Mobility Viewer: An Eulerian Approach for Studying Urban Crowd Flow

Yuxin Ma, Tao Lin, Chen Li, Zhendong Cao, Fei Wang, and Wei Chen

Abstract—Studying human movement citywide is important for understanding the mobility and transportation patterns. Rather than investigating the trajectories of individuals, we employ an Eulerian approach to analyze the crowd flows among a geographical network and a social network, which are extracted from the mobile phone data. We design a suite of visualization techniques to illustrate the dynamic evolutions of the flow over the networks. We contribute the design and implementation of a visual analytics system, called MobilityViewer, that supports situation-aware understanding and visual reasoning of human mobility. We exemplify our approach with a real citywide dataset of 7 millions users in two months.

Index Terms—Human Mobility, Visual Analysis, Data-driven Intelligent Transportation System

✦

1 INTRODUCTION

HUMAN mobility data has been studied for many years. Researchers have proposed methods to study the human mobility records and achieved a deep understanding of human behaviour in physical space [1], [2]. The insights obtained from the datasets are used extensively for discovering city dynamics, evaluating public policies and improving urban planning [3], [4], [5], [6].

In this paper, we examine a particularly valuable dataset that contains both crowd mobility and social ties: mobile phone records. The dataset contains the records of several millions of mobile phone users of a mid-sized city, including the locations of users on the basis of the cell towers, the mobile phone calls, and profiles (ID, coordinates, semantics) of the cell towers. While the human mobility information is not geographically accurate, it does provide a daily observation for the urban crowd flow. Social networks among the users can be extracted from the phone calls. Thus, physical dynamics and social connections of populations in the city can be studied, e.g. the crowd flow patterns among cell towers as well as the social ties of users. In addition, the interconnections among users derived from the calling network also facilitates the capability of studies on social transportation [7], [8], [9].

In analyzing the data, one major challenge is to enable the analyst to understand and gain insights from the crowd flow between cell towers. For example, by monitoring and exploring the crowd flow, the analysts in the city government can discover the area with high population density or mobility, and then determine what is happening in the area and make plans about the event. Unlike the GPS data in which the trajectories are collected continuously at the geographical resolutions of meters [10], [11], the cell tower-based location records have a low and inconstant accuracy ranging from hundreds of meters to several kilo-meters, making the analytical algorithms that rely on accurate locations infeasible [12], [13]. From the perspective of motion representation, the study of trajectories utilizes the Lagrangian description of human mobility, which focuses on the trace of movement. However, for the cell tower-based record data we explore the crowd flow in the Eulerian view. The rationale of our choice is that the accurate trajectories is difficult to retrieve from the dataset, thus we utilize the cell towers to display the flow patterns. In this scenario, each cell tower can be considered as a fixed observation point, and the flow patterns are depicted on these observation points.

In addition to studying the flow patterns, there is a dire need to utilize the mobile phone calls to analyze the correlations between the social relations of a crowd and their physical movements. Previous studies [13], [14] on connecting social and physical spaces have found that the geography in a city influences the social ties of the mobile phone users. To a certain degree, existing methods mainly focus on extracting the correlations and predicting the social relations among phone users in daily life with automated algorithms, while less effort is conducted to provide an interactive exploration approach to empower the analyst find subtle and latent dynamics hidden in the data. For instance, the correlation patterns between the mobile phone network and the crowd flow can reflect the distribution of residence areas and work places, thus guide the analysts to improve the city planning.

To address these challenges, we have designed a suite of interactive visualization methods with which the intelligence, experiences and inspiration of the analyst can be integrated with analytical algorithms. We believe that equipping the

- *Yuxin Ma, Fei Wang and Wei Chen are with State Key Lab of CAD&CG, Zhejiang University. E-mail: mayuxin@zju.edu.cn, wolflyecn@gmail.com, chenwei@cad.zju.edu.cn*
- *Tao Lin, Chen Li and Zhendong Cao are with College of Computer Science and Technology, Zhejiang University. E-mail:{nblintao,lichen28.work,ryecao}@gmail.com*
- *Wei Chen is also with Cyber Innovation Joint Research Center, Zhejiang university.*

analyst with a visual analytics interface will enable a better understanding of dynamic, random, and interleaved human transportation and communications, as witnessed by numerous work done for visual analysis of traffic data [4], [5], [15], [16], [17], [18]. Specifically, our interface design focuses on three objectives:

- Q1: How to enable the analyst to be better informed of the flow patterns among cell towers in a city, including the flow volume and the temporal patterns?
- Q2: How to depict the flow directions of the cell towers dynamically?
- Q3: How can the analyst disclose the correlation between social relations and the physical movements of the crowd?

