

# DiffusionInsighter: Visual Analysis of Traffic Diffusion Flow Patterns\*

SUN Guodao<sup>1</sup>, LI Si<sup>1</sup>, CAO Dizhou<sup>1</sup>, LIU Chunhui<sup>1</sup>, JIANG Xiaorui<sup>2</sup> and LIANG Ronghua<sup>1</sup>

(1. College of Information Engineering, Zhejiang University of Technology, Hangzhou 310012, China)

(2. Department of Computer Science and Mathematics, Aston University, Birmingham B4 7ET, UK)

**Abstract** — Traffic jam has become a severe urban problem to most metropolises in the world. How to understand and resolve these traffic problems has become a global issue. In the new era of big data, visualization and analysis with traffic-related data are increasingly appreciated. This paper presents DiffusionInsighter, a web-based visual traffic analysis system, that allows users to explore the traffic flow and diffusion patterns with different spatial and temporal granularity. The DiffusionInsighter first applies a visual data cleaning and filtering component to remove dirty data and remain available ones for further analysis. A set of carefully designed interaction and visualization tools including geographical view, pixel map view, chord diagram and network diffusion view is proposed in the DiffusionInsighter to support level-of-detail exploration of diffusion patterns of the traffic flow. Different views are collaborated together and are integrated into geographic map. A series of real-life case studies are conducted using a large GPS trajectory dataset of taxis in Hangzhou.

**Key words** — GPS trajectories, Visual analysis, Traffic dynamics, Spatio-temporal aggregation.

## I. Introduction

Traffic congestion has become a major problem in modern metropolises<sup>[1]</sup>. Severe traffic congestion problems cause great economic losses, aggravate air contamination and increase mental discomfort. Traffic administrators have been struggling for years for solutions to alleviate the problem of traffic congestion. Recently, with the wide permeation of sensing technologies, analysts and researchers have made much progress in developing intelligent transportation techniques<sup>[2-4]</sup>. Many aspects of the traffic congestion problem have been studied, such as the scalable detection of traffic congestion sites in Refs.[5,6], the relationship between subsequent events of traffic congestions

in Ref.[7], the correlation between different sites of congestions in Ref.[8], and traffic flows across a specific road and its adjacent roads in Refs.[9,10]. While data-driven intelligent traffic analysis is already in a relatively advanced status, traffic administrators are calling for more demands of a comprehensive understanding of the city's traffic situation as well as a detailed insight of a regional district traffic flow.

Understanding the traffic situation and city's movement is hindered by following major challenges. Firstly, most previous studies work on the cleaning and filtering of traffic data offline. These work do not show emphasis on evaluation and quality of the cleaned/filtered result visually. Thus, users need to iteratively process the raw data and import them into the system, which decrease the analysis efficiency. Secondly, how to recover a multi-level traffic flow and diffusion patterns from sparse traffic data remains a major challenge, since different levels of spatial regions need different visualization approaches. The geographic nature of the problem limits the visualization space because the map usually takes up most space of the screen. In addition, the fact that target users are not IT professionals introduces extra challenges to the choices of visualization schemes.

To overcome these challenges, this paper proposes a visual analysis system, namely DiffusionInsighter, with a set of carefully designed interaction and visualization tools to support level-of-detailed exploration of diffusion patterns of the traffic flow. Major contributions are summarized as follows.

- We develop a visual assistant system to perform visual traffic data cleaning and filtering. Different collaborative views are provided to help users understand the

---

\*Manuscript Received Mar. 16, 2017; Accepted July 9, 2017. This work is supported by the National Natural Science Foundation of China (No.61602409), Zhejiang Provincial NSFC (No.LR14F020002), joint project Data-Driven Intelligent Transportation between China and Europe announced by the Ministry of Science and Technology of China (No.SQ2013ZOC200020), and the Open Projects Program of Key Laboratory of Ministry of Public Security based on Zhejiang Police College (No.2016DSJSYS003).

statistical distribution of the raw traffic data and conduct data processing for further usage.

- We design DiffusionInsighter, a visual system that can assist analysts in their investigation of the complex dynamics of diffusion patterns of the traffic flow, including temporal traffic flow patterns, global traffic diffusion patterns of the road network, and local traffic diffusion patterns within a road cross.

- We offer profound insights into the dynamics of traffic diffusion patterns by using real taxi traffic datasets, including synoptic examination of the traffic in a large area and local inspection of traffic diffusion patterns.

## II. Related Work

### 1. Visual data cleaning

With the development of visualization techniques, visual data cleaning becomes an important method of data cleaning. This section reviews related work on visual data cleaning with respect to issues such as dirty data classification, uncertainty analysis. Ref.[11] proposed 33 kinds of dirty data, and claimed that considering many types of dirty data will improve the cleaning efficiency. Regarding direct visual cleaning, Ref.[12] proposed a pixel-oriented visualization method, whose main principle is to increase the data density of a point to show the value of the point. Another way to visualize large data is sampling. Dynamic clustering technique in Ref.[13] visualized the dynamic cluster of sampling, and used interactive queries to select an arbitrary interval data. For visualization of missing data, some researchers<sup>[14]</sup> found that with the use of default values to replace missing values, missing data will be hidden. In Ref.[15], the authors proposed a uncertainty data visualization model and the challenges we need to face. In addition, the article also pointed out that different areas have different definitions of uncertainty.

