Multi-scale and Interactive Visual Analysis of Public Bicycle System

Xiaoying Shi*, Yang Wang, Fanshun Lv, Xiaohang Yang, Qiming Fang, Li Zhang
School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, China
shixiaoying@hdu.edu.cn, wanggle55@gmail.com, 172050041@hdu.edu.cn, yangxiaoxing088@gmail.com,
fangqiming@hdu.edu.cn, zhangli@hdu.edu.cn
*Corresponding author: Xiaoying Shi

Abstract

Public bicycle system (PBS) is a new emerging and popular mode of public transportation. PBS data can be adopted to analyze human movement patterns. Previous work usually focused on specific scales, and the relationships between different levels of hierarchies are ignored. In this paper, we introduce a multi-scale and interactive visual analytics system to investigate human cycling movement and PBS usage condition. The system supports level-of-detail explorative analysis of spatio-temporal characteristics in PBS. Visual views are designed from global, regional and microcosmic scales. For the regional scale, a bicycle network is constructed to model PBS data, and an flow-based community detection algorithm is applied on the bicycle network to determine station clusters. In contrast to the previous used Louvain algorithm, our method avoids producing super-communities and generates better results. We provide two cases to demonstrate how our system can help analysts explore the overall cycling condition in the city and spatio-temporal aggregation of stations.

Keywords: Visual analysis, Public Bicycle System, human movement, multi-scale, spatio-temporal visualization

1. Introduction

Bike-sharing programs offer an environmental friendly, healthy and inexpensive form of public transportation, which have grown rapidly in the past decade [1][2]. Users can pick up or drop off bikes at docking stations in cities which have been equipped with public bicycle system (PBS). Utilizing information technology, most third-generation public bicycle systems record the real-time bike usage data. These data containing abundant time and space information provides great opportunities for understanding human movement and system operation condition.

For analyzing public bicycle data, the main challenge lies in how to extract useful information from massive data effectively. **First**, the volume of data is huge, containing thousands of stations and tens of millions of trip records. Traditional analytical methods are done in a black-box sort of way. Analysts are unable to participate in the analysis process, and obtain intuitive visual feedbacks. Visual analytics methods combine the computational power of machine and the cognitive ability of human, supporting better knowledge discovery by providing various visual clues [3][4]. Previous work usually focused on specific scale, such as system overview [5][6], station clusters [7][8], or single station status [9][10]. The relationships between different levels of hierarchies (city, region, and microcosmic) are ignored. Analysts cannot integrate multiple geographical scales for collaborative analysis. **Second**, due to the huge number of trip records, showing all trajectories directly would lead to visual clutter. Data abstraction methods were used to process massive PBS data. Stations were clustered by *k*-means [11], hierarchical clustering [8] and community detection methods [7][12][13][15]. A community is a set of densely connected stations, and has sparse connections with other sets. Researchers used modularity-based method to identify station communities [7][12][13][15], which ignored the interdependencies of bicycle flows in the system and tended to produce super-communities containing a large fraction of nodes.

In order to solve the above problems, we propose a multi-scale and interactive visual analytics system for public bicycle data. The system tightly couples computation, visual representation and interaction to support a top-down analysis workflow. Data from city-wide, regional, and microcosmic levels are integrated, aiming to characterize the behavior of the full system. The global overview summarizes the characteristics of bicycle flows from city-scale. For the regional scale, PBS is modeled as a bicycle network, and an effective coarse-grained description of how cycling information flows on the bicycle network is adopted. The station communities are calculated for demonstrating the flow relationships among different regions. In the microcosmic scale, the fine-grained usage patterns for interested stations or grids are revealed. Analysts can discover the potential spatio-temporal pattern of PBS from multiple levels collaboratively. The real public bicycle data in Hangzhou is used for analysis. The results of case studies demonstrate the effectiveness of our method.

The main contributions include: 1) A visual analytics system that supports a progressive understanding and level-of-detail explorative analysis of the spatio-temporal characteristics in PBS is developed. The system incorporates a set of visualization and interaction designs to facilitate overall pattern discovery and detected communities understanding. 2) A flow-based community detection method is adopted to aggregate stations into communities, which concentrates on the essence of flow interdependence in bicycle network. Compared with the previous used Louvain algorithm, our method avoids producing super-communities and generates better results.

The paper is organized as follows: Section 2 introduces the related work. Section 3 describes data form and system pipeline. In Section 4, we introduce the system design from different scales. Case studies are discussed in Section 5. Section 6 concludes the paper.

2. Related Work

With the promotion of public bicycle systems, researchers devoted to study citizens' travel patterns and system operating conditions based on PBS data.

Station clustering: Researchers extracted features from historical usage records to cluster stations [8][11], which ignored the bicycle exchange among stations. To consider the flows among stations, community detection algorithms were adopted. Borgnat et al. [12][13] studied the Lyon's bicycle sharing system, and used Louvain algorithm [14] to aggregate stations. Austwick et al. [15] adopted a hierarchical clustering algorithm to explore the station clusters in five different cities and used modularity to select an appropriate number of clusters for each city. Zhou et al. [7] constructed a bike flow similarity graph, and detected spatial communities of bike flows based on modularity. The above methods all used modularity-based algorithm to find stations communities, and the analyzed station sizes were relative small (less than 400). The modularity-based methods only considered the edge weights, but ignored the flow interdependence in the bicycle network.

Visual analysis: Visual analysis methods have been proposed to mine PBS data for different purposes. Specific visual components have been designed to show data characteristics. A table-like diagram was employed to compare and sort the rental numbers under different filter condition [16]. PCLS (Parallel coordinates with line and set) was designed to analyse the influence of multi-dimensional factors on bicycle rentals [10]. Flow map with curved flow symbols showed the overview flow structure in PBS, which were unable to highlight important flows [9]. A spatial gridded dashboard [9] and a pixel-oriented visualization component [17] are designed to monitor station status. Corcoran et al.[5] designed 'flow-comap' to visualize the effects of weather and calendar events on cycling patterns. Bargar et al. [6] revealed differences in ridership between cities through a web-based visual analytics application. A design study using visualization techniques to get insights from bike-sharing systems is described in [18]. Aiming to different analytic tasks, suitable visual components and abundant interactions between components should be well designed.

Other applications: PBS data was adopted to analyze the behaviors of cyclists, including the different usage characteristics between female and male [19], the change of user profiles [20], and the user typology [21]. Prediction methods were proposed to forecast the bike numbers renting from/returning to each station [22] or station cluster [23] in a future time period. Nair et al. [24] studied the system utilization pattern and flow imbalance. The data was also used to identify potential commuting cyclists' job and housing locations [25][26], user activity patterns [27] and social activities [28]. Other than the above analytic purposes, the focus of this article is on multi-scale system usage condition and human cycling patterns, by discovering and combining the mined knowledge from different scales.

