Information Center, Hangzhou, China
E-mail: zhengk_hz@zjt.gov.cn

1 Introduction

The traffic zone can greatly reduce the complexity of urban traffic system such as traffic flow analysis, management and control. The problems that seriously restrict urban development such as traffic congestion and large-scale congestion in a city can be relaxed by a well-defined traffic zone division scheme. Many cities made their own traffic zone division scheme; however, the division scheme may differ with respect to different scenarios. A single division scheme may not apply to all situations. Moreover, our investigation and research reveals that traffic zone may not play a key role in understanding urban movements. The main reason behind this issue is due to the inflexible division scheme (e.g., merely depends on political boundaries), which hinders its usage in real-world scenarios. Therefore, it is vital to build a reasonable traffic zone division scheme for urban administrators who want to utilize the scheme for better urban management.

Traffic zone division scheme could be mainly divided into two approaches, namely traditional division approach and data-driven division approach. The traditional division approach generates the traffic zones based on human experience and many static conditions in the urban road network, such as roads, rivers and railways. However, such traffic zone scheme is not verified by real-world movement data and may not well reflect real urban behavior. Thus, rationality of the division scheme may be questioned. The data-driven division approach generates traffic zone scheme based on various types of data of urban movements, such as population movement data or taxi Global Positioning System (GPS) data. This approach does not reflect domain knowledge, and the shape of most zones generated is complex and not easy to be recognized by people.

To help urban administrators solve urban traffic problems by employing traffic zone concept, we aim to build a reasonable and accurate traffic zone division scheme.

We studied and summarized the characteristics of both traditional and data-driven division approaches. In this paper, we propose a new traffic zone division approach that combines the features of the traditional and data-driven division approaches. It generates traffic zone division schemes not only based on the traffic hot spots and hidden traveling patterns among different regions, it also considers the realistic and static constraints such as road network. Our research employs spatiotemporal origin–destination (OD) data which is extracted from the bicycle sharing record dataset to test our division approach and further conduct case study. We propose four indicators for each traffic zone and five indicators for each division scheme for better understanding of the approaches

We design a visual analytics system named TZVis to help users select optimal division scheme according to their interests. The system provides a map overview to show the traffic zone division result, a matrix view to show the traffic relationship of zones, a parallel coordinates plot for displaying the specific traffic records, a LineUp view to display all the division results and a radar chart to show the features of each zones. In the TZVis, users can interactively generate multiple division schemes and visually explore and analyze different division schemes. TZVis allows users to evaluate the division scheme, analyze the traffic relationship between different zones and explore the spatiotemporal characteristics of each zone.

The key contributions of this work are as follows:

- We propose a new traffic zone division approach that integrates the features of traditional and data-driven division approaches.
- We design TZVis, a visual analytics system that helps users analyze division schemes and support decision-making.
- We offer empirical findings based on real-world data to test the effectiveness of our zone and scheme indicators as well as the visual analytics system.

2 Background and related works

This section reviews the related works on the traditional, data-driven traffic zone division approaches and visualization about the traffic zone.

2.1 Traditional division approach

Allsop (1974) and Whitson et al. (1973) are the first group of people to study the traffic zone. The early division approaches (traditional division approaches) divide traffic zones according to real physical characteristics such as mountains, rivers, road networks, administrative boundaries and other divisional principles. Traditional division approaches use these physical characteristics as the regional boundaries of traffic zones. Many urban administrators have been using the traditional division approaches to divide traffic zones. However, the traditional division approaches have limitations such as identifying the function of areas imprecisely, lacking the consideration of the relationship between human social activity and urban functional regions, reflecting the impact of human activities on regional function inaccurately.

Although the traffic division scheme obtained by traditional division approaches is more in line with people's intuitive cognition, since the traditional division approaches rely a lot on human experience. However, the result is not verified by real urban movement data, its rationality may be questioned by domain experts. In addition, as the traffic congestion situation becomes more and more serious, the traditional division approaches can no longer meet the demands, thus the research direction gradually shifts to the employment of data-driven division approaches.

2.2 Data-driven division approach

With the generating of massive traffic data from mobile devices (e.g., mobile phone or wearable device), monitoring data, the GPS data of public transportation (e.g., taxis, buses and shared bicycles), data-driven traffic zones division approaches are becoming more popular in many urban planning tasks. The massive mobile data reflect realistic needs and results are more authentic. As a result, a large number of data-driven division approaches which are used in different data have emerged in recent years. For example, Dong et al. (2015) proposed an integrated traffic zone division approaches for mobile communication data development, which has the advantages of high coverage, cost effectiveness and real time. Martínez et al. (2009) proposed a new methodology and algorithm for the definition of traffic zones embedded in geographical information systems software, improved the base algorithm with several local algorithms and comprehensively analyzed the obtained results. You et al. (1998) proposed a method for improving traffic zone division based on GIS. They analyzed and evaluated their method through spatial data analyses. Chao-qun et al. (2007) proposed a dynamic traffic zone division scheme based on game theory. It introduced the conception of traffic zone core based on traffic similarity and took each section node in the road network as a rational game participant, who seeks the most similarity payoff between itself and the traffic zone. Xing et al. (2014) developed a comprehensive approach of traffic zone division on mobile billing data that the land use information obtained from phone call volume and commuting volume as well as spatial obstacles represented as Voronoi distance is taken into account in similarity measurements.

The traffic zone division results from the data-driven division approaches are based on real-world data; however, one limitation is that the shape of each zone may be too complicated and unaccountable. Thus, traditional rule-based division scheme should be further incorporated.

2.3 Area-based traffic visualization

In recent years, more and more scholars have used visual methods to analyze traffic because the visualization is convenient and effective for abstracting traffic data. Sun et al. (2018) proposed "DiffusionInsighter," a visual analysis system for interactive visualization, analysis and understanding of traffic flow as well as traffic diffusion patterns. They also presented a novel visualization technique called route-zooming that can embed spatiotemporal information into a map seamlessly for occlusion-free visualization Sun et al. (2017). Adrienko and Adrienko (2010) proposed an interactive visualization method, which realizes spatial division based on the spatial distribution of trajectory feature points, and visualizes it through flow maps and transformation matrices. Zheng et al. (2015) visualized the mobile data by integrating visualization schemes of three visualization modules, presenting clear features from the starting point to the destination, and achieving comparative analysis of the two locations effectively. At the same time, these visualizations combine multiple visualizations for visual analysis based on multiple views (Sun et al. 2018, 2017; Adrienko and Adrienko 2010; Zheng et al. 2015). For the visualization schemes of the traffic zones, it is very important to consider trustworthiness. Trustworthy visual analytics with effective uncertainty modeling and visualization enables users to explicitly consider the uncertainty information, so that informed decisions can be made (Sun et al. 2013). Proposing technical indicators could be a persuasive way to verify the effectiveness of the visualization system.

We propose a novel division approach for generating traffic zone based on real traffic data as well as the constraints of road network. We also design TZVis, a visualization system that helps users generate and analyze traffic zone division schemes. Four regional indicators and five scheme indicators are further proposed to display zone and scheme features.

3 Workflow

This section describes the pipeline of our TZVis. As shown in Fig. 1, the system is divided into four main parts, namely "Data Preprocessing," "Scheme Generation," "Indicators Extraction" and "Visual Analysis."

The Data Preprocessing part extracts spatiotemporal OD data from raw shared bicycle data and performs data filtering and cleaning. At the same time, we collect and preprocess the road network data as the constraint in the process of traffic zone division.

The Scheme Generation part proposes a new method to divide the traffic zone based on the hot spots of OD data and the constraint of road network. Boundary gaps are further fused in this part to help determine the final traffic zone schemes.