### 2. Visual analysis of traffic data

Traffic data visualization is to describe the space location and other attributes of objects changes along with the time, and visual analysis of Traffic data can reveal many interesting phenomenons of distribution pattern and hidden events<sup>[16]</sup>. The activity ringmap proposed in Ref.[17] shows the change of the strength of different activities over time. Liu *et al.*<sup>[9]</sup> proposed a similar design to represent the number and velocity of taxis versus time. Guo *et al.*<sup>[16]</sup> uses the ThemeRiver technology to show the change of different types of traffic flow through an intersection over time. Above are the works that cluster the traffic data based on the time dimension. However, spatial clustering is also a critical part in visual analysis of traffic data. In urban visual traffic congestion analysis, Wang *et al.*<sup>[7]</sup> organize congestion events in different areas to generate congested propagation map according to the spread relationship of congestion. In addition, they propose a variety of methods such as comparison of two-way traffic

flow, flow velocity and topology distribution, and analyze the characteristics of urban taxi driving mode.

## III. Data and Design Overview

### 1. Data description

In the past few years, the Hangzhou municipal government has been committed to the implementation of intelligent transportation systems. The traffic administrative department has installed positioning device for every taxi in Hangzhou to collect traffic location data, which leads to a large amount of traffic positioning data.

The raw data used in this paper consist of GPS records of about 11,000 taxi cabs collected between Oct 2012 and May 2016 in Hangzhou. A taxi GPS trajectory is defined as a series of GPS records the start and end points have different loading indicators. This definition results in more than 400,000 trajectories in one day and more than 144+ million trajectories in a whole year.

### 2. Visual analysis requirements

Through an interview with a traffic administrator from Hangzhou Transportation Management Bureau and a senior research staff at Hangzhou Traffic Satellite Positioning Corporation, the visualization and analysis tasks of traffic flow and diffusion patterns are summarized as follows.

R1: The system should provide a step-by-step procedure of visual assistant way towards the cleaning and filtering of raw traffic data. Most of the visualization tools needs cleaned/filtered data to feed in, however, most of them do not provide a visual way to help users understand both the spatial and temporal distribution of the raw data. It is crucial and a basic requirement for the system to support subsequent mining and visualization of the traffic flow.

R2: The system should provide a bird-eye view of traffic status of the road network for locating regions of interests. The experts interviewed have a consensus that, even as visual analytics system for specific purposes, the system should support an abundant set of general-purpose visual traffic analysis tools not only for analysing traffic distribution, dynamics and congestions, but also helping analysts locating and grasping an overview of the regions of interests for further analysis.

R3: The system should provide extended interaction tools for analyzing the detailed diffusion patterns among different connected road segments. Most urban visualization contains the feature of the traffic flow visualization, while the traffic diffusion pattern visualization with respect to the whole road network and road cross is hardly seen. Traffic flows of roads that are connected in the road network influence each other, and traffic flow on one road will propagates downstream to adjacent roads and the whole road network.

## IV. Data Preprocessing

In order to match the GPS points to specific roads, we need to obtain a clean and vector-based road network. The road network data were collected from the free geospatial data service OpenStreetMap (see Fig.1). The raw data contain directional ways, which we coin as *osmWay* in the form of vector of road points. The basic information for road point used in this paper is its geographic coordinates  $\langle \text{latitude}, \text{longitude} \rangle$ . Typically each *osmWay* also have several tags, such as road name, road layer and number of lanes *etc.*, which are very useful for recovering knowledge about the road network for better data pre-processing.



Fig. 1. Road network pre-processing with “virtual” intersections and level-of-road annotations

The raw road network file has versatile elements which are useless for traffic analysis. We filter the map file so that only *osmWays* tagged by trunk, primary, secondary, tertiary and unclassified are kept. However, several steps of pre-processing are further performed. As a real-world road usually correspond to several parallel *osmWays*, we need to make a projection of each intersection from the *osmWay* it resides on to the parallel *osmWays* if necessary. This step is vital for accurate map-matching and traffic diffluence calculation. For the sake of visualizing traffic dynamics directly in the geographic map, we further annotate the map file as follows. Each road (*i.e.*, each *osmWay* corresponding to this road) is annotated with a level tag with three values. Level-3, Level-2 and Level-1 for roads with more than 3 lanes, roads with two lanes and tiny streets. Using the annotated map file, we segment each *osmWay* into a series of so-called *splitWays*, each of which consists of a series of road points where only the two ends of the *splitWay* could be road intersections.

## V. Visual Data Cleaning and Filtering

As show in Fig.2, our visual traffic data cleaning and filtering system contains several basic and crucial processing steps including temporal reordering of trajectories, sampling of the reordered data and geographical offset correction of the positioning points.