3. System overview

3.1 Data Description

PBS data of Hangzhou from April to June in 2014 was obtained. The data contains two kinds of records. Every journey made between the original and destination station was recorded. The

trip records contain *start station*, *end station*, *start time*, *end time*, *bike ID*, *user ID*, etc. The precise trajectory between each origin and destination was unknown. The station records contain *station ID*, *station name*, *longitude*, and *latitude*. From these records, we can know where and when journeys were made. The dataset includes 21,162,572 original journeys occurring in 2759 bicycle stations.
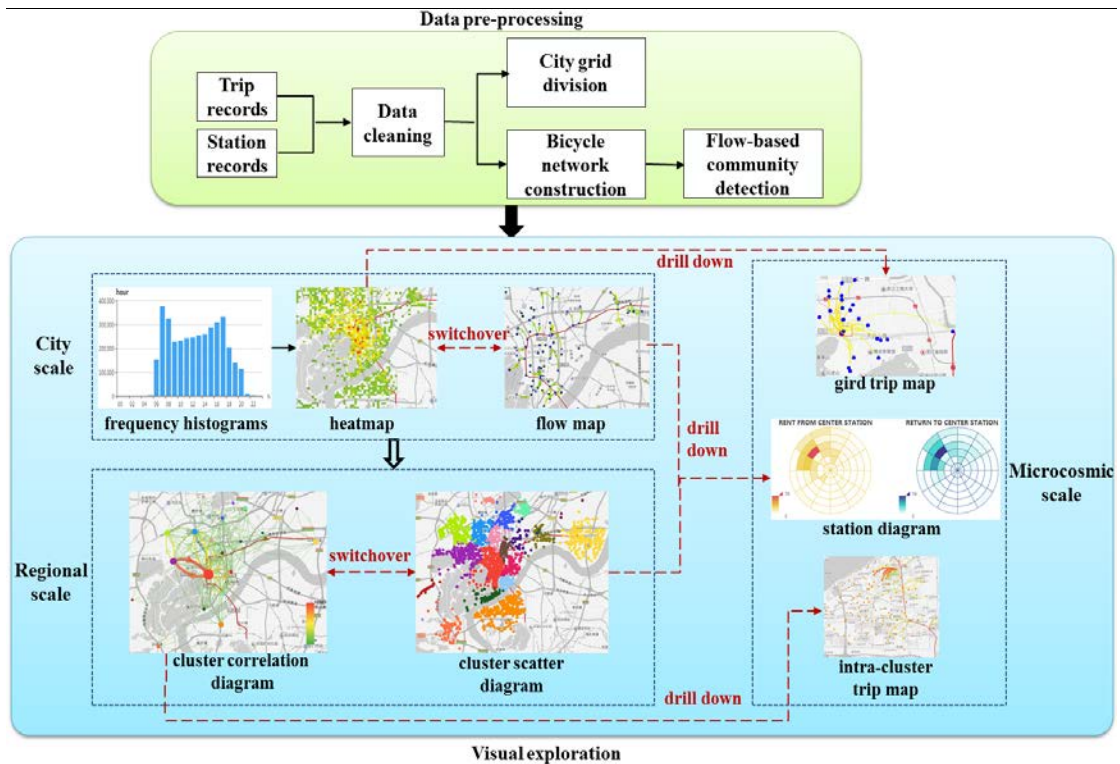

Fig. 1. System pipeline

3.2 Design goal & System pipeline

For comprehensive understanding the PBS operation from different scales, we summarize the design goals (DG) in advance.

    •**DG1**：**Find popular cycling areas/trips and cycling features of people.**

    •**DG2**：**Understand interactions among PBS and other public transit.** People ride bikes and transfer to other public transit. The connection between them should be detected.

    •**DG3**：**Show stations' community distribution**. To investigate the regional cycling patterns and avoid visual clutter, we need to categorize stations with close rental connections.

    To fulfill the above goals, we introduce a visual analytics system to support explorative analysis of PBS data. It consists two phases: data pre-processing and data visual analysis (**Fig. 1**).

    During data pre-processing, incorrect records are deleted from the original dataset, including: 1) *start station* or *end station* is null. 2) *start time* >= *end time*; 3) Duplicate trip records generated by device malfunction. After that, some results are pre-computed to speed up the subsequent calculation. Because the exact cycling trajectories between original and destination stations are unknown, the Euclidean distances of two stations are computed

for simplicity. The *distance*/*angle* between two stations and the *duration* of each journey are calculated and stored. The temporal aggregations of data by hourly intervals for each station are also stored in advance. To be more specific, *start/end day*, *start/end hour* are extracted from *start/end time*; for each station, we group by *start day*, *start hour*, *start station*, and *end station*, to compute the total bike numbers renting from start station and returning to the end station in each day and hour. To summarize the overall cycling condition in the city, the entire urban area of Hangzhou is divided into grids, and all stations are mapped to the index of the grids. To obtain station communities, a dynamical bicycle network is constructed, and a flow-based community detection algorithm is applied to the network, which will be explained in Section 4.2 in detail.

In the visual analysis phase, multiple visual components are designed from different scales. From the city scale, frequency histograms, heat map and filtered trip map are used to present system-wide information satisfying pre-defined conditions. From the regional scale, according to the community detection results, the cluster scatter diagram and cluster correlation diagram are rendered to show the station clusters and relationships among clusters. To assist the analysis of city and regional scale, detailed information of grid/station/cluster is visualized in the microcosmic scale. The grid trip map and intra-cluster trip map show the exact trips occurring in the grid or cluster respectively. The station diagram is drawn to investigate station usage pattern. Analysts can interact with various visual components, and the red dashed lines present the interactive mode. Rich and coordinated interactions are provided for collaborative analysis.

4. System Design

In this section, we explained the system design from three scales in detail.

4.1 City-scale Design

To grasp the overview condition of PBS, three visual components are designed, which enables analysts to explore the aggregated data in a highly interactive fashion. The frequency histograms present overall attributes of cycling behaviours. In view of the frequency histograms, analysts can switchover heat map and trip map to explore flow pattern in spatial context by filtering out unimportant attributes.

VIS Design 1: Frequency histograms. Based on the pre-computed results, the frequency histograms present *distance*, *duration* and *start hour* attributes of cycling behaviors. The vertical axes represent number of rentals, while the horizontal axes stand for cycling distance, duration and 24 hour in a day respectively (**Fig. 2(b)-(d)**).

VIS Design 2: Heat map. To summarize the city-wide cycling condition, we uniformly divide the entire urban area of Hangzhou into 100×100 grids. All stations are mapped to the index of grids according to their locations. Based on the city grids, the stations in the same grid are aggregated. The total usage amounts of each grid in time period T are computed. The heat map shows the popularity of each gird by adopting a gradient color scale (**Fig. 2(a)**). Red color stands for higher usage numbers, while green color for smaller numbers. We keep the same color encoding in other visual views. Analysts can further input a parameter *range* to find popular regions. The grids whose rental numbers are smaller than *range* don't shown.

VIS Design 3: Filtered trip map. Since the heat map demonstrates popular regions, the precise important flows among stations remain unshown. The trip map is a complement with respect to heat map. To avoid visual clutter, a multi-attribute query interface is provided for

filtering original trips. Three attributes are supported for filtering in view of the frequency histograms, including *duration*, *distance* and *start/end hour*. Trips meeting all constraints will be displayed on the trip map (**Fig. 4**).