In the Indicators Extraction part, we come up with four regional indicators including regional attraction, influence, openness and stability to help users obtain insights of each traffic zone. We also propose five scheme indicators including scheme average attraction, influence, openness, stability and usability to help users understand traffic zone division schemes.

In the Visual Analysis part, we design a visual analysis system to display the traffic division zone schemes. This system could help users to analyze the traffic zone and division schemes through indicators and multiple views. Users can define the generation parameters to generate division schemes during the analysis process.

4 Data description

This section introduces the data used in our work and indicators of traffic zones and division schemes.

4.1 Shared bicycle data

The traffic division approach is tested on spatiotemporal OD data extracted from the shared bicycle data. We employ shared bicycle data because of its three advantages, namely **large number of driving records**, **short driving distance** and **rich driving routes**. The large number of driving records can make the regional division more fully validated. The short driving distance is beneficial to make a small and accurate traffic zone scheme. The complex and varied routes cover as many urban road network as possible. Compared with the single route of the bus and the long driving distance of the taxi, the shared bicycle record data is more conducive to our work.

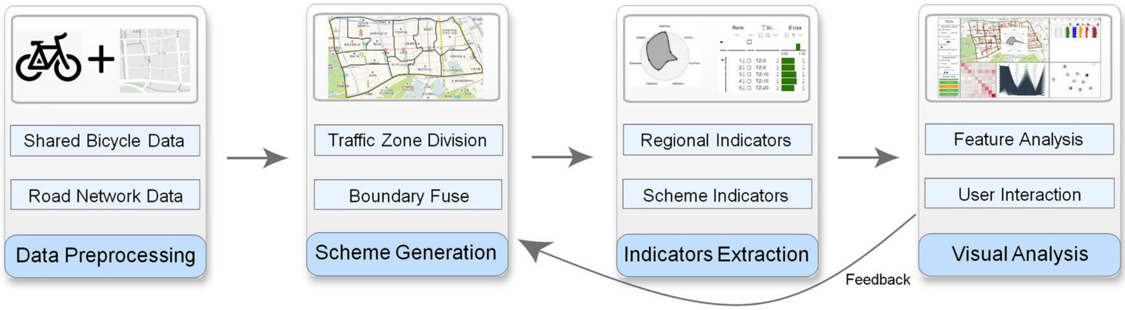


Fig. 1 The pipeline of our visual analysis system, which includes Data Preprocessing, Scheme Generation, Indicators Extraction and Visual Analysis

In this work, we employed the shared bicycle dataset which contains more than 10 million records. In one raw shared bicycle data records, the most vital parameters are [DepartTime, DepartLoc, ArriveLoc, DBTime]. "DepartTime" is the unlocking time of the shared bicycle, while the "ArriveTime" is the locked time. "DeapartLoc" is the unlocked location coordinates, and "ArriveLoc" is the locked location coordinates.

These raw shared bicycle data were preprocessed to improve system efficiency and performance of the query function. Preprocessing of data included data quality analysis, cleaning and conversion, and the rules used in the preprocessing stage are listed as follows.

- We only reserve the following data fields for this study, including CompanyID, DepartTime, ArriveTime, DeapartLoc and ArriveLoc. The other data fields were removed during preprocessing.
- We only reserve long-time records. Short-time (e.g., < 1 min) records are likely to be generated by test cases or data noise and are removed to reduce disturbance to the traffic zones division.

4.2 Road network data

Our division approach considers realistic constraints such as road network. The road network data consist of multiple roads and crossings. We assume that a road is a sequence of points. Each point in the sequence is a point on the road, and multiple points form a road in order. Crossings are where the roads cross and also can be represented by points. Keeping only the turning points not only preserves the road network information, but also reduces the storage pressure of the road network data. Reducing the points of the road network does not affect the intersection.

In our work, we only collect the major roads and the crossings where a major road cross another major road. We assume that the major road refers to important roads in commercial areas or residential areas where traffic is relatively large and important, such as *Tianmushan* Road or *Moganshan* Road in Hangzhou. The constraints from the major roads are strong, while from subordinate roads such as sidewalks or pedestrian street is weak. Considering the whole roads leads the constraints to be extreme complicated. In order to avoid this problem, we collected and extracted road network data according to the following rules.

- Only major roads are remained while the minor roads are neglected.
- Only turning points of a road are reserved, while other unimportant points are removed.
- Only crossings among major roads are reserved.

4.3 Regional indicator

Each traffic zone has real traveling information such as inflow data (V_{in}), outflow data (V_{out}) and self-flow data (V_{self}). In a specific period, a zone's inflow data are computed by all the shared bicycles that arrive at. Its outflow data are all the shared bicycles that leave from the zone, and its self-flow data are the one that depart from and arrive at itself. Regional traveling patterns can be obtained from analyzing these traveling information.

To help users analyze and evaluate the patterns in the traffic zone, we propose four regional indicators that extracted from raw traffic flow data. These regional indicators are regional attraction, influence, openness and stability.

4.3.1 Attraction

We assume that the product of the flow and distance between the two regions is the mutual attraction between the two regions. The attraction of a region is the numerical addition of the attraction of this region to all other regions. R^{m}_{attr} is the regional attraction of the zone m, and it is proportional to the inflow of zone m and the distance to zone m. The bigger R^{m}_{attr} is, the more attractive the zone m is. R^{m}_{attr} is calculated by formula $R^{m}_{\mathrm{attr}} = \sum_{i=1}^{n} V^{im}_{\mathrm{in}} * \mathrm{Dis}_{im}$, where V^{im}_{in} is the traffic inflow from zone i to zone m, Dis_{im} is the Manhattan distance from zone i to zone m.

4.3.2 Influence

We set the product of the flow and distance between the two regions as the influence between the two regions. The influence of one region is the sum of the influences between all other regions. Regional influence R_{inf}^{m} of the zone m is proportional to the flow about zone m and the distance to zone m. It is calculated by formula $R_{\text{inf}}^{m} = \sum_{i=1}^{n} \left(V_{\text{in}}^{im} + V_{\text{out}}^{mi}\right) * Dis_{mi}$, where V_{in}^{im} is the traffic inflow from zone i to zone m, V_{out}^{mi} is the traffic outflow from zone m to zone i, Dis_{mi} is the Manhattan distance from zone m to zone i.

4.3.3 Openness

We present the openness of a region to account for the proportion of traffic in this region and other regions to its total traffic. The larger the ratio, the higher the openness of the region. The higher the degree of openness, the closer the relationship is to the outside world. Regional openness R_{open}^{m} of the zone m is proportional to the sum of inflow and outflow, and inversely proportional to the total flow. It is calculated by formula $R_{\text{open}}^{m} = \frac{V_{\text{in}}^{m} + V_{\text{out}}^{m}}{V_{\text{in}}^{m} + V_{\text{out}}^{m} + V_{\text{self}}^{m}}$, where V_{in}^{m}, V_{out}^{m} and V_{self}^{m} are inflow, outflow and self-flow of zone m, respectively.

4.3.4 Stability

Regional stability R_{sta}^{m} of zone m is the degree of its stability. We regard the $\frac{\left|V_{\text{in}}^{m} - V_{\text{out}}^{m}\right|}{V_{\text{in}}^{m} + V_{\text{out}}^{m} + V_{self}^{m}}$ as a variable. The bigger this variable, the region is easier to have excess or too little shared bicycles in a short period. In another word, this region is unstable. So the regional stability is calculated by formula $R_{\text{sta}}^{m} = 1 - \frac{\left|V_{\text{in}}^{m} - V_{\text{out}}^{m}\right|}{V_{\text{in}}^{m} + V_{\text{out}}^{m} + V_{\text{self}}^{m}}$, where V_{in}^{m}, V_{out}^{m} and V_{self}^{m} are respectively the inflow, outflow and self-flow of the zone m.