This paper presents our effort in visually assisted knowledge discovery and sense-making from massive, dynamic, and multi-variate data. We contribute the design and implementation of a web-based visual analysis system that supports situation-aware exploration and visual reasoning of crowd flow in an integrated visual interface. We exemplify our approach with case studies on a real citywide urban dataset. The main contributions include:

- A suite of interactive visualization techniques for exploring crowd flow volumes, directional dynamics and spatio-temporal patterns;
- A visual analysis approach for inspecting human social relations and their physical movements.

The remaining sections are organized as follows. Related work is covered in Section 2. Section 3 presents an overview of the dataset and our analytical pipeline, followed by Section4 where data pre-processing procedures are presented. The next section describes the interface design and visual analysis methods. Section 6 demonstrates the effectiveness of our solution through case studies. We discuss related issues draw conclusions in Section 7.

2 RELATED WORK

The work presented in this paper is related to three broad topics: 1) trajectory visualization, 2) traffic visualization, and 3) mobile phone data analytics and visualization.

2.1 Trajectory Visualization

A spatial *trajectory* is "a trace generated by a moving object in geographical spaces"[19]. It is usually recorded and represented by a sequence of temporal ordered points. In the field of visualization, trajectory data has been widely studied in recent years. In [3], the authors classified the visual analytics techniques into three main categories: direct depiction, summarization and pattern extraction. Scheepens et al. [20] presented a density-based visualization technique to show an aggregate overview of trajectories. VATT (Visual Analytics of Taxi Topics) system [21] was designed to discover the trajectory patterns of taxi. Trajectories are grouped into taxi topics to facilitate exploration of grouping patterns by introducing the Latent Dirichlet Allocation (LDA). Liu et al. [22] proposed a trajectory visualization technique for taxi trajectory data to examine the source/destination pairs that

contain highly diverse routes. Tominski et al. [23] presented an 3D approach to visualize attributes of trajectories. The T-Watcher system [11] allowed for visualization and analysis of complex traffic situations and patterns in taxi trajectories. Zeng et al. [24] presented the Interchange Circos Diagram, a visual representation of interchange patterns contained in massive trajectory dataset. Ferreira et al. [6] modeled a visual trajectory querying scheme to support origin-destination queries of taxi trips.

2.2 Traffic Visualization

Traffic is the flux or passage of motorized vehicles, unmotorized vehicles, and pedestrians on roads, or the movement of passengers or people [25]. Depending on the tasks of the analysis methods, the traffic data visualization can be summarized into the following categories[26]: visual scanning of traffic situations, pattern extraction, context-aware exploration.

For visual scanning of the traffic data, events can be retrieved from the real-time data. Piringer et al. [27] addressed the importance of video surveillance data in road tunnels and designed AlVis which enables visual presentation of the spatio-temporal development of scenarios in real-time.

For the task of pattern extraction, the mobility patterns of objects are discovered from the traffic data which characterize the movements, evolutions and relations with other objects. The TripVista system [28] supported the investigation and analysis of microscopic traffic patterns and abnormalities. Andrienko et al. [29] suggested a visual analysis approach to extract clusters from complex trajectories. First a small subset is retrieved from the big dataset, and the analyst performs clustering algorithm on the subset. Then a classifier is built on the clustering result to attach new trajectories to the existing clusters. Clustering and classification results can be visually inspected and refined.

For context-aware exploration task, the system is required to support querying, exploration, reasoning of the traffic situation. Wang et al. [30] designed a visual analysis system for on-demand complex topological queries of trajectories based on a dynamic road-based trajectory query model and a bi-directional linked hash index scheme.