In the trip map, each station is represented by a blue dot. The flow magnitude between two stations is double encoded by arc thickness and color. The thicker and redder arc indicates larger number of rentals. The arrow on the arc indicates the cycling direction. Users may borrow and return bikes to the same station, thus some stations have circle trips. For these stations, an outer ring is added on the blue dot. The thickness and color of the ring also indicates the number of circle trips in that station. The metro lines are highlighted on the map. To support progressive exploration, analysts can filter *range* of trip numbers after the trip map is first generated, and find important cycling trips conveniently.

4.2 Regional-scale Design

To reveal the underlying community structure in regional scale, we model the PBS as a bicycle network, and apply a flow-based community detection algorithm to it. After that, two visual components are designed to illustrate the station communities.

According to the bi-directional bicycle flows among stations, the bicycle network is constructed at first. A station is a node in the network, and the edges are flows generated by bike movement from one node to another. In order to capture the spatial flow patterns changing with time, the entire time period T is divided into smaller time intervals according to different analytic tasks, such as by day, by hour, or by date calendar (workday, weekend or holiday).

Assuming that N is the station set with a total of n stations, $n_i \in N \ (1 \le i \le n)$ represents one station. τ is a time interval. Let $D = \{(n_i, n_j, \tau)\}$ be the set of trip records, (n_i, n_j, τ) represents one flow from station n_i to station n_j during time interval τ. The flows among stations are regarded as directed edges. $E_\tau = \{e_{ij}\} \ (1 \le i, j \le n)$ is the directed adjacent matrix. The value of matrix element e_{ij} represents the rental number from station n_i to station n_j during τ. Because we differentiate the rental direction, so $e_{ij} \ne e_{ji}$. Based on the above definitions, the bicycle network is expressed as $G = \{G_\tau\} \ (\tau \in T)$. $G_\tau = \{N, E_\tau\}$ is a network snapshot, in which N is the station set, and $E_\tau \subseteq N \times N$ is the edge set during time interval τ. Each snapshot G_τ conveys detailed information in different time granularity.

Then, we apply a flow-based community detection method called 'Infomap' [29] to partition G_τ. For Infomap algorithm, finding optimal community structure is equivalent to minimize the description code length of a random walk on the network. It induces a structured pattern of flow with long persistence times in, and limited flow between communities. More particularly, for dividing n stations into k clusters, the average description length of a single step is given by：

$$L(M) = q_\curvearrowright H(\mathcal{L}) + \sum_{i=1}^{m} p_{\circlearrowleft}^i \mathcal{H}(\mathcal{P}^i)$$

This equation contains two terms: the entropy of the movement between clusters, and the entropy of movements within clusters. $q_\curvearrowright$ is the probability that the random walk switches clusters on any given step. $H(\mathcal{L})$ is the entropy of the cluster names. Suppose c_i is a specific cluster including many stations, $\mathcal{H}(\mathcal{P}^i)$ is the entropy of the within-cluster movements,

including the exit code for cluster c_i. The weight $p_{\circlearrowright}^i$ is the fraction of within-cluster movements that occur in cluster c_i, plus the probability of exiting cluster c_i such that $\sum_{i=1}^{m} p_{\circlearrowright}^i = 1 + q_{\curvearrowright}$. To minimize the equation, a deterministic greedy search algorithm is used to explore the space of possible partitions, and the results are refined by a simulated annealing approach. More details of the algorithm can be refer to [29]. The final partition result in time interval τ is $C_\tau = \{c_i\}\,(1 \le i \le k)$, where k is the number of communities, and c_i represents one community. A group of stations among which bicycle interchanging frequently are aggregated as a single well connected community. The circle trips are ignored during the partition process.

To help understanding the cycling patterns among communities, two diagrams are designed.

VIS Design 4: Cluster scatter diagram (left of **Fig. 7(a)**). To visualize the geographical distribution of station communities, stations in one community are represented by same color on the map. Because the bicycle network G_τ is time-varying, the cluster result C_τ in different time interval is changing. The cluster results between different G_τ (such as G_{τ_i} and G_{τ_j}) need to be compared. In the cluster scatter diagram, if the color values are randomly assigned for clusters in each time interval, the stations in similar geographical regions would have entirely different colors across different time intervals, which is hard for pattern discovery. Therefore, we propose a color assignment strategy. The goal is that if the stations are adjacent in geographical locations, their cluster colors should be as consistently as possible across different time intervals.

To be more specific, we partition the G_{τ_i} first, and obtain the cluster result C_{τ_i}. Each cluster in C_{τ_i} is assigned with a unique color randomly, and these colors are the baseline for subsequent color assignment. Then, the cluster result C_{τ_j} for the next G_{τ_j} is computed. For each cluster in C_{τ_j}, the strategy finds a cluster whose geographical location is closest to it in C_{τ_i}, and assigns the same color. If the cluster number k_j in C_{τ_j} is bigger than k_i in C_{τ_i}, the remaining clusters are assigned colors randomly. By using this color assignment strategy, the cluster visualization results have consistent color distribution across multiple G_τ.

VIS Design 5: Cluster correlation diagram (right of **Fig. 7(a)**). To visualize the relationship among communities, the centroid of each community is computed, whose location is the average longitude and latitude of all stations in that community. Each cluster centroid is represented by a circle on the map. The circle size is proportional to the station number in that community. The arcs between clusters reflect the flow magnitude, whose color and thickness encoding scheme is in line with the trip map.

By observing the two diagrams together, analysts can have a clear understanding of the cluster geographical distribution, and analyze the bicycle flows among clusters.

4.3 Microcosmic-scale Design

The visual concompents in the microcosmic scale are designed to assist the analysis of city and regional scale. The detailed trips in the grids and clusters are visualized. The unique usage pattern of station is also visualized.

VIS Design 6: Grid/Intra-cluster trip map. When a specific grid or cluster is selected, the grid trip map (right of **Fig. 3**) and intra-cluster trip map (**Fig. 8**) demonstrate the exact trips occurring in the grid and cluster respectively. In both trip maps, the stations are represented by dots, the color and thickness of arcs reflect the flow magnitude. Analysts can further filter the *range* of flow magnitude to find significant flows in that grid/cluster.

VIS Design 7: Station diagram. The station diagram includes two sub-diagrams: rental and return flow distribution diagrams (**Fig. 6(a)(b)**). Both diagrams adopt the radial layout design. Stations associated with the analyzed station are aggregated according to *distance* and *angle*. The radius encodes *distance*, while direction encodes *angle*. The *distance* increases by 1 kilometre as the radius increases. Each sector represents the bicycle usage number from that direction. Sector with a larger number will be mapped with a darker color. When clicking a certain sector, the stations belonging to that sector will be displayed on a map.

More specifically, the rental flow distribution diagram (**Fig. 6(a)**) shows the numbers of people renting bikes from the analyzed station and returning to other stations. From which, we can know the most common destination areas for a station. The return flow distribution diagram (**Fig. 6(b)**) represents the numbers of people renting bikes from surrounding stations and returning to the analyzed station. We can know people renting bikes from which areas would most likely return to one station.