We also propose five scheme indicators to evaluate the division schemes. The scheme indicators are average regional attraction $\text{Ave}_{\text{attr}} = \frac{\sum_{i=1}^{n} R_{\text{attr}}^{i}}{n}$, average regional influence $\text{Ave}_{\text{inf}} = \frac{\sum_{i=1}^{n} R_{\text{inf}}^{i}}{n}$, average regional openness $\text{Ave}_{\text{open}} = \frac{\sum_{i=1}^{n} R_{\text{open}}^{i}}{n}$, average regional stability $\text{Ave}_{\text{sta}} = \frac{\sum_{i=1}^{n} R_{\text{sta}}^{i}}{n}$ and a new indicator scheme usability $S_{\text{usa}} = \frac{\text{Ave}_{\text{sta}}}{\text{Ave}_{\text{open}}}$ of a division scheme. The first four scheme indicators are the average of all zones in a division scheme and the last indicator, scheme usability S_{usa}, demonstrates the usability of a division scheme. The average stability of a solution should be as high as possible, and the degree of average openness should be as small as possible.

5 Traffic zones generation

This section introduces the features analysis about traditional division approaches and data-driven division approaches, our division method and boundary fuse after dividing.

5.1 Features analysis

Before conducing traffic zone division, we first summarize and analyze the features of traditional division approaches and the data-driven division approaches.

In the traditional division approaches, the person in charge of division usually divides the same type of functional area into a traffic community, or tries to use the geographical barriers such as railways, rivers and roads as the boundary of the traffic zone. If a planner plans a traffic zone according to the traditional division approaches, each zone will easily understood by people's intuitive perception. Most traffic zone will have rectangular or nearly rectangular shape features and be shown in front of people in a very neat form.

In the data-driven division approaches, the planner conducts the traffic zone division according to the requirements and the data. The division result will reflect the data implicitly. The designed program is rigid and tries to follow users' requirement as much as possible. This also results in the situation that most traffic zones which are from data-driven division approaches have very complicated shapes and be hard to understand.

In summary, the features of traditional division approaches enable people to understand traffic zone easily, while data-driven division approaches allow the traffic divided to be divided from a data perspective. We would like that the traffic zone has both cognitive level comprehensibility and data level rationality.

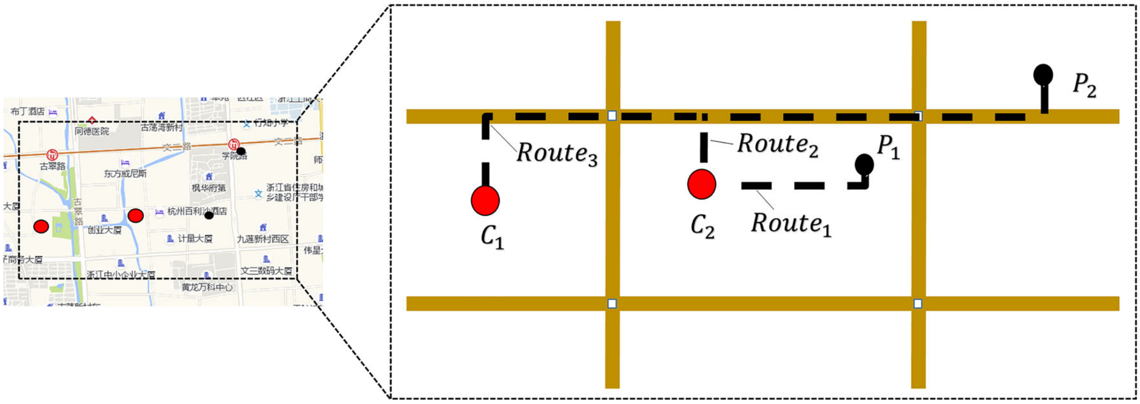


Fig. 2 Meta-centers (red) and grid points (black) in the road network

5.2 Traffic zone division

The existing traditional division approaches and data-driven division approaches have limitations in generating a traffic zone scheme for realistic traffic system. In order to solve these limitations, after discussion with domain experts, we propose a novel traffic zone division approach.

5.2.1 Extract traffic hot spots

Traffic hot spots are the areas which have heavy traffic flow. Extracting them is helpful for urban planning and road network division. In our research, a K-means clustering approach is employed on the spatiotemporal OD data that is preprocessed from the shared bicycle dataset. After clustering, we assume that the hot spots are the centers of the clusters. These hot spots are important for the traffic zones generated by our division approach. A zone has a hot spot is named “meta-center” (see Fig. 2).

5.2.2 Analyze the traffic accessibility under constraints of the road network

Traffic accessibility is the convenience level of reaching a certain area in a traffic system. We employ the concept of traffic accessibility during the division process. Under constraints of the realistic road network, we assume that each meta-center has an area that every point in this area can reach the meta-center faster than the point out this area. This area is defined as the traffic zone of meta-center.

During division, we divide the area into multiple small grids with a side length of 50 meters and further analyze the traffic accessibility among the meta-centers and grid points. For example, we choose two meta-centers and two grid points (see Fig. 2) to introduce the division. We calculate the constrained distance from every meta-center to a grid point in the road network and assign the point to a meta-point according the shortest distance.

In the road network, there are routes from meta-centers to grid points. Each routes will intersect roads or will not. For example, route between meta-center C_2 and grid point P_1 does not cross the road, while the route between meta-center C_2 and grid point P_2 crosses the road.

It is found that each grid point has one of the following cases. The first case is that a grid point can reach the meta-center without crossing the road such as the one between meta-center C_2 and grid point P_1 in Fig. 2. In this case, we could calculate the Manhattan distance between the meta-center C_2 and grid point P_1 without still considering impact of C_1. we consider grid point P_1 is located in the traffic zone of meta-center C_2. The second case is that a grid point cannot reach any meta-center without crossing the road such grid point P_2 in Fig. 2. In this case, we should calculate the distances from grid point P_2 to meta-center on the basis of C_2 the $Route_2$ and to the meta-center C_1 on the basis of $Route_3$. We consider grid point P_2 is located in the traffic zone of meta-center C_2 according to the distances.

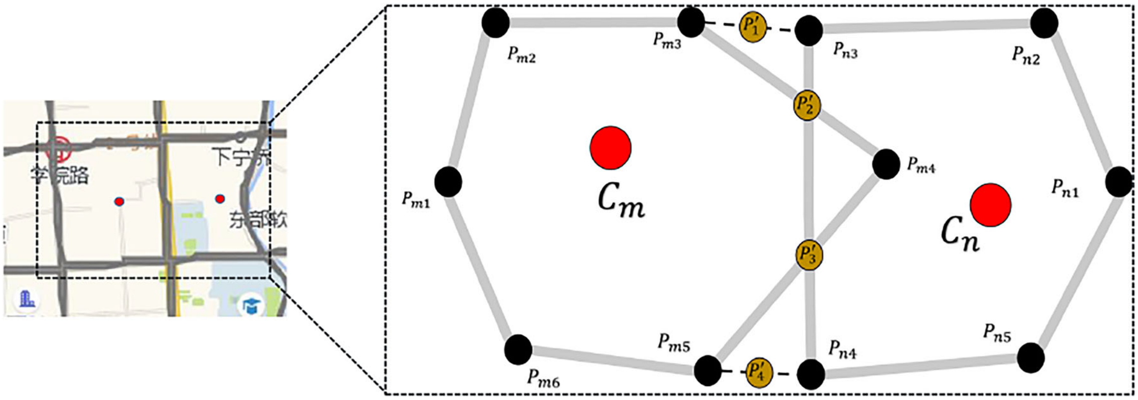


Fig. 3 Adjusting boundaries during the division to eliminate insignificant intersections and blank gaps

5.2.3 Calculating the concave shell

After finishing the above steps, each meta-center has a set of points that the points are in the traffic zone of meta-center. Then, we calculate the concave shell of the each point set to determine the boundary of the traffic zone for each meta-center. The concave shell algorithm we employed was inspired by Park and Oh (2012) and Rosen et al. (2014).