2.3 Mobile Phone Data Analytics and Visualization

Gao et al. [14] proposed an agglomerative clustering algorithm to explore the interpret the patterns in the phone call interactions as well as the mobile phone users' movement. The method utilized the geographical context of mobile phone cell towers and designed an alternative modularity function incorporating a gravity model which is inspired by the Newman-Girvan modularity metric. The relations between social interactions and movement of phone users are further studied in [13], where a similarity measure for mobility and predictability of the movement was introduced. Zhang [12] presented a comprehensive description of the mobility patterns extracted from cellular data traffic network and found that the cellular data network records can provide a finer granularity of location and movement. In [31], [32] it is shown that the predictability of human mobility is high by using call detailed

records (CDRs) from the cellular network. The accuracy of predicting a mobile phone user's trajectory can be 93% by measuring the trajectory entropy. [33] designed a visual analysis system to visualize the phone call records and support the analysis of mobility patterns.

3 OVERVIEW

3.1 Data

The dataset is collected from a city with 14 million citizens, which contains three parts: cell tower profiles, trajectory records, and call detailed records (CDRs). The cell tower profile is listed in Table 1. The Location Area Code (LAC) is used as an identification of a set of cell towers which are grouped together to optimize signalling, and the Cell ID is a unique number to identify a cell tower within a location area [34]. Hence, a cell tower in the city is uniquely identified by its <LAC, Cell ID>. The function type of a cell tower is a semantic label indicating the function type of the area around the cell tower, such as "Residential Area", "Business District", "Industrial Estate", etc. The trajectory data comprises a unique mobile phone user ID, a timestamp and a Cell ID. A record is stored once a mobile phone user enters the cover range of a cell tower. The CDRs contain a timestamp, a calling user ID, a called user ID and the connected Cell IDs of the two mobile phone users when the phone call starts. In the cell tower profiles, there are over 28,000 individual cell towers distributed in the city, and the locations are shown in Figure 1. The trajectory records were collected from Dec. 16, 2013 to Dec. 22 and from Jan. 14, 2014 to Feb. 27, 2014. The number of trajectory records are over 14 billion with over 7 million phone users, and the size is about 1.8 terabytes. The date of CDRs covers from Dec. 16, 2013 to Dec. 22, 2013 and contains more than 4.5 million records in each day.

TABLE 1
The cell tower profile.

Field	Description
Location Area Code (LAC)	Identification of the location area
Cell ID	Identification of the cell
Latitude	The latitude coordinate of the cell
Longitude	The longitudinal coordinate of the cell
Function type	City function type of the surrounding area

Some preliminary statistics of the trajectory and CDRs are performed based on the data on one day. Figure 2 (a) shows the number of trajectory records in every 10 minutes on Jan. 14, 2014. We can find that the number of records increases in the morning hours indicating the start of activities in the morning. During the work time (9 a.m. ∼ 12 a.m. and 2 p.m. ∼ 5 p.m.), the quantity remains at a relatively high level, while from about 8 p.m it begins to fall. A similar rise and fall trend is found in the figure for CDRs (Figure 2 (b)).

From the trajectory records, the flow volume of a specified cell tower can be derived by simply calculating the summation of inward and outward transitions of the tower.

Fig. 1. The distribution of the cell towers indicates the density of population citywide.
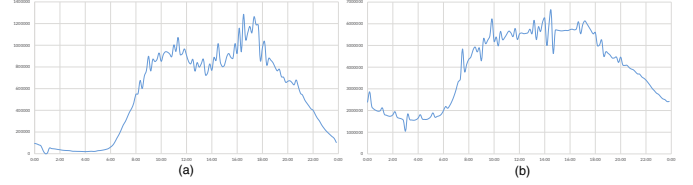

Fig. 2. (a) The distribution of the number of trajectory records on Jan. 14, 2014. (b) The distribution of the number of CDRs on Dec. 18, 2013.

3.2 Design Considerations

Before crafting the analysis pipeline and visual design, we sought principles to guide our design process. After studying the preliminary statistics and analysis of the dataset, we arrived at the following considerations:

- **Simplification by summarization** As the volumes of trajectory records and CDRs are extremely large, it is infeasible and unnecessary to depict every single records in the views, thus simplification should be performed before visualizing the records.
- **Spatio-temporal analysis** The trajectory records contains both locations and timestamps, thus the two main properties should be considered for visual design.
- **Combination of multiple datasets** The information and relationship between the flows and the social relations should be presented in the views.