(a)   (b)

(c)   (d)

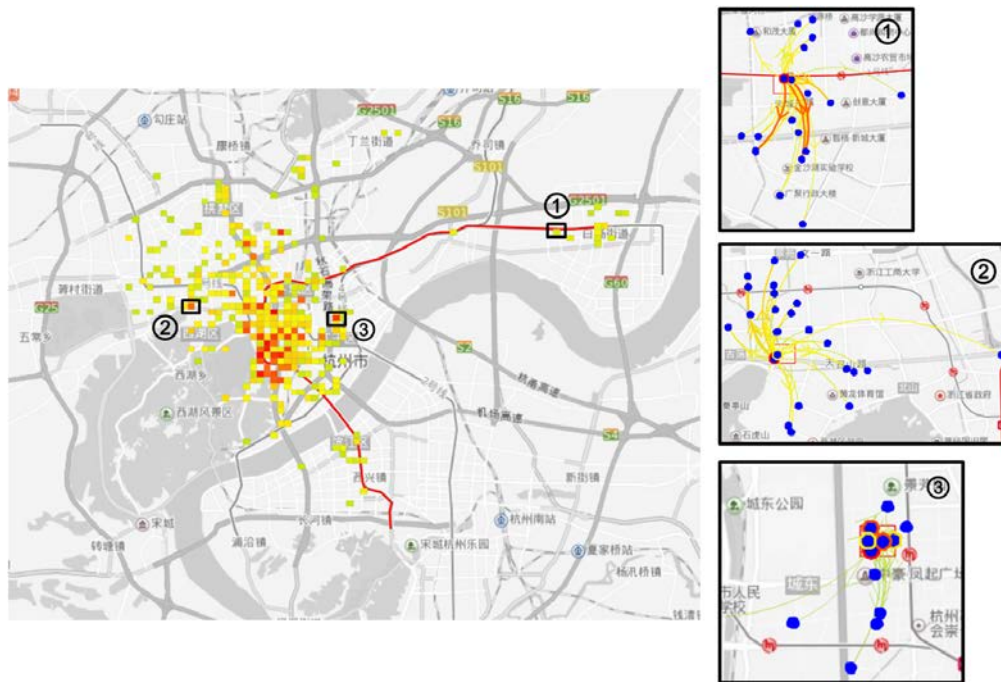**Fig. 2.** The overall usage condition of PBS

Fig. 3. Heat map (*range*>20000)

5. Case studies

This section demonstrates how our system can be used to find reliable cycling patterns. Two case studies are carried out to show different functional aspects of our system. Algorithm comparison and the limitation of our method are also discussed.

5.1 Case study 1: analysis of overall bicycle usage pattern

We study the overall usage condition of PBS in Hangzhou at first (**DG1**). Seen from **Fig. 2**, the cycling trips are widespread in the city, and the main cycling areas concentrate on downtown. As shown in the frequency histograms, we find that the *distance* curve is unimodal. The peak appears around 1 *km*, and the usage numbers are large from 0.3 *km* to 3 *km*. People rarely ride bikes over 7 *km*. For trip *duration*, most journeys take time within one hour, since a journey lasting more than one hour would charge fees. The usage amounts are large within 30 minutes. For rental *hours*, the morning peak appears during 7AM to 8AM, and the afternoon peak is around 4PM to 5PM. After 6PM, the rental amounts decrease obviously.

In order to observe the popular regions more clearly, we filter out grids with flow numbers smaller than 20000 (left of **Fig. 3**). By clicking the grids and viewing the grid trip map, we find that the hot grids can be divided into two categories roughly. For grids in the first category, they contain stations near subway or bus stops. Their trips are radiating from the center stations to the surrounding regions, such as grid ① and grid ②. Grid ① contains bicycle stations near JinShaHu metro station, and grid ② contains bicycle stations near GuDang central bus station. For the second category, large circle trips are found in one grid. People often start from one grid and go back to it.

Afterwards, we turn our focus to compare the trip maps in different time periods. To filter out journeys that don't really happen, for example, people borrow a bike and immediately

return it causing by bike malfunction, we set *duration>=5min*. The minimum value of *range* is set to be 50 to ignore journeys with relatively smaller usage numbers. **Fig. 4** shows the popular trip distribution during the morning peak. Several regions are magnified to show the arrow directions more clearly. From this figure, we find that journeys with larger rental amounts have short distances. Different from premeditation, the two largest journeys (red color) do not appear in downtown, but in XiaSha district. One trip is from "South door of China Jiliang University" to "China Jiliang University" (trip 1), which is from the dormitory area to teaching area for one university. Another journey is from "Bus WenSu Terminal" to "Metro WenZe Station" (trip 2), which is from a bus station to a subway entrance.

We also find the connection of PBS and metro system (**DG2**). Most popular trips are along the metro line, conforming to the design goal of PBS: 'solve the last mile problem'. Seen from the enlarged figures, most arrow directions point to metro stations, which means users ride bikes to stations near subway entrance, and transfer to subway in the morning.

Then, we observe the trip distribution during afternoon peak (**Fig. 5**). The trips with larger usage amounts also like the morning rush hour, causing by college students' cycling between teaching areas and dormitories. Compared to the morning peak, more stations have larger circle trips, which may be related to citizens' shopping behavior after work. From the enlarged figures, we find that the main rental directions are opposite to the morning peak along the metro line.
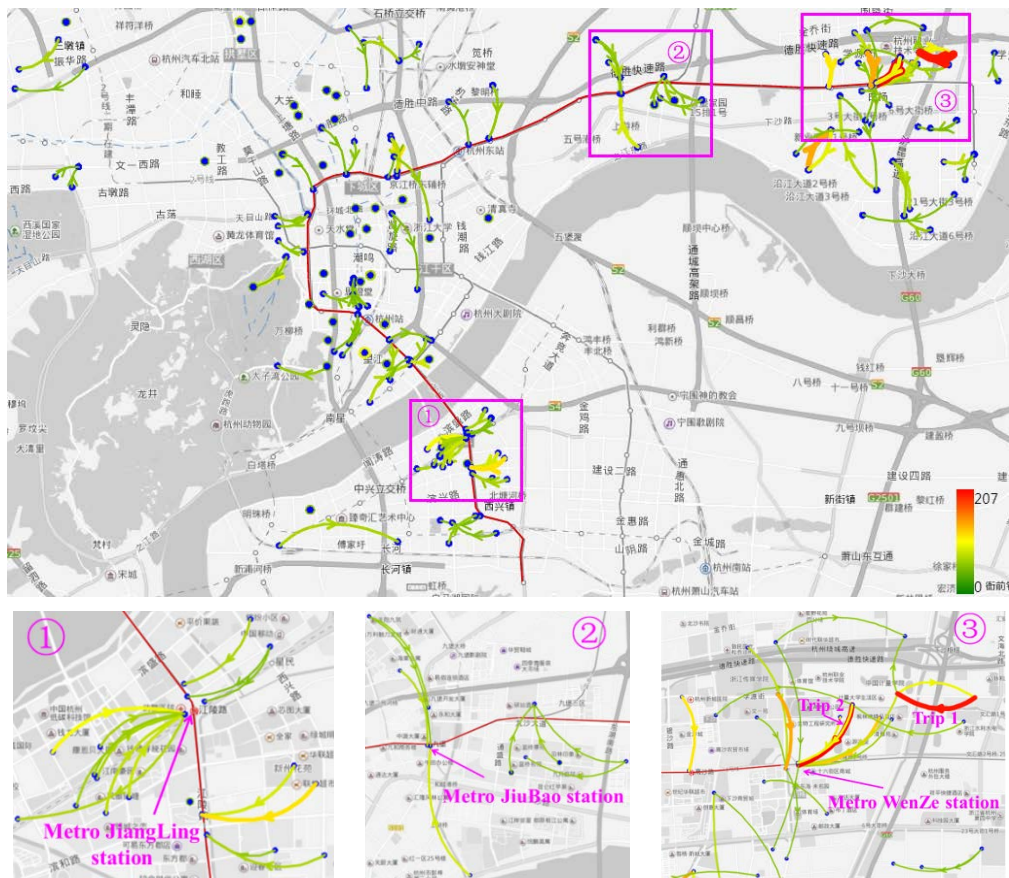