5.3 Boundary fuse

After calculating the concave shell and regarding it as the boundary of the meta-center, we find out that boundaries of two adjacent traffic cells may be different. There are still small intersections and blank areas (see Fig. 3) on adjacent boundaries because we employ the concave shell algorithm on the grid points to generate the boundaries. We propose a method to adjust the boundaries.

As shown in Fig. 3, the boundary is in form of a set of points in order. The traffic zone boundary of meta-center C_m is $[P_{m1}, P_{m2}, P_{m3}, P_{m4}, P_{m5}, P_{m6}]$, while $[P_{n1}, P_{n2}, P_{n3}, P_{n4}, P_{n5}]$ is the traffic zone boundary of meta-center C_n.

The approach to adjust the boundary is elaborated as follows. *Remove boundary points in other zones* Firstly, there may be several boundary points that are inside the other traffic zones. For example in Fig. 3, P_{m4} is a boundary point of *Zone* C_m, while it is inside the boundary of *Zone* C_n. We calculate the intersection points P'_2 and P'_3 and regard them as new boundary points of these two zone. At the same time, the P_{m4} should be removed in the boundary of *Zone* C_m. As a result, the boundary of *Zone* C_m is changed to $[P_{m1}, P_{m2}, P_{m3}, P'_1, P'_2, P_{m5}, P_{m6}]$ and the boundary of *Zone* C_n becomes $[P_{n1}, P_{n2}, P_{n3}, P'_1, P'_2, P_{n4}, P_{n5}]$.

5.3.1 Locate middle point of two near points

There exists a boundary point subset near to other zone boundaries. Taking *Zone* C_m as an example (see Fig. 3), we traverse boundary points to find the first point (P_{m3}) which is near to the boundary of *Zone* C_n, and traversal boundary points reversely to find the last point (P_{m5}) which is near to the boundary of *Zone* C_n. The boundary point subset $[P_{m3}, P'_1, P'_2, P_{m5}]$ should be adjusted. Then, we find the nearest point in the boundary of *Zone* C_n for each point in this boundary point subset and calculate the middle point as a new boundary point while remove these two near points. For example, P_{m3} and P_{n3} are the nearest points to each other and the middle point P'_1 is a new boundary point while removing P_{m3} and P_{n3}.

In the end, the boundary of *Zone* C_m is changed to $[P_{m1}, P_{m2}, P'_1, P'_2, P'_3, P'_4, P_{m6}]$ and the boundary of *Zone* C_n becomes $[P_{n1}, P_{n2}, P'_1, P'_2, P'_3, P'_4, P_{n5}]$.

6 Visual design

This section describes the user requirements collected from our domain experts and derives a set of design goals. After our discussion, we introduce our visualization design based on the design goals.

6.1 User requirements

We collaborated with our domain experts for this study. We aimed to propose a traffic zone division approach to relax the disadvantages of traditional division approaches and data-driven approaches. With domain experts, we tried to employ visual analysis for the traffic zone scheme generated by our division approach and we derived research questions from our domain experts, which are described as follows.

Q1 What is the result of the traffic zone division scheme? What is the shape and location of each traffic zone?
Q2 What is the traveling flow relationship between the traffic zones over a period of time? How does the relationship vary over the time?
Q3 How do the spatiotemporal feature of the traffic flow between traffic zones evolve over a period of time?
Q4 What is the difference between the traffic zone division schemes? How to display the difference between the division schemes effectively?
Q5 What is the similarity between traffic zones? What is the connection between similar traffic zones?

These requirements helped us derive the appropriate design principles and make judicious decisions on our visual design.

6.2 Design goals

We define following design principles to guide our visual design based on user needs:

G1 *Show the traffic zone division scheme* The final design should provide a map view to show the traffic zone division scheme for users. It helps users recognize the shape and geographical features of each traffic zone (**Q1**) and analyze the regional information from a cognitive perspective (**Q1** to **Q5**).
G2 *Encode the traffic flow* The design should effectively encode the traffic flow to help users analyze the specific traffic flow and relationships between the traffic zones (**Q2** and **Q3**). Encoding traffic flow can help users analyze the regional features, difference and similarity specifically (**Q2** to **Q5**).
G3 *Display relationships between traffic zones* The design should provide a view to help users analyze traffic flow relationships between traffic zones (**Q2**). It is also helpful for analyzing the spatial and temporal features of traffic flows (**Q3**).
G4 *Extract and display the regional features* The design should extract the regional features or indicators from the traffic flow and display them to users intuitively (**Q2**).
G5 *Calculate the similarity between traffic zones* The design should calculate the similarity (**Q5**) between traffic zones and show it to users to help them analyze the division schemes (**Q4**).
G6 *Provide a contradistinction of division schemes* The design should provide a view that shows the features of divisions (**Q4**) to help users select the optimal division scheme for the real traffic system.

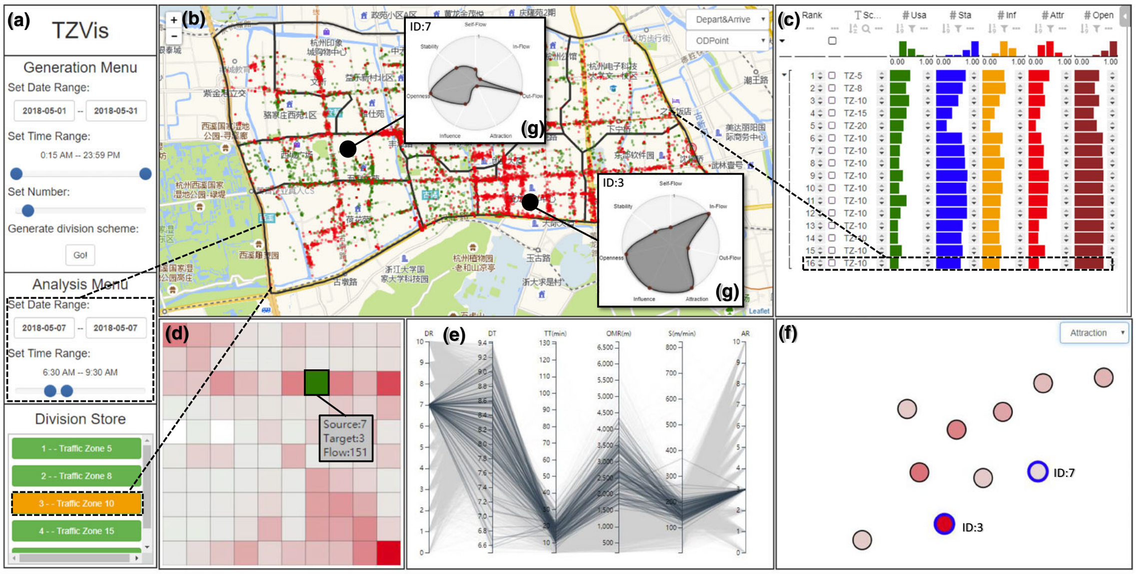
6.3 Visualization techniques

This section introduces our interactive visualization techniques that are designed based on the aforementioned design goals.