### 1. Geographical view

The Geographical map view intuitively displays the raw traffic data on the map (shown in Fig.2). Each GPS

point is represented by a dot on the map, with different colors distinguishing different kinds of data. We classify the raw data in to five categories including time stamp duplication data, property loss data, abnormal data, suspicious data and normal data. A pie chart is displayed on the map to present the proportion of different kinds of data, and is automatically updated when the each filtering operation is performed. In addition, we also provide a box selection tool to ensure that users can obtain data within an accurate geographic area, such as a region of interest, or some important roads and segments. If wrong kinds of dirty data is among the several types that we define in advance, every time system will automatically select one type of the wrong points. The corresponding points on the map will disappear to give users a visual hint of the filtering results.

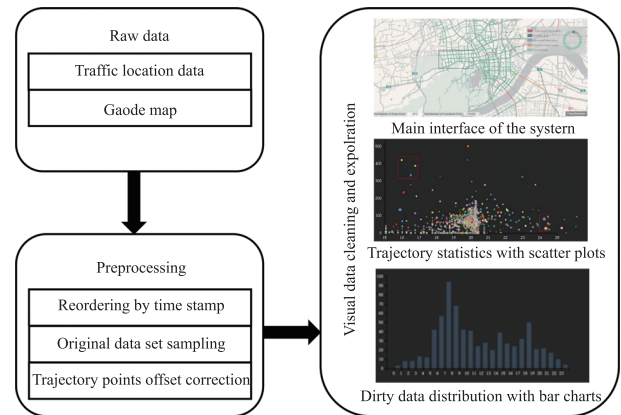


Fig. 2. Our visual data cleaning component contains two parts: raw data preprocessing and visual exploration

### 2. Statistical view

Simply relying on the geographical view hinders deeper exploration of the dynamic properties of the data such as the abnormal speed of a taxi. Thus, we present a scatterplot view and barchart view to present underlying properties of taxi drivers. As shown in the scatterplot in Fig.3, this interface includes a map view for presenting the raw data, a pie chart indicating proportion of different kinds of data, a scatterplot view showing the distribution of distance and time of each trajectory, and a bar chart showing the error distribution. The  $X$  axis represents the mean time interval of all two adjacent points in one trajectory, and  $Y$  axis represents the mean distance interval of all two adjacent points in one trajectory. For each point in the scatter plot, it represents a certain trajectory of a certain vehicle. It is clear that the points in the upper left corner of scatter plot represents the vehicles that has larger mean velocity, while the points in the lower right corner of scatter plot represents the vehicles that has smaller mean velocity. This view could help users identify the wrong GPS points in the data which leads to the abnormal velocity. In addition, a bar chart showing

the number of errors of each time period within a day is further presented (shown in the lower right corner of Fig.3). This view could help identify whether there exists an outbreak of data error, and locate the parts of the data in batch.

### VI. Traffic Diffusion Visualization

This section introduces our interactive visualization techniques that are designed based on the aforementioned user requirements.

Traffic visualization has become a common tool for personal travel and traffic management department. For the management department, simply monitoring the real-time traffic is inadequate and less effective, since each road have their own respective traffic patterns during different time period, which is affected by many factors such as the surrounding infrastructure and POIs. Thus, it is necessary to analyze the historical data of the road, summarize the patterns from historical data, and locate the similarities of traffic flow in different time periods, roads segments and areas.

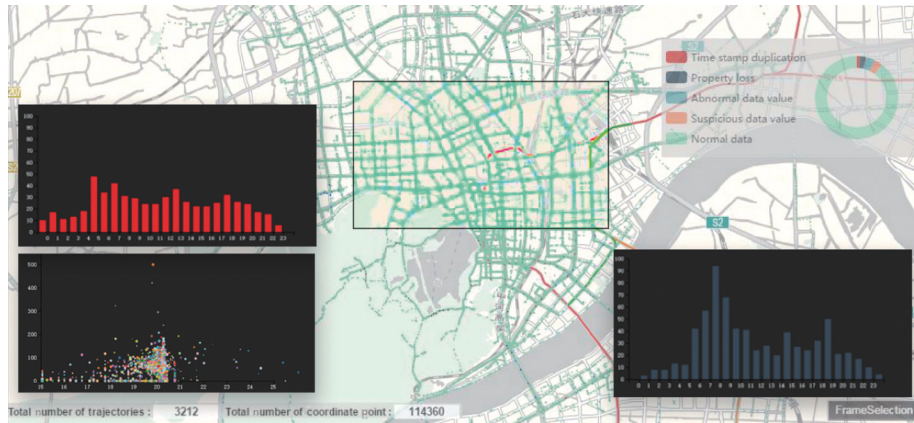


Fig. 3. The user interface for traffic data cleaning and filtering

Directly plotting traffic flow on the roads has been applied in many places such as real-time traffic display provided by map service providers. However, when the traffic is moving within the city, the analyst not only would like to observe the road traffic flow statistic, but also have the demand to understand how traffic flow diffuse among different road segments. The diffusion of traffic flow includes going straight from one road to another, turning left/right or turning round, and these behaviors will directly affect the situation of other road traffic situation. However, a comprehensive understanding of the traffic diffusion pattern is missed in the related work, which will be one of the focused topics of this paper.