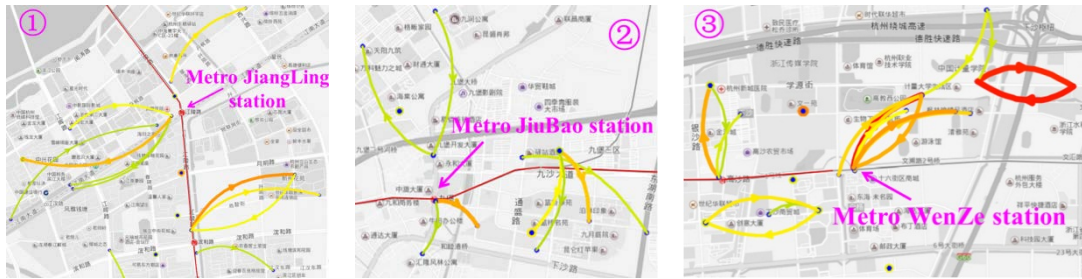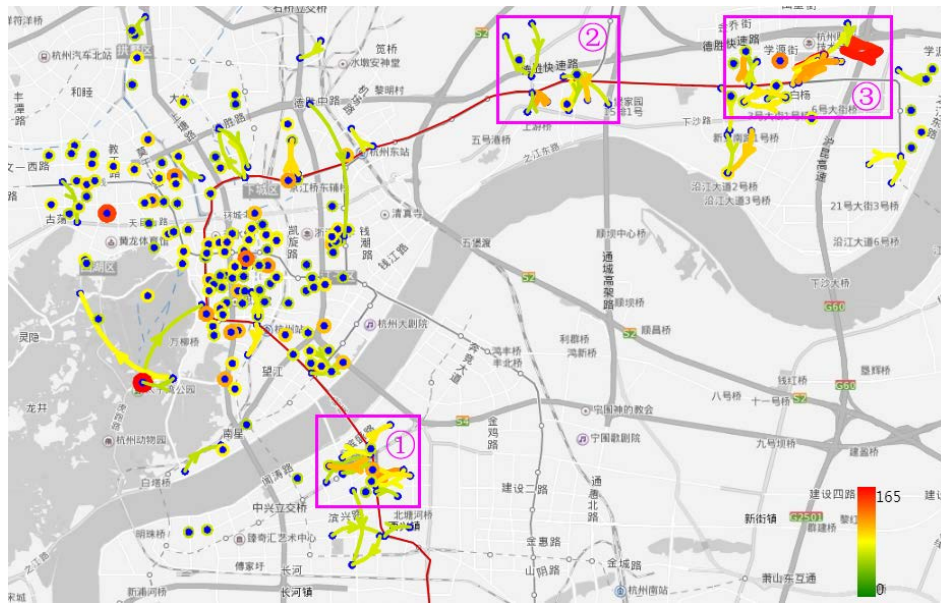**Fig. 4.** Filtered trip map (*hour*=[7AM,8AM], *duration>=5min*, *range>=50*)

Fig. 5. Filtered trip map (*hour*=[4PM,6PM], *duration*>=5*min*, *range*>=60)

(a) rental, hour=[7AM, 8AM]   (b) return, hour=[7AM, 8AM]

(c) rental, hour=[4PM,6PM]   (d) return, hour=[4PM,6PM]
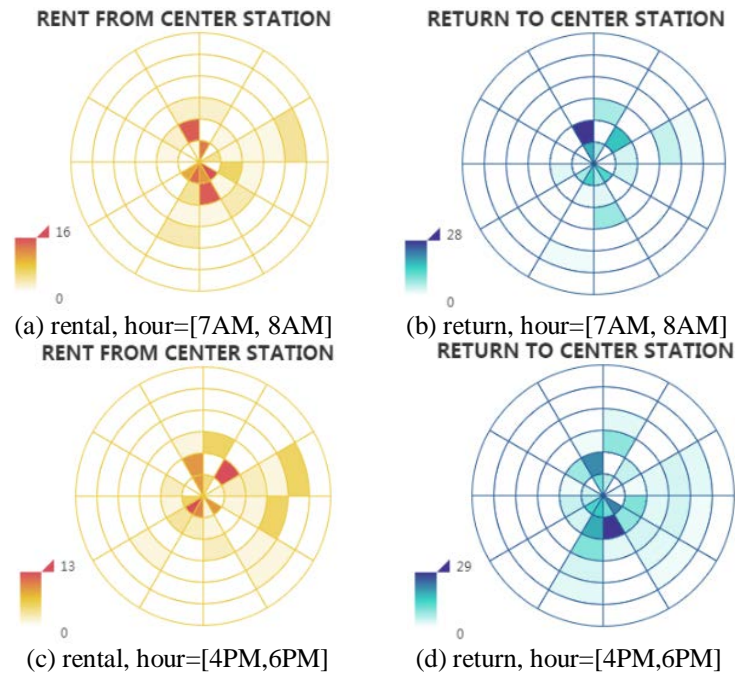
Fig. 6. The rental/return flow distribution diagrams of "Metro WenZe station" in one week