6.3.1 Traffic zone visualization

We designed the TZVis, a visual analysis system for users to analyze the traffic division scheme. TZVis provides several coordinated views such as a map overview to display the traffic zone and shared bicycle record on the map, a matrix view to show the traveling flow volume among traffic zones, a parallel coordinates plot to show the shared bicycle records in detail, a LineUp view to show the features of all division schemes, a radar chart to show the features of each traffic zone and a t-SNE plot to show the similarity between the traffic zones. We introduce these visualizations views in the following contents.

Map overview TZVis provides a map view (Fig. 4b) to display the results of traffic zone division and the spatiotemporal OD data extracted from traffic flow data, which can help users to obtain an overview of the traffic zone division scheme. The regional boundaries on the map help users recognize the shape and



Set Date Range: 2018-05-01 -- 2018-05-31
Set Time Range: 0:15 AM -- 23:59 PM
Set Number:
Generate division scheme: Go!

Set Date Range: 2018-05-07 -- 2018-05-07
Set Time Range: 6:30 AM -- 9:30 AM

1 - - Traffic Zone 5
2 - - Traffic Zone 8
3 - - Traffic Zone 10
4 - - Traffic Zone 15

Source:7
Target:3
Flow:151

Fig. 4 Analyze the traffic features of traffic zones based on the shared bicycle data of during 6:30–9:30, May 7th, 2018 in TZVis: **a** a menu bar includes "Generation Menu," "Analysis Menu" and "Division Store"; **b** a map overview of the division scheme and traffic data; **c** a LineUp view to display the indicators of multiple division schemes based on user-defined data; **d** a matrix chart of the traffic flow information among traffic zones; **e** a parallel coordinate plot of specific traffic flow information between traffic zones; **f** a t-SNE chart presenting similarity of different traffic zones; **g** a radar chart shows the indicators of a traffic zone

geographical location of each traffic zones. In order to help users understand the distribution of traffic flow, TZVis allows users to choose the style (OD points or hot spots) in the map overview. In OD point map (Fig. 4b), the green dot on the map indicates the O-Point in the OD data, and the red dot indicates the D-Point. This can help users analyze the hot spots where the traffic OD data occur and analyze the spatial distribution of traffic flow in the traffic zone.

Matrix view TZVis provides a matrix view (Fig. 4d) to show the traveling flow volume among traffic zones. Each grid in the matrix represents the traveling flow between two traffic zones. The value of traveling flow is encoded by color (**G2**). The redder the color is, the higher the value of the traveling flow is. The column is the source of the traveling flow, while the row is the target of the traveling flow. For example, the grid of Column 4, Row 3 indicates that there are 93 traveling flow records from the *Traffic Zone 4* to the *Traffic Zone 3*. The grid on the main diagonal shows the self-flow of a traffic zone. TZVis allows users to

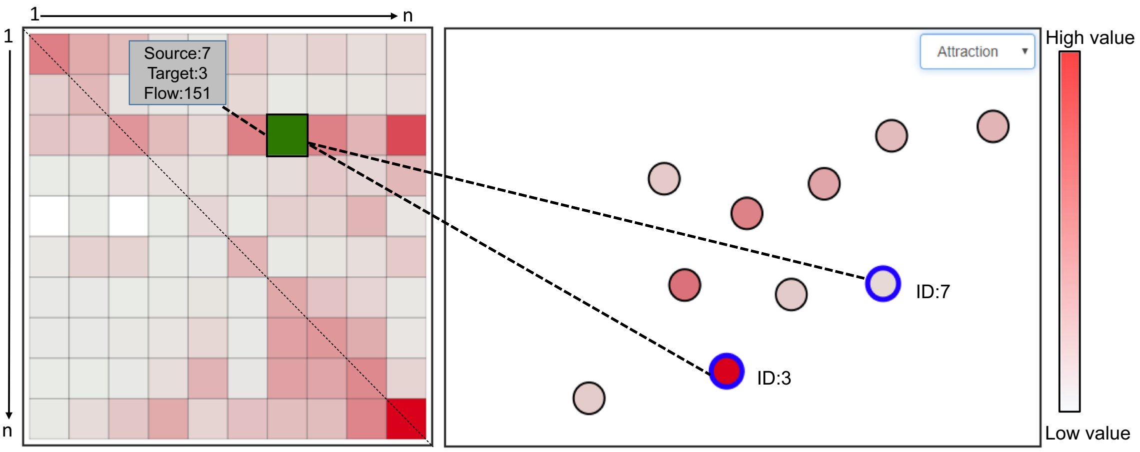
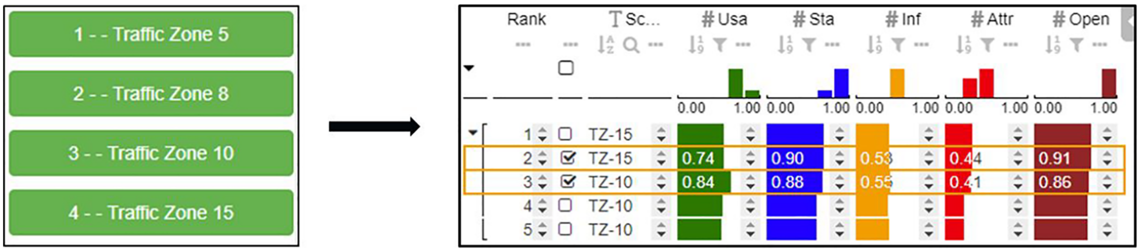


Fig. 5 Matrix view that displays traffic flow between traffic zones (left) and t-SNE plot that shows the similarity and attraction of traffic zones (right)



Rank		Sc...	Usa	Sta	Inf	Attr	Open
1	☐	TZ-15					
2	☑	TZ-15	0.74	0.90	0.53	0.44	0.91
3	☑	TZ-10	0.84	0.88	0.55	0.41	0.86
4	☐	TZ-10					
5	☐	TZ-10					

Fig. 6 LineUp view that shows indicators of division schemes, and each division scheme (left) has their indicators in the LineUp view (right)

analyze the traffic flow relationships and features of traffic zones in the matrix view (**G2** and **G3**) and users could "click" the grid of interest to focus and show relevant information in other views. The matrix view and the t-SNE plot are linked. When a user clicks on a grid on the matrix view, the grid will turn green, and the two traffic cells associated with the grid will be highlighted in the t-SNE view. As shown in Fig. 5, the user clicks on the grid indicating the *Traffic Zone 7* to the *Traffic Zone 3*, and the value of flow is displayed as 151. At the same time, T-SNE plot highlights corresponding zones.

Parallel coordinates plot TZVis provides a parallel coordinates plot (Fig. 4e) to show the shared bicycle records in detail. Users can analyze the time and speed features about traffic flow of the traffic zones (**G4**). In the parallel coordinates plot, each traffic record is encoded as a line (**G2**) that consists the information such as departing region (**DR**), departing time (**DT**), traveling time (**TT**), traveling Manhattan distance (**TMD**), speed (**S**) and arriving region (**AR**). In the parallel coordinates plot, users can filter data according to their interests.