#### 1. Traffic flow computation

To present the diffusion flow for different road segments, we need to take each road as a unit to compute the individual traffic flow. Due to the deviation of GPS points, we need to match the GPS points to a specific roads. We first match temporal position of each vehicle to specific roads through shortest path algorithm, and we will obtain the amounts of position points for each road. For each crossing in the road network, traffic diffusion direction includes turning left, turning right, going straight and U-turn (shown in Fig.4), thus we need to handle different situation one by one. In Fig.4, we denote two adjacent roads by road  $\vec{a}$  and road  $\vec{b}$ . The first step

is to calculate the angle between two vectors representing the two roads. If the angle is less than 30 degrees, we regarded it as straight line. If it is greater than 150 degrees, we regarded it as U-turn. Since the angle can not directly express the direction of turning left or right, we need to calculate the cross product of two vectors. If the result is positive, the vehicle turns to left, and vice versa. After the computation above, we can obtain the amount of diffusion flow among different road segments.

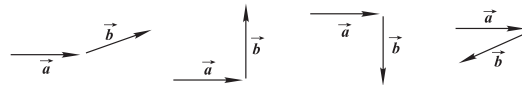


Fig. 4. The relative spatial position of four kinds of road segments, with  $\vec{a}$  and  $\vec{b}$  representing two roads vectors

#### 2. Visual traffic flow and diffusion design

1) Overall road traffic flow visualization: After the computation of the overall and temporal traffic flow of each road, we can directly draw them on each road. We apply a most intuitive and acceptable line drawing way to present the traffic flow and average velocity of each road. As shown in the Fig.8 (left), the redder a road is, the more congested of a roads' traffic is. Similar color encoding is applied to present the magnitude of velocity, the redder a road is, the less average velocity of a roads' traffic is.

In order to better carry out target analysis, visualiz-

ing the detailed patterns of each road is necessary. Thus, our system support presenting the temporal traffic flow and average velocity of each road. Users could click on a specific road, and a pixel map will popup (shown in Fig.5). The  $x$ -axis represents 24 hours in one day, and the  $y$ -axis represent one week. Each rectangle in the pixel map represents the traffic flow and average velocity of a specific road. The rectangle is split into upper and lower symmetrical triangle, and the upper triangle represents traffic flow volume and the other is average velocity.

This visual design unites the two attributes in one visual space and is easier for users to relate them. We further add an interactive function, and it allows users to switch between the view of traffic and speed. If an analyst only concerns about the flow change pattern, he can click the hide speed button to switch to the speed view, and vice

versa.

2) Traffic diffusion in a crossroad: To better analyze the relationship between different road segments, simply relying on the visualization of traffic flow for individual roads is not enough. Thus, we need to present the traffic flow diffusion patterns among different road segments. If moving into or out one road is considered separately, there are seven kinds of diffusion situations for a crossroad connected by four roads. Thus, there will 28 kinds of situation for four roads since most of the roads have two directional lanes. For example, for a crossroad, the traffic flow that travelling into a road segment A include  $A \rightarrow A$ ,  $B \rightarrow A$ ,  $C \rightarrow A$  and  $D \rightarrow A$ . Traffic flow travelling out from road segment B include  $B \rightarrow A$ ,  $B \rightarrow B$ ,  $B \rightarrow C$  and  $B \rightarrow D$ .

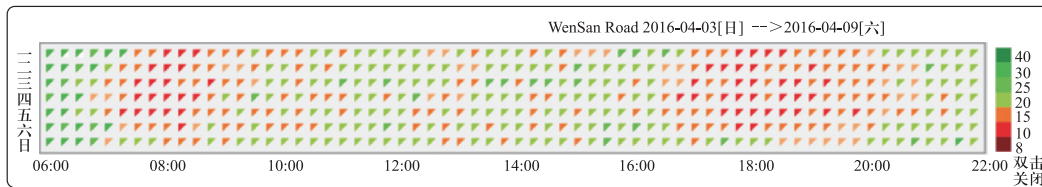


Fig. 5. Pixel map for showing the traffic flow and average velocity of a specific road during a specific time

We can see that case of  $B \rightarrow A$  is repeated. Therefore, we adopt a chord diagram to present the diffusion patterns among different roads, and we can reduce the number of repeated situation from 28 to only 16, which could further reduce the visual clutter of the final visualization.