(a)   5 April to 7 April, holiday (Tomb-Sweeping Day) , *k*=32

(b)   14 April to 18 April, workday, *k*=25

(c)   19 April to 20 April, weekend, *k*=29

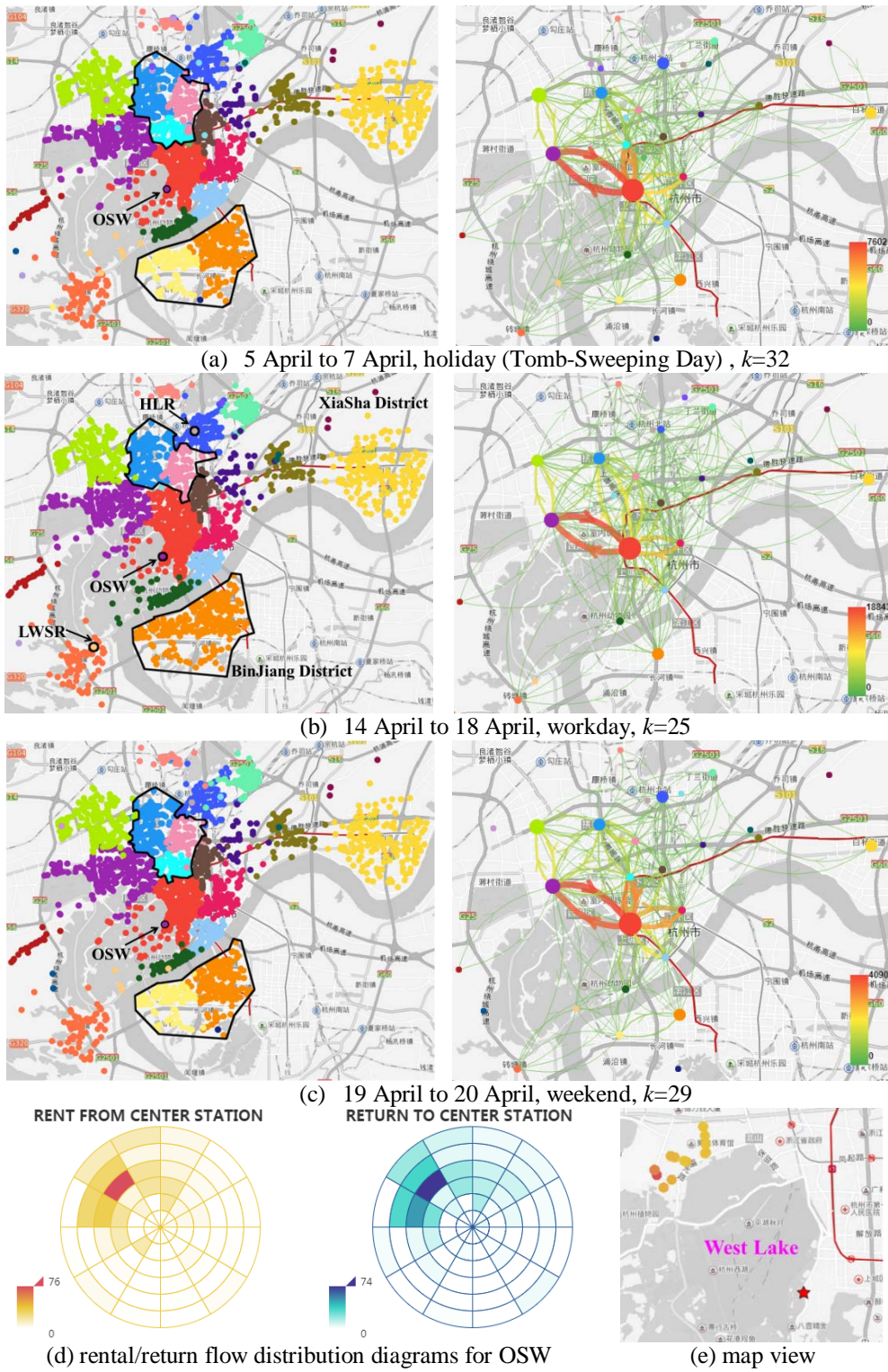(d) rental/return flow distribution diagrams for OSW   (e) map view

Fig. 7. The community detection results under different calendar attributes by using Infomap

In order to verify our speculation, we further choose the "Metro WenZe station" to compare its usage condition during the morning peak and afternoon peak. From **Fig. 6(a)-(d)**, we find that the distribution in **Fig. 6(a)** and **Fig. 6(d)** are consistent. During the morning peak, some users get out of the subway and borrow bikes to go somewhere else, may be company (**Fig. 6(a)**). Accordingly, these users ride to subway entrance from other places and return bikes during the afternoon peak (**Fig. 6(d)**). A similar pattern can be found from **Fig. 6 (b) (c)**. Users ride to subway entrance from other places in the morning and borrow bikes near subway entrance in the afternoon.

The above results all indicate that people borrow public bicycles to transfer subway during rush hours, which is also the green travel mode advocated by Hangzhou government. We also compare the common cycling trips with short distance and long distance, and find that most cycling trips with short distance are along subway. The longer cycling trips have a great change, most of which are along the West Lake also with longer duration. People may ride bikes for sightseeing. In conclusion, for the purpose of commuting, people usually ride a short distance to connect with other faster means of transport to save time. While travelling, they are more willing to ride a longer distance to enjoy the landscape.

5.2 Case study 2: analysis of regional bicycle usage pattern

The second case study analyzes the flow correlation inter and intra communities (**DG3**).

Fig. 7(a)-(c) show the community detection results under different calendar attributes by using Infomap. We look at the cluster scatter diagram at first. On the whole, the stations in one cluster are geographically adjacent, regardless of the changing of calendar attributes (workday, weekend, holiday). This indicates that many trips are local and adjacent stations exchange more bicycles than distant stations. The cluster results also reflect underlying geography. The deep yellow region is corresponding to XiaSha district, while the orange region to BinJiang district (**Fig. 7(b)**). Several regions have slight aggregation and partition (drawn by black line), such as the BinJiang district is divided into two sub-regions sometimes, meaning the internal flows of the sub-regions are more closely during those time intervals. In general, the cluster scatter diagrams show that the operation of public bicycle system is stable. The station partitions based on cycling flows remain the same on the whole, and the bicycles in the same cluster are exchanged more frequently, which provides guidelines to bicycle scheduling.

We also find some singular stations. Take the station "Orioles Singing in the Willows (OSW)" for example, it is far away from the cluster center. We check the flow distribution diagrams to investigate the station usage pattern. From **Fig. 7(d)**, we find that most related areas appear in the north-western direction, which are 3*km* to 4*km* far away from the center. We can see the station distribution in that area more clearly from the map view (**Fig. 7(e)**). Since this station belongs to one of ten scenes of West Lake, people usually ride bikes to other sides of West Lake, so this station is divided to the purple area. Some stations are partitioned as one cluster, such as the "LongWangSha Road(LWSR)" and "HuiLong Road(HLR)". By using the flow distribution diagrams, we find these stations have few flows during the observed time period.

Next, we observe the cluster correlation diagram in **Fig. 7**. The red cluster is central region all the time, corresponding to city heart. The red region keeps close connections to other regions, and the quantities of bicycle inflow and outflow of this region are very large. The purple region located in the western part of city maintains high traffic volume with the red region, indicating human movements between the city heart and the western part are very frequently. The arcs among most regions are green, meaning these regions have few

correlations with other regions. These regions form independent subsystems. Within these regions, the bicycles are exchanged frequently, while they seldom exchange bicycles with other regions.