LineUp view We employ a LineUp view Gratzl et al. (2013) to show the difference among multiple division schemes, and we have borrowed the experience of visualization work (Wenchao et al. 2015; Lu et al. 2015, 2017; Liu et al. 2016) about the usage of "LineUp." TZVis provides a LineUp view (Fig. 4c) to show the features of all the division schemes. Users can analyze the difference of division schemes with LineUp view (**G6**). We derive the indicators of the division scheme by calculating the average value of the traffic zone indicators. As shown in Fig. 4c, the indicators are average attraction (Attr), average stability (Sta), average influence (Inf), average openness (Open) and usability (Usa) of the division scheme. All the scheme indicators are introduced in Section 4.3. The left side of Fig. 6 is the traffic cell division scheme, and the right side is the LineUp view that shows the division scheme parameters. In TZVis, the LineUp view attempts to provide indicators of the traffic zone division scheme based on user-defined analysis data. As shown in Fig. 6, "Sc" is the name of the division scheme, and each division has five indicators which are represented by a bar of a different color. The LineUp will highlight user's selected scheme and display the exact indicator value. For the example in Fig. 6, the "*TZ-10*" and "*TZ-15*" schemes are selected and the indicators value are shown such as the usability of "*TZ-10*" is 0.84, while the usability of "*TZ-15*" is 0.74.

Radar chart TZVis provides a radar chart (Fig. 4g) to show the regional indicators of each traffic zone. We want to use a radar chart to show the indicators of each traffic zone so that users can directly get the indicators of the traffic zone where they are interested. As shown in Fig. 4g, the indicators consist major traffic information (e.g., traffic inflow, outflow and self-flow) and the extracted indicators such the stability, attraction, openness and influence.

The t-SNE plot In this view, we would like that through the dimension reduction method, we can see the similarity among different zones. We considered three dimensionality reduction methods what are commonly used by others, such as multidimensional scaling (MDS) (Torgerson 1958) projection, principal component analysis (PCA) (Jegou et al. 2010), and t-distributed stochastic neighbor embedding (t-SNE) (Maaten and Hinton 2008). In comparison, t-SNE can solve the congestion problem of the other two dimensionality reduction methods. If there are hundreds of traffic zones, the other two methods may be not suitable for discovering the cluster patterns between these traffic zones. Thus, we use t-SNE to find the similarity and hidden patterns between zones based on seven attributes such as [V_{in}, V_{out}, V_{self}, R_{attr}, R_{inf}, R_{open}, R_{sta}]. Each node in two-dimensional plane represents a traffic zone, and the t-SNE algorithm arranges similar traffic zones as near as possible. The redder the color of a node is, the larger value it represents. We also have the option to allow users customize the value represented by the color. When a certain area is selected, the border of the circle representing this zone will become thick and blue. For example in Fig. 5, nodes representing *Traffic Zone 3* and *7* are selected, and their boundaries are thickened and turned blue.

6.3.2 User interactions

TZVis supports various advanced and basic interactions such as zooming and highlights.

Show overview first and details-on-demand TZVis displays the traffic zone division scheme and spatiotemporal OD data on the map so that users can obtain an overview quickly. Then, users can zoom in/out the area of interest by mouse and analyze the geographical features of the traffic zone or the spatiotemporal OD data. In addition to interacting with the map overview, users can also manipulate other views by mouse to mine details of interest.

Highlight POI TZVis highlights the related information of the POI to help users explore and analyze traffic zone. For example, when a user is interested in the traffic flow between two traffic zones in the matrix view, the user can click on this grid to be green by the mouse. The matrix view will highlight the grids related to these two traffic zones. At the same time, TZVis will highlight these two traffic zones in the map view, highlight the traffic records between these two zones in the parallel coordinates and also highlight the dots that represented these two traffic zones in the t-SNE plot. Then, the user can analyze the spatial distance through the map view, analyze the temporal features of the traffic flow through the parallel coordinates and understand the similarity through the t-SNE plot.

Customize data of interest TZVis allows users to customize the data of interest and generate traffic zone scheme based on the customized data. When selecting the best traffic zone division scheme, users can customize multiple traffic datasets through the menu panel of the TZVis. After that, TZVis generates multiple traffic zone division schemes for users based on these datasets. Finally, the user analyzes each traffic zone division scheme with the interactive visualization means provided by the TZVis.

7 Case study and evaluation

We used TZVis to conduct experiments and case study to verify the effectiveness of our traffic zone division approach, and we also invited domain experts of urban planning to evaluate our system.

7.1 Case study

In this section, we use TZVis to conduct experiments and case study based on several questions derived from domain experts (Q1-Q5).

We first selected an area of Hangzhou, which is surrounded by four major roads (Zijingang Road, Yuhangtang Road, Moganshan Road and Tianmushan Road). Data collection and preprocessing strictly follow the rules described in Section 4. Regarding the spatiotemporal OD data extracted from the shared bicycle data, we only reserve the data that the OD events occur in this area. Regarding the road network data, we only collect important roads in this area. The horizontal roads include Yuhangtang Road, Wenyi Road, Wener Road, Wensan Road and Tianmushan Road. The longitudinal roads include Zijingang Road, Gudun Road, Fengtan Road, Gucui Road, Xueyuan Road, Jiaogong Road and Moganshan Road. The above data collection and preprocessing steps strictly follow the principles outlined in Sect. 2. Then, we use TZVis to generate five traffic zone division schemes with traffic zone number of 5, 8, 10, 15 and 20.

Based on the above data, we conducted two case studies for the TZVis. The first one is to analyze the traffic features of traffic zones, and the second is to compare and analyze the difference schemes.

Analyze the traffic features of traffic zone As shown in the “Analysis Menu” and “Division” in Fig. 4a, we selected the third option, which divides the area into 10 traffic zones, and we perform visual analysis of this scheme based on the shared bicycle data of during 6:30–9:30, May 7th, 2018. As can be seen from the map overview (Fig. 4b), *Zone 3* has a large amount of inflow and *Zone 7* has a large amount of outflow in the period of this time. Considering that May 7th, 2018, 6:30–9:30 is the peak hour of a working day and

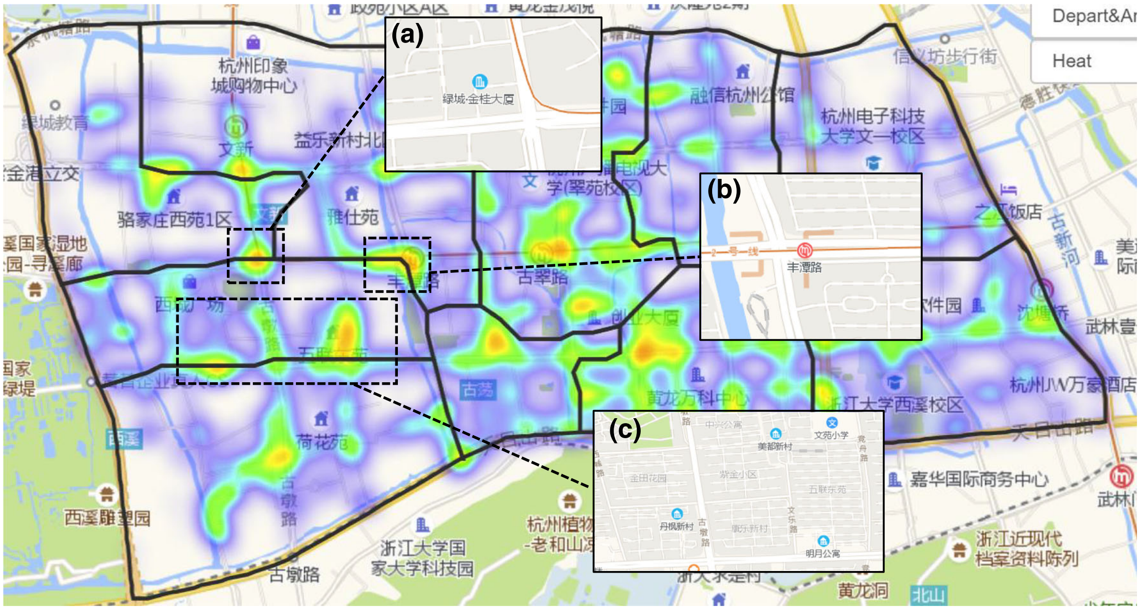


Fig. 7 The heat map of Zone 7 shows that the usage of bicycles assembled in office building (a), underground station (b) and residential area (c)