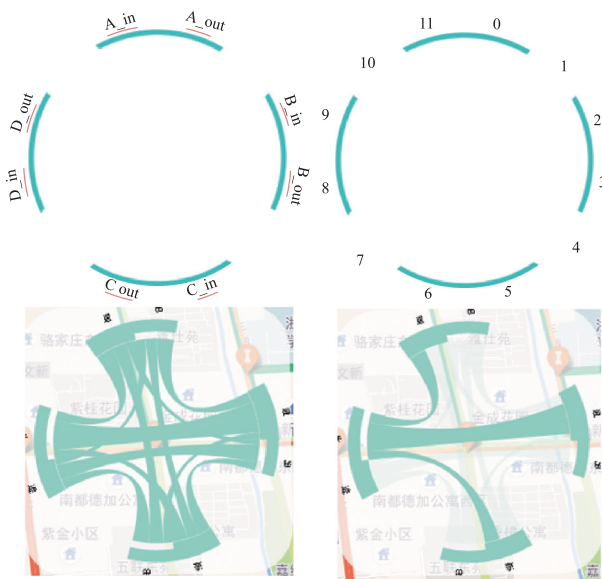


Fig. 6. The illustration of chord diagram showing the traffic diffusion pattern

The original data source for chord diagram is a matrix, and the contents of the matrix is the traffic flow from each road segment to another road segment, including a road segment itself. Thus we need to calculate the pro-

portion of each element, and extend to the entire matrix, *i.e.*, the chord diagram. For roads on the map, the direction of each arc chord diagram should correspond to the position on the road, and every road should correspond to two arcs, which represents the entering and exiting of the traffic volume of that road. Adjacent middle of the road section was added blank to help distinguish different road segments. Each intersection contains three road segments, then a chord diagram was divided into 12 segments (see Fig.6). We employ the width of the link (chord) to encode the traffic flow among different road segments. When users hover the mouse over an arc, our system will present the diffusion the flows into/out from the arc (see Fig.6 below). In addition, arc representing road segments are added with blank space to help distinguish different road segments.

We recreate the matrix underlying the chord diagram according to the illustration of the chord diagram in Fig.6. Though there are only four road segments in the illustration figure, each road is split into two groups (both directions). Thus, there will be a total of eight situations. In order to improve the aesthetics of chrod diagram, we added one blank gap between two chord (for example, the gap between 10 and 11 in Fig.6). Thus the number of situations increase to twelve. Finally, the  $4 \times 4$  matrix has become a  $12 \times 12$  matrix.

3) Traffic diffusion within a road network: Visualization of the traffic diffusion in a crossroad could help users understand how traffic moves around that area, however,

this may not be enough for a traffic management staff, who needs a more comprehensive understanding of the whole city movement. The flow distribution in a city is not a phenomenon in a road intersection. Understanding the propagation of the traffic through the whole network is more important. When observed the problem from a higher level, the path of a specific vehicle is simply a connected road segments. Thus, we propose a abstract road network diffusion view to present the traffic flow across the whole road network.

Fig.7 shows the relationship of the road traffic flow and the traffic diffusion graph of the road network. In the diffusion graph, each node encodes a road in the geographic map, and each curved edge stands for the traffic flow between two certain roads. The thickness of the curve indicates the size of traffic flow between adjacent road segments. Thus, the analyst can obtain an overview of the traffic diffusion patterns with the diffusion graph. We employ hover interaction to help link the node in the diffusion graph and the road in the geographic map. When user hovers over a node, the corresponding road segment on the map will be highlighted.

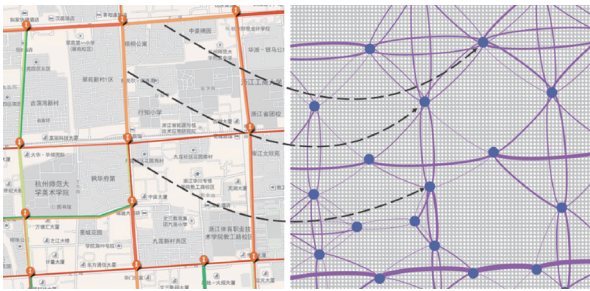


Fig. 7. The traffic flow visualization of an area and its corresponding road network diffusion patterns

In order to intuitively encode the diffusion direction of traffic flow, we comply the traffic rules of driving in the right. Thus, for the visualization of traffic diffusion of two road segments, we put the diffusion that travels from a left road segment to a right one onto the lower part of the edges connecting two corresponding nodes, and put the diffusion that travels from a lower road segment to an upper one onto the right part of the edges connecting two corresponding nodes.

In the case where the traffic junction is not very regular, the traffic flow diffusion map may be distorted to a certain degree. However, our system support modifying the curvature of the link in the network diffusion view manually to better reflect the patterns. In addition, we control the visible number of road segments by adjusting the zoom level of the map. In the larger scale, we hide the information of the branch road and help analyst analyze the traffic diffusion from a macro perspective. It reduces the interference caused by irregular road network and too much branch roads.