3.3 Pipeline

Our analysis pipeline contains two stages, as shown in Figure 3. In the data pre-processing stage, the data is cleaned and used to compute the time-varying crowd flow among cell towers. The details of the pre-processing stage is described in Section 4. In the visual exploration stage, four steps are iteratively performed:

Flow Volume Analysis provides an intuitive summarization of the flow volumes in each cell tower. We employ a heatmap, a glyph-based sunburst view and a 3-D terrain view to visualize the flow volumes in multiple perspectives and allow for visual exploration and comparison..

Flow Link Analysis focuses on the flow exchange patterns between cell towers. We apply a dynamic density-based visualization method to display the direction and the quantity of
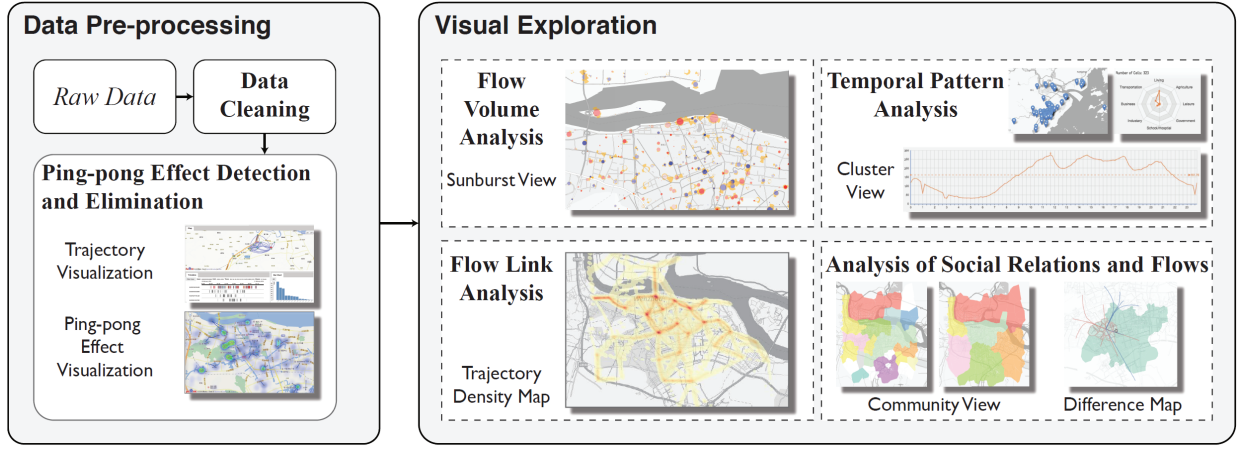
Fig. 3. The visual analysis pipeline of our approach.

the flows between pairs of cell towers. The analyst can explore the flow patterns of the mobile phone users in a specified range of time.

Temporal Pattern Analysis is used for presenting the time-varying patterns of flow volumes on each cell tower along time. For the cell towers in the city, the analyst can compute the correlations of the flow volume series between pairs of cell towers, and then perform clustering algorithm by utilizing their correlation. Illustrating the clusters of cell towers visually reveals the general temporal patterns and the abnormality. The geographical relations of the clusters can also be visually investigated.

Analysis of Social Relations and Flows facilitates discovering the social communities of the mobile phone users based on the CDRs and investigating the relations of trajectories among community members. The analyst can visually investigate a specified community and summarize the commonalities and abnormality of the mobile phone users.

The design of our analytical process is inspired by the visual information-seeking mantra ("Overview first, zoom and filter, then details-on-demand." [35]). In the Cell Flow Volume Analysis step and Flow Link Analysis step, the analyst can obtain an overview of the flow volumes and the flow link patterns among cell towers. In exploring detailed patterns for specified cell towers, the Temporal Pattern Analysis step supports the filtering of cell towers with similar characteristics and visualization methods to display the details of cell towers. The Analysis of Social Relations and Flows step proposes another filtering and exploration strategy by using the social relations of the mobile phone users.

4 DATA PRE-PROCESSING

In the trajectory records and CDRs there are several types of dirty records, demanding a cleaning and organization pre-process.

- Missing values: records containing missing fields.
- Invalid values: records with Cell IDs which are not listed in the cell tower profile or with meaningless timestamps.
- Duplicated records: records which are completely equal to each in all fields.

After cleansing and deduplication of the dirty records mentioned above, about 8 percent of the records are deleted from the dataset.