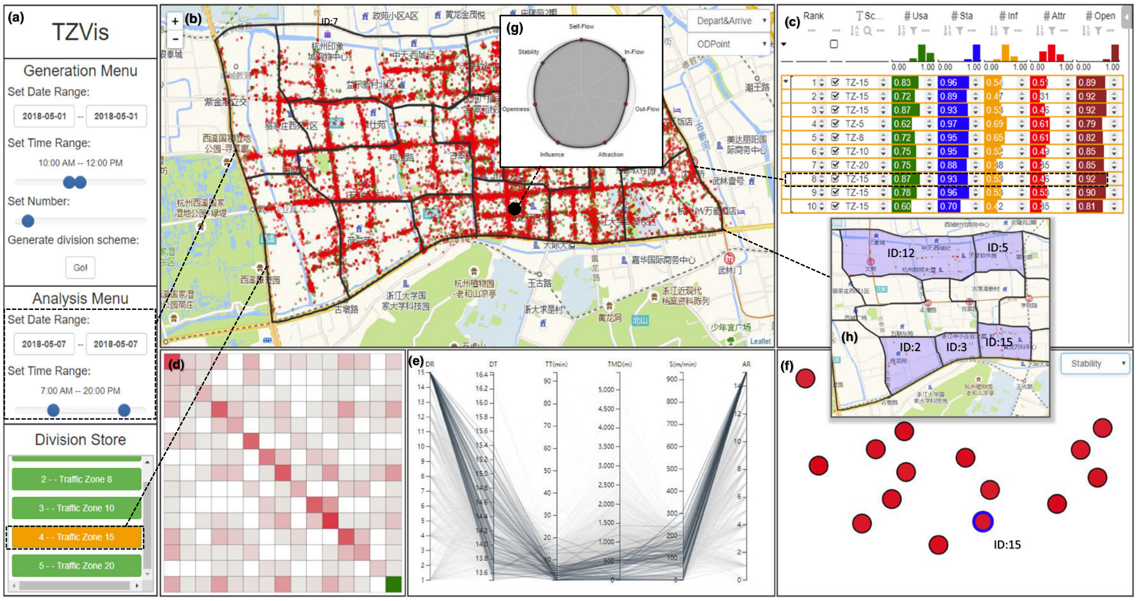
Zone 3 is a typical work area, while *Zone 7* is a typical living area in reality, the traffic flow features of these two zones are consistent with human cognition.

There are two radar charts showing the features of *Zone 3* and *Zone 7*. The inflow, attraction, openness and influence of *Zone 3* are very high, and the stability, self-flow and outflow are relatively small. The outflow and openness of *Zone 7* is very high, the attraction and inflow are almost zero, and other indicators and flows are similar. From Fig. 4g, we can find high openness indicator for both zones, which indicates that there are a large number traffic flows among these two zones and the other zones. These two zones are also very unstable during this time period. If this state is maintained, the supply and demand of the shared bicycle in these two areas will be seriously imbalanced. From Fig. 4f, g, we can get the opposite attraction indicators for these two regions. Based on the analysis of these attributes, we find that the similarity between the two regions is very low. At the same time, we can also find from Fig. 4f that the similarity of the two zones is not high. For the *Zone 3*, during this time period, a large number of people ride the shared bicycle and arrive at this zone from other zones, so the attraction indicator of the *Zone 3* is very high among all zones. Regarding *Zone 7*, there are a large number of people riding bicycles from this zone to other zones, while few people arrive at this zone, so the attraction indicators of *Zone 7* are the lowest. Based on the matrix view (Fig. 4d), we can obtain the traffic flow from *Zone 7* to *Zone 3* is 151. At the same time, based on Fig. 4e, we find that these traffic flows generally occur between 8:00 and 9:00.

As shown in Fig. 7, we found that the origin location of the shared bicycle occurred in the lower half of the zone, and the destination location of the shared bicycle occurred near the upper boundary of the zone. By comparing and analyzing the real situation, we found that the unlocking location about the *Zone 7* is basically concentrated in the residential area or at the entrance of the residential area. The locked-up locations for *Zone 7* are mainly concentrated in two places. One is the Fengtan Road subway station, and the other is the office building. This explains that people not only use shared bicycle as a travel tool, but also integrate travel mode of shared bicycles and subways.

Analyze the difference of schemes In Fig. 8, we analyze different division schemes. As shown in the "Analysis Menu" and "Division" (Fig. 8a), we divided the area into 5, 8, 10, 15 and 20 traffic zones, and conducted visual analysis of these schemes based on the shared bicycle data during 7:00–20:00, May 7th, 2018. After several iterations of data customization and visualization, we found that the availability of the division schemes of the 15 traffic zones is very stable compared to other division schemes.

In the LineUp view Fig. 8c, we calculate the scheme indicators based on the shared bicycle dataset and show them. We have tested ten sets of data for these schemes, and the indicators of "*TZ-15*" are relatively better than other schemes. The availability of the "*TZ-15*" is highest, reaching 0.87, while the usability of



(a) TZVis

Generation Menu

Set Date Range: 2018-05-01 -- 2018-05-31

Set Time Range: 10:00 AM -- 12:00 PM

Set Number:

Generate division scheme: Go!

Analysis Menu

Set Date Range: 2018-05-07 -- 2018-05-07

Set Time Range: 7:00 AM -- 20:00 PM

Division Store

2 - - Traffic Zone 8

3 - - Traffic Zone 10

4 - - Traffic Zone 15

5 - - Traffic Zone 20

(c)

Rank		Sc...	Usa	Sta	Inf	Attr	Open
1	☑	TZ-15	0.83	0.96	0.54	0.51	0.89
2	☑	TZ-15	0.72	0.89	0.47	0.31	0.92
3	☑	TZ-15	0.87	0.93	0.53	0.46	0.92
4	☑	TZ-5	0.62	0.97	0.69	0.61	0.79
5	☑	TZ-8	0.72	0.95	0.65	0.61	0.82
6	☑	TZ-10	0.75	0.95	0.52	0.49	0.85
7	☑	TZ-20	0.75	0.88	0.38	0.35	0.85
8	☑	TZ-15	0.87	0.93	0.53	0.46	0.92
9	☑	TZ-15	0.78	0.96	0.53	0.53	0.90
10	☑	TZ-15	0.60	0.70	0.42	0.35	0.81

Fig. 8 Analysis of traffic zone division scheme *TZ-15* and *Zone 15* based on daytime data, and *TZ-15* has the highest usability, *Zone 15* has stable state

other division schemes did not reach 0.8. Its stability also reaches a high value of 0.93. In the LineUp view, we can get that this division scheme has high usability indicators under multiple custom data.

At the same time, the nodes in the t-SNE (Fig. 8f) view indicate that the nodes are relatively red, indicating that their stability is relatively high. Through the t-SNE view, we found that almost all the zones have relatively high stability. This means that these zones do not have a supply and demand imbalance of the shared bicycle for a long time.

Figure 8g shows the detailed features of the *Zone 15*, Fig. 8d shows traffic flow relationships of the traffic zones, and Fig. 8e shows the self-flow of the *Zone 15*. Referring to the indicators and through visual analysis of these division schemes, the traffic division scheme "*TZ-15*" could be the best division scheme candidate among these schemes. We analyze whether the traffic division scheme is in line with people's perceptions.