Finally, we investigate the station relationship within a specific cluster. Non-significant journeys are filtered out first. For BinJiang district, we find that the cycling trips are along subway on workdays (**Fig. 8(a)**). On 1 May, in addition to the trips along subway, there have obvious flows from subway (Metro BinKang station) to "BaiMa lake station" (**Fig. 8(b)**). During that time, the China International Cartoon & Animation Festival was held near BaiMa lake, so this trip was more popular. We also find the predominant stations of the purple region in **Fig. 7** are central bus stations, indicating the connection of bus and public bicycle.
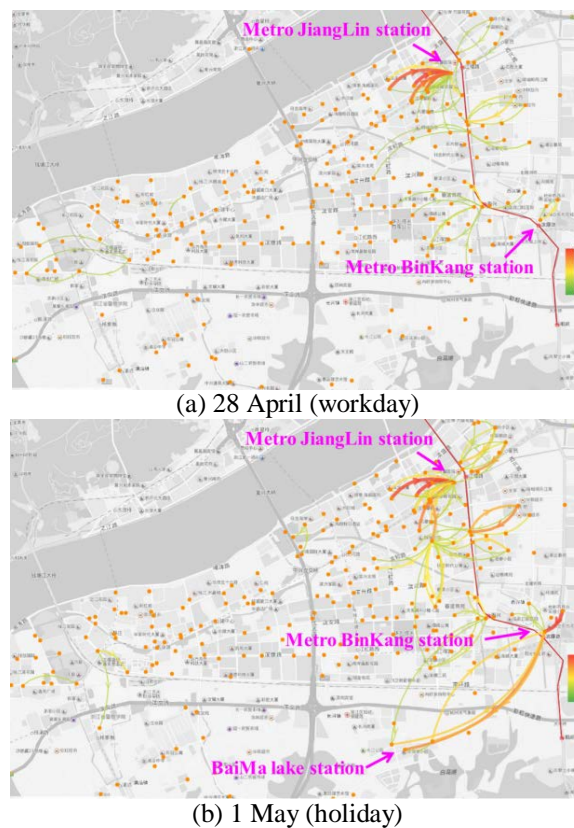

(a) 28 April (workday)

(b) 1 May (holiday)

Fig. 8. The intra-cluster trip map of BinJiang district during different days

5.3 Discussion

We compare the infomap algorithm with the previous used Louvain algorithm [13]. **Fig. 9** illustrates the station partition results. The two methods reflect different senses of what it means to have a network. The flow-based method (infomap) considers the essence of a network is the flow pattern induced by its structure, which minimizes the description code length. The modularity-based method (Louvain) divides clusters according to the topological properties of network links, which maximizes the modularity. As shown in **Fig. 9(b)**, the Louvain algorithm has a tendency to produce super-communities containing a large fraction of nodes. The infomap algorithm generates a better result in the view of code length and modularity.
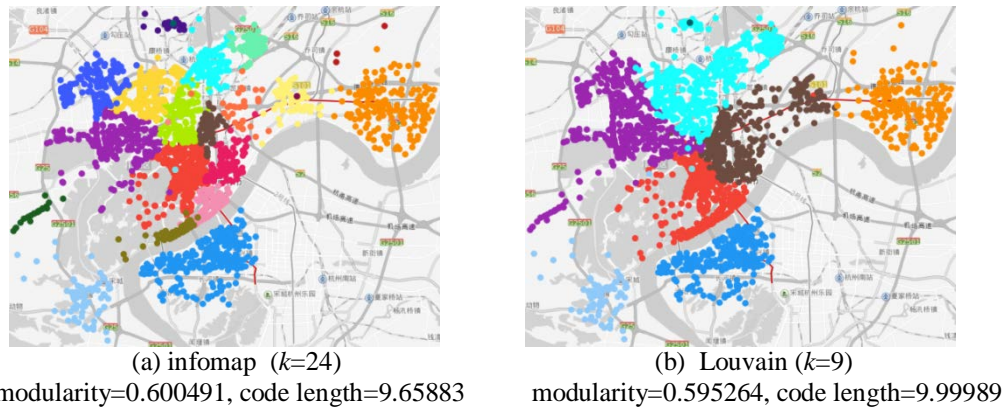
(a) infomap (*k*=24)	(b) Louvain (*k*=9)
modularity=0.600491, code length=9.65883	modularity=0.595264, code length=9.99989

Fig. 9. The station partition results by infomap and Louvain for PBS data

There are still some issues not well resolved in this paper, which will be addressed in the future work. **First**, althrough a multi-attribute query interface has provided for filtering out uninterested trips, the result may generate overlapping flows. Sometimes, the arrows on the arcs are not clear. In the future work, we would like to design new scheme to represent flows with no loss of geographic information and flow magnitude accuracy. **Second**, a flow-based communitiy detection algorithm has been adopted to cluster stations, which has been demonstrated better than Louvain algorithm. The cluster scatter diagram and cluster correlation diagram present the community detection results. However, we can not judge the changing of belonged communities for one station under different constraints immediately, especially for the stations along the margins of several communities. In the future work, we will design a compact representation of community categories, to allow analysts to compare the station clusters under different date attributes more intuitively.

5. Conclusion

In this paper, we present a multi-scale visual analytics system to investigate human cycling pattern and system usage condition based on PBS data. Multiple coordinated views are designed to explore the overview PBS operation, spatio-temporal aggregation of stations, and station usage pattern. Through case studies, we conclude several insights found by using our system: 1) The connection of PBS and subway is very close especially for commuting purpose. 2) People usually ride a short distance to connect with other faster means of transport to save time when commuting, while they are more willing to ride a longer distance to enjoy the landscape when travelling. 3) The operation condition of public bicycle system is stable, and the station clusters reflect underlying geography. The bicycles in the same cluster are exchanged more frequently, while the bicycle exchanging numbers among clusters are relatively few. The infomap algorithm performs better than Louvain according to the measure criterions of code length and modularity. It detects communities with closer rental correlation, which can provide guidelines to bicycle scheduling.

Although PBS data can be used to effectively explore human movement pattern, cycling is only one kind of public transportation mode. Other people may take bus, taxi or car. Therefore, the patterns discovered in this work only represent the habits of cycling users. In the future, we plan to combine multi-source data, such as taxi data, bus data and social media data, study the data fusion methods and design more intuitive visual tools to explore user activity pattern and city dynamics comprehensively.

References

[1] E. Fishman, "Bikeshare: A review of recent literature," *Transport Rev*, vol. 36, no.1, pp.92-113, 2016. Article (CrossRef Link).

[2] M. Ricci, "Bike sharing: A review of evidence on impacts and processes of implementation and operation," *Research in Transportation Business & Management*, vol. 15, pp. 28-38, 2015. Article (CrossRef Link).

[3] N. Andrienko, G. Andrienko, "Visual analytics of movement: An overview of methods, tools and procedures," *Information Visualization*, vol. 12, no.1, pp.3-24, 2013. Article (CrossRef Link).

[4] M. Lu, S. Chen, C. Lai, et al, "Frontier of Information Visualization and Visual Analytics in 2016," *J Visual-Japan*, vol. 20, no.4, pp.667-686, 2017. Article (CrossRef Link).

[5] J. Corcoran, T. Li, D. Rohde, et al, "Spatio-temporal patterns of a Public Bicycle Sharing Program: the effect of weather and calendar events", *J Transp Geogr*, vol. 41, pp. 292-305, 2014. Article (CrossRef Link).