Then we focus on the detection and elimination of ping-pong effects. The trajectory records of an individual mobile phone user forms a sequence of Cell IDs with timestamps. However, when the user stays in a region covered by signals from multiple cell towers, the mobile phone may handoff among those cell towers frequently but not steadily connect to a single one. This phenomenon is called ping-pong effects [36], [37], which is illustrated in Figure 4 (a). Under this circumstance, multiple redundant records are generated even though the mobile phone user remains unmoved.

Our idea of detecting and removing the ping-pong effects is inspired by the n-gram representation for text analysis. We sort the trajectory records by timestamp and line up the cell tower of each record to form a sequence of Cell IDs. Then we check all the n-grams in the Cell ID sequence. If a consecutive set of n-grams which only contains n different cell towers and the interval of corresponding timestamps are short adequately, it can be considered that the ping-pong effect happens in this subsequence. Algorithm 1 illustrates the detection and elimination procedure.

We provide an interactive visual exploration interface of the trajectory records illustrated in Figure 4 (b). The record view on the lower left side depicts the mobile phone users' trajectory records along the time axis. The analyst can filter and rank the users by date and the number of records. When a specific user is selected in the record view, the bar chart view on the lower right side displays the number of cell towers in his/her trajectory records from the highest to the lowest. The trajectory of the selected user is shown as a series of line segments in the map view as well.

Finally we slice one day into intervals with equal lengths and aggregate the transitions of mobile phone users in each interval to derive the flow volume. The structure of the transitions is essentially a dynamic graph, and is stored as a series of adjacency lists in our implementation. The length of the interval can be adaptively selected on the basis of the requirement of the details for analysis and computational cost.
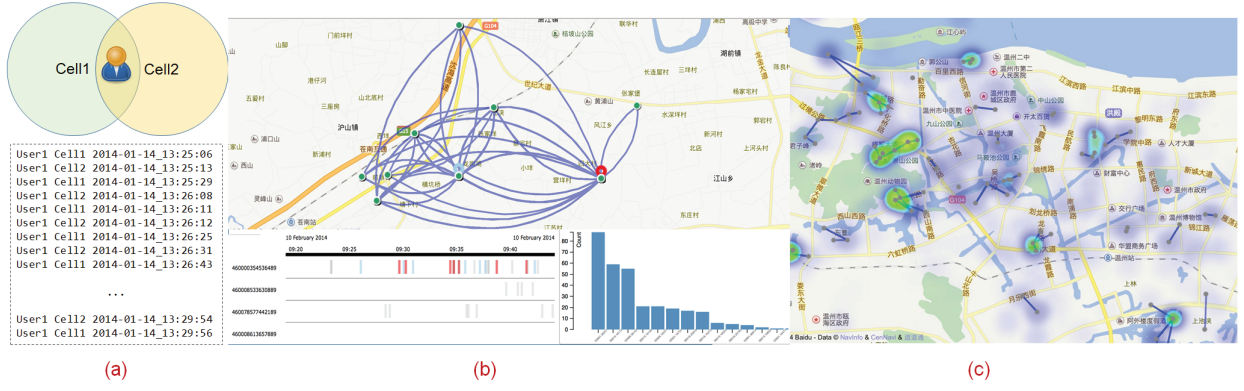
Fig. 4. (a) Illustration of the ping-pong effect. (b) Visualizing trajectory records. (c) Heat map visualization of the ping-pong effect frequency.

Algorithm 1 Detection algorithm of the ping-pong effect

Require:

The list of trajectory records with length m, T;

The length of n-gram, n;

Maximum duration between two consecutive n-grams, t_{max};

Ensure:

A set of ping-pong effect subsequences in T, P;

1: $P = \emptyset$;
2: $p_{start} = 1, p_{end} = n$;
3: $S = \{T_1.Cell\_ID, T_2.Cell\_ID, \ldots, T_n.Cell\_ID\}$;
4: **for** $i = 2$ to $m - n - 1$ **do**
5:     $S_{n\text{-gram}} = \{T_i.Cell\_ID, \ldots, T_{i+n-1}.Cell\_ID\}$;
6:     **if** $S \subseteq S_{n\text{-gram}}$ **then**
7:         **if** $T_{i+n-1}.timestamp - T_{i+n-2}.timestamp < t_{max}$ **then**
8:             $p_{end} = p_{end} + 1$;
9:         **end if**
10:     **else**
11:         Add the subsequence $[T_{p_{start}}, \ldots, T_{p_{end}}]$ into P;
12:         $S = \{T_{p_{end}+1}.Cell\_ID, \ldots, T_{p_{end}+n}.Cell\_ID\}$
13:     **end if**
14: **end for**

In our paper, we set the length of interval to 10 minutes, namely 144 intervals in a day.