Figure 8h shows the traffic cell division scheme of "*TZ-15*." The highlighted blue zones are *Zone 2*, *Zone 3*, *Zone 5*, *Zone 12* and *Zone 15*. As shown in Fig. 8h, the borders of the zones that are highlighted in blue are basically along the road network. Their shape is very neat and easy to understand, such as *Zone 5*. For zones that are not highlighted, they have partitive boundaries along the road network.

7.2 Evaluation

We introduced our methods and visual analysis system of dividing traffic zones to two experts who are in the urban planning field and asked for their opinions. One expert is from city transportation bureau, and the other one is from academic field who focuses on urban planning.

We received many positive feedback and suggestive comments. They all agreed that that our approach can not only relax the traditional approaches on the problem of data verification, but also complement the shortcomings of the data-driven method that may result in the complicated shape. One of our experts said that "combining road network constraints and data-driven methods for traffic zone division is solid idea and commonly recognized in urban planning field. The visual analysis system can help user to conduct good traffic characteristics analysis and spatial feature analysis for traffic zones." He would like to employ such system to conduct work related to regional planning. The other expert said that the data-driven traffic zone is more in line with the real data distribution and the road network constraints make the traffic zone look neat. The idea of this traffic zone division method had great significance in updating the traffic zone of city.

The experts also raised some suggestions for improvement. The city often uses traffic zones to do a lot of analysis, such as population census, travel analysis, traffic control analysis, etc. For example, the population census is always related to administrative boundaries. Therefore, experts suggest that our division method consider constraints such as administrative boundaries, population distribution, and land use functions.

8 Discussion and future work

The case studies and evaluation demonstrate the usability and effectiveness of the system. This study has three important implications.

First, we propose a traffic zone division approach that combines the advantages of traditional division approaches and data-driven division approaches. It is capable of generating traffic zones with road network constraints based on user-defined traffic data. It not only relax the determination that the traditional approaches has no data verification, but also complement the shortcomings of the complex and unexplainable shape of the data-driven approaches.

Second, our traffic zone division approach is equally applicable to other data such as taxi travel data, population movement data and point of interest (POI) data, etc. Users may apply our approach to divide traffic zone for demographics, economic analysis, traffic control and land function analysis. Our approach can provide support for iterative updates of urban traffic communities.

Third, we design the TZVis to interactively generate and visually analyze the traffic zone division scheme. Users can customize their interested data to generate a new traffic zone division scheme according the analysis. This method may increase the efficiency in urban planning.

However, the present work has four limitations. Firstly, we use a single source of data. The fusion of multiple real data can make the traffic zone division result more in line with the authenticity of the data. Secondly, we still need to incorporate more constraints, such as the constraints of rivers, mountains administrative boundaries, population distribution, and land use functions. Thirdly, the time complexity of our algorithm is still relatively high, and it takes much time to generate traffic cells. These limitations are what we need to solve in future work.

9 Conclusion

This paper presents TZVis, a visual analysis system that solves the shortcomings of traditional and data-driven traffic zone division approaches through a seamless integration of a novel traffic zone division approach and a set of visualization techniques. The proposed traffic zone division approach combines the advantages of the traditional rule-based methods and data-driven division approaches and employs the data voting concept to generate the traffic zone scheme under realistic constraints such as road network. Interactive visualization techniques are further proposed to display the characteristics of the traffic zone division scheme and help users visually compare the division scheme with their own interest.

Acknowledgements This work is partly supported by National Natural Science Foundation of China (No. 61972356), Zhejiang Provincial Natural Science Foundation of China (No. LY19F020026), National Natural Science Foundation of China (No. 61602409), Zhejiang Provincial Key Research and Development Program of China No. 2019C01009) and Fundamental Research Funds for the Provincial Universities of Zhejiang (No. RF-C2019001).

References

Adrienko N, Adrienko G (2010) Spatial generalization and aggregation of massive movement data. IEEE Trans Vis Comput Graph 17(2):205–219

Allsop RE (1974) Some possibilities for using traffic control to influence trip distribution and route choice. In: Proceedings of the transportation and traffic theory, vol 6

Dong H, Wu M, Ding X, Chu L, Jia L, Qin Y, Zhou X (2015) Traffic zone division based on big data from mobile phone base stations. Transp Res C Emerg Technol 58:278–291

Gratzl S, Lex A, Gehlenborg N, Pfister H, Streit M (2013) Lineup: visual analysis of multi-attribute rankings. IEEE Trans Vis Comput Graph 19(12):2277–2286

Jegou H, Douze M, Schmid C, Perez P (2010) Aggregating local descriptors into a compact image representation. Proc CVPR 238(6):3304–3311

Liu D, Weng D, Li Y, Bao J, Zheng Y, Qu H, Wu Y (2016) SmartAdP: visual analytics of large-scale taxi trajectories for selecting billboard locations. IEEE Trans Vis Comput Graph 23(1):1–10

Lu M, Lai C, Ye T, Liang J, Yuan X (2017) Visual analysis of multiple route choices based on general GPS trajectories. IEEE Trans Big Data 3(2):234–247

Lu M, Wang Z, Yuan X (2015) TrajRank: exploring travel behaviour on a route by trajectory ranking. In: 2015 IEEE pacific visualization symposium (PacificVis). IEEE, pp 311–318

Ma C, Wang R, Wang Y, Yan B (2007) Calculating method of traffic zone radius in city based on inner trip proportion. J Traffic Transp Eng 7(1):68–72

Maaten Lvd, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9(Nov):2579–2605

Martínez LM, Viegas JM, Silva EA (2009) A traffic analysis zone definition: a new methodology and algorithm. Transportation 36(5):581–599

Park JS, Oh SJ (2012) A new concave hull algorithm and concaveness measure for n-dimensional datasets. J Inf Sci Eng 28(3):587–600

Rosen E, Jansson E, Brundin M (2014) Implementation of a fast and efficient concave hull algorithm. Uppsala Universitet, Uppsala

Sun GD, Wu YC, Liang RH, Liu SX (2013) A survey of visual analytics techniques and applications: state-of-the-art research and future challenges. J Comput Sci Technol 28(5):852–867

Sun G, Liang R, Qu H, Wu Y (2017) Embedding spatio-temporal information into maps by route-zooming. IEEE Trans Vis Comput Graph 23(5):1506–1519

Sun G, Li S, Cao D, Liu C, Jiang X, Liang R (2018) DiffusionInsighter: visual analysis of traffic diffusion flow patterns. Chin J Electron 27(5):942–950

Torgerson WS (1958) Theory and methods of scaling. Biometrika 46(3/4):247–297

Wenchao W, Jiayi X, Haipeng Z, Yixian Z, Huamin Q, Bing N, Mingxuan Y, Ni LM (2015) TelCoVis: visual exploration of co-occurrence in urban human mobility based on telco data. IEEE Trans Vis Comput Graph 22(1):935–944

Whitson R, White B, Messer C (1973) A study of system versus isolated control as developed on the mockingbird pilot study. Prepared for the City of Dallas. Texas Transportation Institute, College Station and Dallas, Texas

Xing X, Huang W, Song G, Xie K (2014) Traffic zone division using mobile billing data. In: 2014 11th international conference on fuzzy systems and knowledge discovery (FSKD). IEEE, pp 692–697

You J, Nedović-Budić Z, Kim TJ (1998) A GIS-based traffic analysis zone design: implementation and evaluation. Transp Plan Technol 21(1–2):69–91

Zheng Y, Zhao G, Liu J (2015) A novel grid based k-means cluster method for traffic zone division. In: Second international conference on cloud computing and big data in Asia. Springer, pp 165–178