## VII. System Architecture and User Interaction

This system is built based on B/S architecture. In the server side, we first clean up the original GPS data, apply GPS point matching algorithm and other data preprocessing. The front end user will obtain filtered data from back-end server for the use of visualization technology to draw different visualization views. We employ Hadoop platform to store data in a distributed storage and process the data in parallel to improve computational efficiency. In addition, we employ spatial index technology to improve query efficiency when a user send request to a specific region. Different views could coordinated with each other, and could interact with the server side in real time.

The whole user interface contains a date selection, a time selector, a search bar and a rectangular lasso that facilitates user locating a specific area. After selecting a date and time on screen, they can view the global traffic flow (traffic volume or speed) situation on the map. Each crossroad is labelled with a marker automatically, users could click on the marker to examine the detailed diffusion patterns. Chord diagram will appear and rotate to a certain degree to match the angel of each crossroad of the real world map. Users can click on the road and a pixel map will popup to present the temporal traffic flow of that road. When users put the mouse on the pixel map view, corresponding road will be highlighted.

## VIII. Evaluation

### 1. Case 1: Synoptic view of the traffic in a large area

In this case study, we demonstrate that our system allows users to quickly identify interesting spatio-temporal patterns through the examination and comparison of multiple traffic flow in different local regions of interest on the map.

We randomly select an evening peak in 2016, and overlay the traffic flow visualization directly onto the roads (see Fig.8). We can see that a huge imbalance of traffic flow visualization existed on Wener Road and Wensan Road, which indicates that these two roads are one-way streets. In addition, in the area D of Fig.8, we can see that the traffic flow of Wensan Road decreases from east to west. We then examine the traffic diffusion patterns along the Wensan Road, and find most of the traffic flow spread towards surrounding roads (highlighted in red in area D of Fig.8). We can get a detailed view of the traffic diffusion patterns in a specific road cross through the chord diagram (indicated by E), most of the traffic flow from south travels into the westward road (*i.e.*, the Wenyi Road). We surmise that during the evening peak, most

people are travelling from the city center (east) to home (west), and the traffic flow on Wensan Road becomes saturated. Thus some drivers may choose other roads (*e.g.*, the Wenyi Road) which causes the traffic diffusion above.

Through the visual examination of traffic flow for a large area, the system can help relevant departments in road planning, and ease the congestion of during specific time periods.

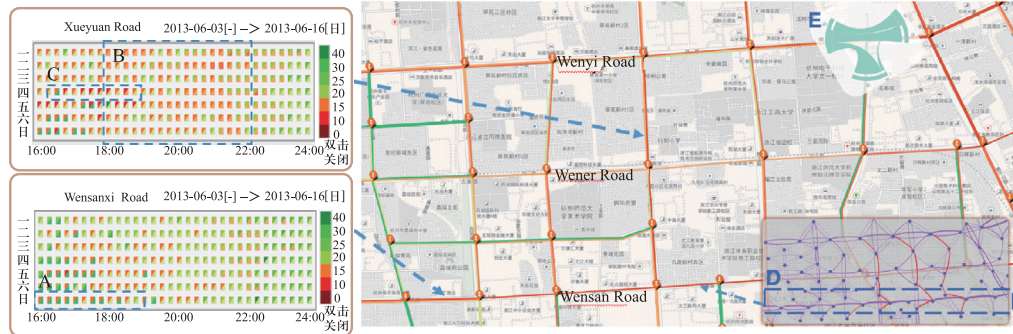


Fig. 8. Synoptic View of traffic flow in a large area. The traffic flow are directly plotted on the map to indicate the traffic volume of each road

The pixel map view in the left of Fig.8 illustrates one-week traffic flow of Wensanxi Road and Xueyuan Road. We can find that the average traffic velocity in Xueyuan Road is relatively slow and lasts for longer time (indicated by area B in Fig.8). The pixel map view can help identify the traffic patterns with high traffic flow and also relatively high traffic speed. For example, we can see that the traffic situation indicated by area A and C in Fig.8 appears to have high traffic flow (red in the upper left triangle), and high traffic speed (green or yellow in lower-right triangle). Examining the story behind this pattern will help traffic administrators analyzing traffic conditions.

## 2. Case 2: Local view of traffic diffusion patterns

This section will introduce how to employ the system to perform traffic flow and diffusion analysis in a specific local region. In this case, we select one month of traffic data in June, 2013 through the data selection bar in the upper left corner of the user interface. We monitor the daily traffic flow statistics in this area of Huanglong Sports Center of Xihu. The visualization results are shown in Fig.9.

On June 25, we find there is an obvious increase of traffic flow near Huanglong Sports Center (see the area A and E in Fig.9). To further understand the traffic diffusion patterns among different road segments, we click one of the road crosses to examine the pair-wise traffic diffusion (see the area C in Fig.9). We found that the traffic flow from JiaoGong road mainly spreads into HuangLong road, which may lead to the larger traffic flow in that region. We can also observe that the traffic flow is relatively high in that time period (indicated by area D in Fig.9), which proves a situation of congestion until 22:00. We found that the famous soprano singer Sarah Brightmans world tour was held in Hangzhou Huanglong Sports Center in that

evening through web search. We speculate this may be the reason that causes severe traffic congestion at that time. This proves that the system is helpful in assisting users in exploring a regional traffic flow information accurately and effectively.