5 VISUAL ANALYSIS

Exploring flow volumes, flow links, temporal patterns and the correlations between social relations and flows are tightly integrated into an interactive visual analysis process.

5.1 Flow Volume Analysis

We utilize a sunburst view to present the inflow, outflow and the distribution of flow directions of a cell tower. A 3-D terrain view is deployed to visualize the flow volumes with the dimension of height.

Sunburst View A cell tower is visually presented as a sunburst glyph shown in Figure 5 (a). The radius of the inner circle encodes the total volume of inflow and outflow and

opacity the absolute value of the difference between the inflow and outflow. The filled color is set to be blue when the inflow is larger than the outflow red when outflow is larger. The equally-placed sectors in the outside layer represent the flows from the corresponding direction, which can be set to inflows or outflows. As only properties on cell towers are utilized, we did not add links between glyphs which may cause visual clutter in the view.
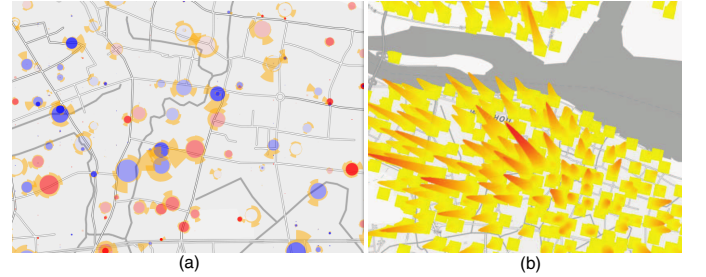

Fig. 5. (a) The sunburst view. Here only top 100 cell towers with the largest flow volumes are displayed. (b) The 3-D terrain view.

3-D Terrain View In this view the flow volumes are represented by the height of the peaks over the cell towers on a street map. First the flow volume values are regularized into the range of $[0, 1]$. Then the map is divided into equally-sized grids, and the value on each grid is estimated by using kernel density estimation based on the flow volume values of all the cell towers to compute a density field on the street map. We transform the density field into a height field. A threshold is assigned to filter out the grids with very low height values and set them to be zero. Finally the surface of the height field is rendered as shown in Figure 5 (b). The analyst can choose to display the inflows or outflows of the cell towers.

5.2 Flow Link Analysis

The flows between pairs of cell towers depict time-varied patterns. We employ a density map-based trajectory visualization method to deliver an intuitive view of the exchange patterns.

Given a specific time t, we compute the position of a mobile phone user by the following interpolation scheme:

1) Retrieve the trajectory record r_{t_1} with timestamp t_1 which is right before time t and cell tower c_1, and the one r_{t_2} with timestamp t_2 which is just after t and cell tower c_2;

2) Compute the speed value $\frac{t-t2}{dist(c_1,c_2)}$ where $dist(c_1,c_2)$ is the distance between c_1 and c_2.

3) If the speed value is lower than $5km/h$, we consider that the user has not left c_1. If not, the user's latitude and longitude are set to be:

$$latitude : lat(c_2) - \frac{(t-t_2) \times 5km/h}{dist(c_1,c_2)}(lat(c_2) - lat(c_1))$$

$$longitude : lng(c_2) - \frac{(t-t_2) \times 5km/h}{dist(c_1,c_2)}(lng(c_2) - lng(c_1))$$

where $lat(\cdot)$ and $lng(\cdot)$ are the latitude and longitude of the cell tower, respectively.

Essentially, our interpolation scheme evaluates each user's position during moving from one cell tower to another. Moreover, the judgement of the speed value in step 2 is designed to avoid the situation in which a user is assumed to keep on moving in a very long time period. For example, if a user has a record in cell tower A at 8 a.m. when he/she arrives at the office, and another record in cell tower B at 12 a.m. going for lunch, it is unreasonable to consider that the user continuously moves from A to B during the four hours. Hence we assume that the user walks to cell tower B with the preferred walking speed of humans, i.e. 5km/h [38], and derive the leaving time reversely.