[6] A. Bargar, A. Gupta, S. Gupta, et al, "Interactive visual analytics for multi-city bikeshare data analysis," in *Proc. of 3rd Int. Workshop on Urban Computing*, August 2014. Article (CrossRef Link).

[7] X. Zhou, "Understanding spatiotemporal patterns of biking behavior by analyzing massive bike sharing data in Chicago," *PloS one*, vol. 10, no.10, pp. e0137922, 2015. Article (CrossRef Link).

[8] J. Froehlich, J. Neumann, N. Oliver, "Sensing and Predicting the Pulse of the City through Shared Bicycling," in *Proc. of Int. Conf. on IJCAI*, pp. 1420-1426, July 2009. Article (CrossRef Link).

[9] J. Wood, A. Slingsby, J. Dykes, "Visualizing the dynamics of London's bicycle-hire scheme," *Cartographica: The International Journal for Geographic Information and Geovisualization*, vol. 46, no. 4, pp. 239-251, 2011. Article (CrossRef Link).

[10] X. Shi, Z. Yu, J. Chen, et al, "The visual analysis of flow pattern for public bicycle system," *J Visual Lang Comput*, vol. 45, pp. 51-60, 2018. Article (CrossRef Link).

[11] X. Shi, Z. Yu, H. Xu, et al, "Clustering the stations of bicycle sharing system," *Journal of Donghua University*, vol. 33, no. 6, pp.968-972, 2016. Article (CrossRef Link).

[12] P. Borgnat, P. Abry, P. Flandrin, et al, "Shared bicycles in a city: A signal processing and data analysis perspective," *Adv Complex Syst*, vol. 14, no.3, pp.415-438, 2011. Article (CrossRef Link).

[13] P. Borgnat, C. Robardet, P. Abry, et al, "A dynamical network view of lyon's vélo'v shared bicycle system," *Dynamics On and Of Complex Networks*, vol. 2, pp. 267-284, 2013. Article (CrossRef Link).

[14] V. Blondel, J. Guillaume, R. Lambiotte, et al, "Fast unfolding of communities in large networks," *J Stat Mech-Theory E*, vol. 2008, pp.P10008, 2008. Article (CrossRef Link).

[15] M. Austwick, O. O'Brien, E. Strano, et al, "The structure of spatial networks and communities in bicycle sharing systems," *PloS one*, vol. 8, no. 9, pp. e74685, 2013. Article (CrossRef Link).

[16] X. Shi, Q. Zhou, X. Qu, et al, "Visual Analysis of Station Usage Patterns in Public Bicycle System," in *Proc. of Int. Symposium on Computational Intelligence and Design*, pp.132-135, December 2016. Article (CrossRef Link).

[17] G. Oliveira, J. Sotomayor, R. Torchelsen, et al, "Visual analysis of bike-sharing systems," *Computers & Graphics*, vol. 60, pp.119-129, 2016. Article (CrossRef Link).

[18] J. Wood, R. Beecham, J. Dykes, "Moving beyond sequential design: Reflections on a rich multi-channel approach to data visualization," *IEEE T Vis Comput Gr*, vol. 20, no.12, pp. 2171-2180, 2014. Article (CrossRef Link).

[19] R. Beecham, J. Wood, "Exploring gendered cycling behaviours within a large-scale behavioural data-set," *Transport Plan Techn*, vol. 37, no.1, pp. 83-97, 2014. Article (CrossRef Link).

[20] A. Goodman, J. Cheshire, "Inequalities in the London bicycle sharing system revisited: impacts of extending the scheme to poorer areas but then doubling prices," *J Transp Geogr*, vol. 41, pp. 272-279, 2014. Article (CrossRef Link).

[21] M. Vogel, R. Hamon, G. Lozenguez, et al, "From bicycle sharing system movements to users: a typology of Vélo'v cyclists in Lyon based on large-scale behavioural dataset," *J Transp Geogr*, vol. 41, pp. 280-291, 2014. Article (CrossRef Link).

[22] Z. Yang, J. Hu, Y. Shu, et al, "Mobility modeling and prediction in bike-sharing systems," in *Proc. of Int. Conf. on Mobile Systems, Applications and Services*, pp.165-178, June 2016. Article (CrossRef Link).

[23] Y. Li, Y. Zheng, H. Zhang, et al, "Traffic prediction in a bike-sharing system," in *Proc. of Int. Conf. SIGSPATIAL on Advances in Geographic Information Systems*, pp.33, November 2015. Article (CrossRef Link).

[24] R. Nair, E. Miller-Hooks, R. Hampshire, et al, "Large-scale vehicle sharing systems: analysis of Vélib'," *Int J Sustain Transp*, vol.7, no.1, pp. 85-106, 2013. Article (CrossRef Link).

[25] X. Shi, Z. Yu, Q. Fang, et al, "A Visual Analysis Approach for Inferring Personal Job and Housing Locations Based on Public Bicycle Data," *ISPRS International Journal of Geo-Information*, vol. 6, no.7, pp.205, 2017. Article (CrossRef Link).

[26] R. Beecham, J. Wood, A. Bowerman, "Studying commuting behaviours using collaborative visual analytics," *Comput Environ Urban*, vol. 47, pp. 5-15, 2014. Article (CrossRef Link).

[27] Y. Yan, Y. Tao, J. Xu, et al, "Visual analytics of bike-sharing data based on tensor factorization," *J Visual-Japan*, vol.21, no.3, pp. 495-509, 2018. Article (CrossRef Link).

[28] L. Chen, D. Yang, J. Jakubowicz, et al, "Sensing the pulse of urban activity centers leveraging bike sharing open data," in *Proc. of Int. Conf. UIC-ATC-ScalCom*, pp.135-142, August 2015. Article (CrossRef Link).

[29] M. Rosvall, C. Bergstrom, "Maps of random walks on complex networks reveal community structure," in *Proc. of the National Academy of Sciences*, vol. 105, no.4, pp.1118-1123, 2008. Article (CrossRef Link).

Xiaoying Shi received the Ph.D. degree in control theory and control engineering from Zhejiang University of Technology in 2014. She is currently a lecturer at School of Computer Science and Technology, Hangzhou Dianzi University, China. Her research interests include information visualization, data mining and urban transportation.

Yang Wang is an undergraduate student at School of Computer Science and Technology, Hangzhou Dianzi University, China. His research interest includes information visualization.

FanShun Lv received the B.S. degree from Anhui University of Finance and Economics. He is currently pursuing the M.S. degree in computer technology with the School of Computer Science and Technology, Hangzhou Dianzi University. His research interests are in geographic information visualization and spatial data mining.

Xiaohang Yang is an undergraduate student at School of Computer Science and Technology, Hangzhou Dianzi University, China. His research interest includes information visualization.

Qiming Fang received the Ph.D. degree in computer science and technology from Tsinghua University in 2010. He is currently a lecturer at School of Computer Science and Technology, Hangzhou Dianzi University, China. His research interests include data mining, parallel computing and wireless sensor network.

Li Zhang received the Ph.D. degree in communication and information system from Zhejiang University in 2008. Her is currently a lecturer at School of Computer Science and Technology, Hangzhou Dianzi University, China. Her research interests include image processing, computer vision and intelligent transportation system.