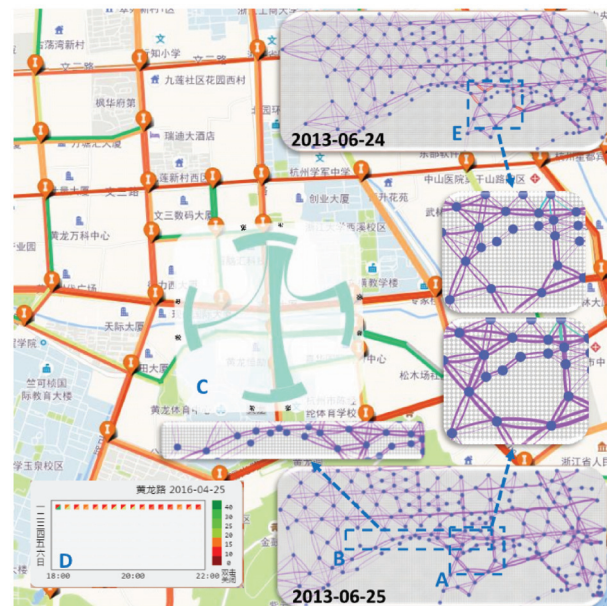


Fig. 9. Traffic diffusion patterns of a specific local region

## 3. User evaluation

We invite ten users to evaluate the proposed system. The participants finished a questionnaire with seven questions. For each question, users are required to rate their satisfaction.

Fig.10 presents these questions as well as the average user rating result. A seven-point Likert scale is used to present the result. We can find that all the questions receive positive feedback with average rating larger than

five. We can further observe that the highest score is the efficiency of the system (Q5). This is attributed to our efficient background data processing with Strom framework to ensure the efficiency of data processing. The lowest score lies on the comprehensive understanding of global patterns (Q4). We speculate that users haven't seen such visualization that switch the meaning of nodes and edges (*i.e.*, nodes for roads, and edges for pair-wise traffic diffusion strength).

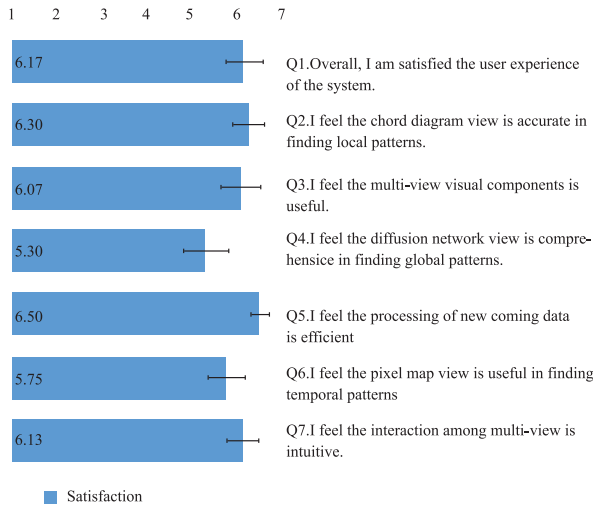


Fig. 10. User evaluation chart: all the questions receive positive feedback with average rating larger than five

## IX. Discussion and Conclusions

This paper proposes DiffusionInsighter, a visual analysis system for interactive visualization, analysis and understanding of traffic flow as well as traffic diffusion patterns. A visual data cleaning and filtering component is included in the DiffusionInsighter to help users remove non-effective data and keep track of available ones visually. Visual analysis components including geographical view, pixel map view, chord diagram and diffusion network view are tightly linked and collaborated together to support level-of-detailed analysis tasks. The effectiveness and efficiency are proven by two case studies with real world datasets.

There are multiple possible avenues for future work. The current system uses offline datasets for analysis and exploration. At the same time there are some issues that may affect the results of visual analysis. The data mentioned in this article are mainly the taxi GPS data and vector network data. Though the taxi data can display city traffic flow and speed change to a certain degree, but it may be not enough to express the distribution of the vehicle. With the increase of the numbers of private cars, simply relying on taxi data may not fully reflect the overview of the traffic situation. In addition, currently our system cannot present the temporal changes of network traffic diffusion. Users need to choose a specific time pe-

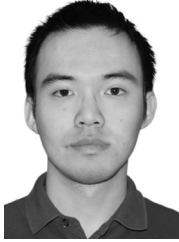
riod, and view the traffic situation. Last but not the least, in the case where the traffic junction is not very regular, the traffic flow diffusion map may be distorted to a certain degree. However, our system support modifying the curvature of the link in the network diffusion view manually to better reflect the patterns. We also plan to apply this approach to additional data from different problem domains.