Once the positions of users are calculated, a density map is rendered onto a street map based on the kernel density estimation result of the positions, which is shown in Figure 6. Within a selected range of time, the density maps are generated in each consecutive 10-minute time interval and rendered frame by frame dynamically. The rationale of choosing density map is that it can present both the directions and the flow volumes among cell towers. In Addition, the temporal patterns can be highlighted by creating animation with the series of density maps generated from consecutive time intervals. To avoid visual clutter in the region with high cell tower density, the density maps are set to have multiple resolution levels. All cell towers are depicted in the lowest level. In higher zoom levels, the cell towers are clustered by using the k-means algorithm based on their spatial coordinates with different k. The analyst can obtain an overview at a high zoom level and zoom in to observe the local flow patterns.

5.3 Temporal Pattern Analysis

As described in Section 4, for each cell tower there are 144 samples of flow volumes in a single day. Here the samples are considered as a time sequence to describe the flow patterns. We first conduct clustering of these time sequences and then visualize the clusters on the map.

5.3.1 Clustering of the Flow Sequences

To perform the clustering algorithm, the similarities between the time sequences are derived. Given a specified day, we apply the Pearson correlation coefficient as the similarity metric. Thus an $m \times m$ similarity matrix is computed where m

Fig. 6. Illustration of the flow links with the density map. The red links indicate high flow volumes between the regions. Note that the cell towers are clustered with $k = 45$.
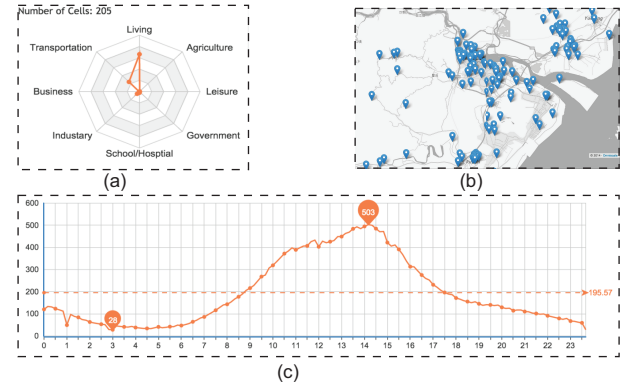

Fig. 7. Views for temporal pattern analysis: (a) the distribution of cell types, (b) the map view and (c) the time sequence view.

is the number of cell towers. Then spectral clustering is done by using the similarity matrix. Additionally, the analyst can choose to use the volumes of inflow or outflow.

5.3.2 Visualization of the Flow Sequences and Clusters

The clustering result is displayed as a list in the interface. The analyst can select a cluster in the list, and in the map view (Figure 7 (b)) the cell towers that in the selected cluster will be highlighted to enable the analyst to explore the spatial distribution. Additionally the distribution of cell types are displayed. In the time sequence view (Figure 7 (c)), a line chart of the average flow volume sequence is displayed as a dash line presenting a summary of the cluster. When hovering on a cell tower point in the map view, the corresponding time sequence is displayed in the time sequence view with a solid line. The colors of the solid line and the dash line represent whether inflow or outflow is selected in the clustering step (orange for inflow and blue for outflow). Moreover, multiple clusters can be activated and displayed simultaneously with different colors in the time sequence view and the map view. The analyst can compare the distributions of the clusters and analyze the relations of the rise and fall patterns among clusters.

5.4 Correlation Analysis of Social Relations and Flows

5.4.1 Community Detection of Cell Towers

The communities extracted from the CDRs represent the social groups of the mobile phone users. In our work we only focus on the spatial distribution of the community. Based on [14], we adopt an enhanced community detection algorithm that incorporates the call records between cell towers and the population on each cell tower. The analyst can set a time period $[t, t']$, and then a weighted directed graph of cell towers is created with the CDRs in which the weights are the number of calls, and the population of cell towers are derived from the positions of all the mobile phone users at time t'.