## References

- [1] D. Helbing, *et al.*, *Traffic and Granular Flow99: Social, Traffic and Granular Dynamics*, Springer Science & Business Media, Berlin, Germany, pp.481–0487, 2000.
- [2] W. Chen, F. Guo and F.Y. Wang, “A survey of traffic data visualization”, *IEEE Transactions on Intelligent Transportation Systems*, Vol.16, No.6, pp.2970–2984, 2015.
- [3] R. Krueger, G.D. Sun, F. Beck, *et al.*, “TravelDiff: Visual comparison analytics for massive movement patterns derived from Twitter”, *2016 IEEE Pacific Visualization Symposium (PacificVis)*, Taipei, China, pp.176–183, 2016.
- [4] G.D. Sun, R.H. Liang, H.M. Qu, *et al.*, “Embedding spatio-temporal information into maps by route-zooming”, *IEEE Transactions on Visualization and Computer Graphics*, Vol.23, No.5, pp.1506–1519, 2017.
- [5] G. Andrienko, N. Andrienko, U. Demsar, *et al.*, “Space, time and visual analytics”, *International Journal of Geographical Information Science*, Vol.24, No.10, pp.1577–1600, 2010.
- [6] G. Andrienko, N. Andrienko and M. Heurich, “An event-based conceptual model for context-aware movement analysis”, *International Journal of Geographical Information Science*, Vol.25, No.9, pp.1347–1370, 2011.
- [7] Z. Wang, M. Lu, X.Y. Yuan, *et al.*, “Visual traffic jam analysis based on trajectory data”, *IEEE Transactions on Visualization and Computer Graphics*, Vol.19, No.12, pp.2159–2168, 2013.
- [8] Z. Wang, T. Ye, M. Lu, *et al.*, “Visual exploration of sparse traffic trajectory data”, *IEEE transactions on visualization and computer graphics*, Vol.20, No.12, pp.1813–1822, 2014.
- [9] H. Liu, Y. Gao, L. Lu, *et al.*, “Visual analysis of route diversity”, *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, Providence, RI, USA, pp.171–180, 2011.
- [10] D.Q. Zhang, L. Sun, B. Li, *et al.*, “Understanding taxi service strategies from taxi GPS traces”, *IEEE Transactions on Intelligent Transportation Systems*, Vol.16, No.1, pp.123–135, 2015.
- [11] W. Kim, B.-J. Choi, E.-K. Hong, *et al.*, “A taxonomy of dirty data”, *Data Mining and Knowledge Discovery*, Vol.7, No.1, pp.81–99, 2003.
- [12] D.A. Keim, “Designing pixel-oriented visualization techniques: Theory and applications”, *IEEE Transactions on visualization and computer graphics*, Vol.6, No.1, pp.59–78, 2002.
- [13] Hellerstein, M. Joseph and P.J. Haas, “Online aggregation”, *ACM SIGMOD Record*, Vol.26, No.2, pp.171–182, 1997.
- [14] R. Twiddy, J. Cavallo and S.M. Shiri, “Restorer: A visualization technique for handling missing data”, *Proceedings of the conference on Visualization '94*, Washington, DC, USA, pp.212–216, 1994.
- [15] A.M. MacEachren, A. Robinson, S. Hopper, *et al.*, “Visualizing geospatial information uncertainty: What we know and what we need to know”, *Cartography and Geographic Information Science*, Vol.32, No.3, pp.139–160, 2005.
- [16] H. Guo, *et al.*, “TripVista: Triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection”, *Proceedings of the IEEE Pacific Visualization Symposium*, Vol.18, No.1, pp.163–170, 2011.



- [17] J.F. Zhao, P. Forer and A.S. Harvey, "Activities, ringmaps and geovisualization of large human movement fields", *Information Visualization*, Vol.7, No.3, pp.198-209, 2008.



**SUN Guodao** is an assistant professor at the College of Information Engineering, Zhejiang University of Technology, Hangzhou, China. He received the B.S. degree in computer science and technology, and Ph.D. degree in control science and engineering both from Zhejiang University of Technology. His main research interests are information visualization and visual analytics. (Email: guodao@zjut.edu.cn)



**LI Si** is a M.S. candidate at the College of Information Engineering, Zhejiang University of Technology, Hangzhou, China. He received the B.S. degree in electronic information engineering from Zhejiang University of Technology. His main research interest is visual analytics. (Email: 2111503076@zjut.edu.cn)



**CAO Dizhou** is a M.S. candidate at the College of Information Engineering, Zhejiang University of Technology, Hangzhou, China. He received the B.S. degree in computer science and technology from Zhejiang University of Technology. His main research interest is information visualization. (Email: cdz@zjut.edu.cn)



**LIANG Ronghua** (corresponding author) received the Ph.D. degree in computer science from Zhejiang University in 2003. He worked as a research fellow at the University of Bedfordshire, UK, from April 2004 to July 2005 and as a visiting scholar at the University of California, Davis, US, from March 2010 to March 2011. He is currently a professor of Computer Science and Dean of College of Information Engineering, Zhejiang University of Technology, China. His research interests include visual analytics, computer vision, and medical visualization. (Email: rhliang@zjut.edu.cn)