5.4.2 Community View

The community view supports deductive exploration. To visualize the communities of the cell towers, the map is partitioned into Voronoi regions based on the location of cell towers, and the regions in the same community are painted with the same color (Figure 8 (a)). Thus by checking the region colors the analyst can understand the spatial distribution of the communities. In addition, the changes of the community distributions in two different time periods may relate to the movement of mobile phone users, hence we design a difference map to display the changes shown in Figure 8 (b). The analyst can select two time intervals $[t_1, t'_1]$, $[t_2, t'_2]$, and the corresponding Voronoi regions of cell towers in different communities between $[t_1, t'_1]$ and $[t_2, t'_2]$ are highlighted with green color. When selecting a region in the difference map, the flows from time t_1 to t'_2 are presented as a node-link diagram where edge-bundling is applied (Figure 8 (c)), enabling the comparison of the user's movement and evolution among regions.
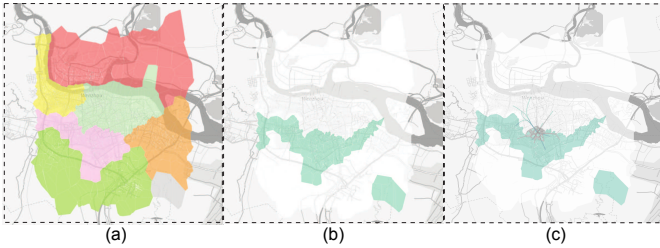

Fig. 8. The community view. (a) The distributions of communities. (b) The difference map. (c) Highlighted region with a red boundary. The bundled blue and red lines indicates the inflows and outflows among other regions.

6 CASE STUDIES

We conduct four case studies to respectively verify the four visual exploration steps. First we check the global flow volume patterns in the city. Then we analyze the patterns of flow links, followed by the exploration of temporal patterns. Finally we extract communities by using the CDRs and investigate the evolution of the community structure with movement of mobile phone users.

6.1 System Implementation

Our web-based system is primarily implemented in JavaScript for front-end UI, which employs OpenStreetMap as the street map library, D3.js as graphic rendering library, jQuery UI for user interface components and Backbone.js as the MVC framework. For back-end computational support, we design a RESTful interface for communication built on Django Web Framework and employ Apache Spark as the data processing engine.

6.2 Case 1: Exploring Flow Volumes

In this case, we use the overview to explore the flow volumes on cell towers. We choose three typical times on the following days: Dec. 18, 2013, Jan. 31, 2014 and Feb. 10, 2014. Note that Jan. 31, 2014 is the day of the Chinese new year when most of the natives go home and take a vacation. Figure 9 (a) compares the flow volumes in the sunburst views. We can notice that on Jan. 31 the radius of the inner circles is generally smaller than those in another two days, which indicates less daily activities of the citizens. Then we specifically investigate the two cell towers with large volumes. In Figure 9 (b) we find that for each of the two cell towers in blue rectangles the direction of its largest outflow is towards the river. In the corresponding satellite image a ferry crossing is presented in the red rectangle, which is not shown in the street map.

6.3 Case 2: Analyzing Flow Links

Next we utilize the flow link density map to investigate the crowd flow among cell towers. Figure 10 shows the density map that are generated by aggregating the flow links in every 10 minutes in a period from 2:00 pm to 4:00 pm, Jan.15, 2014. The resolution of cell towers increases from (a) to (c) in Figure 10. We start with (a), an overview of the entire city area where the cluster number of cell towers are set to be 100. In the downtown area (blue rectangle) we find that the density of the crowd flow is higher than other regions. In region A and B in Figure 10 two salient patterns are shown, thus we further explore the flow links by zooming in the viewport (represented by Figure 10 (c)). In region A, there are two cell towers presenting high interchanging flows. After looking up a detailed map, it is shown that in region A the two cell towers lie in an exit of *G330* National Highway. In region B the radial pattern shows high crowd density in the center area in which the coach station of this city is, thus the population density and the crowd flow should be high.
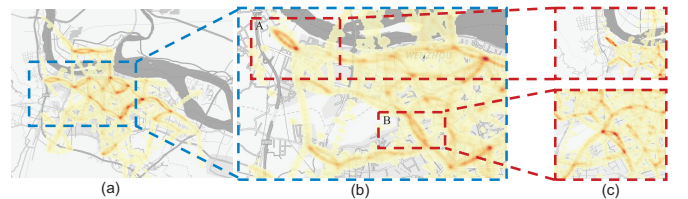

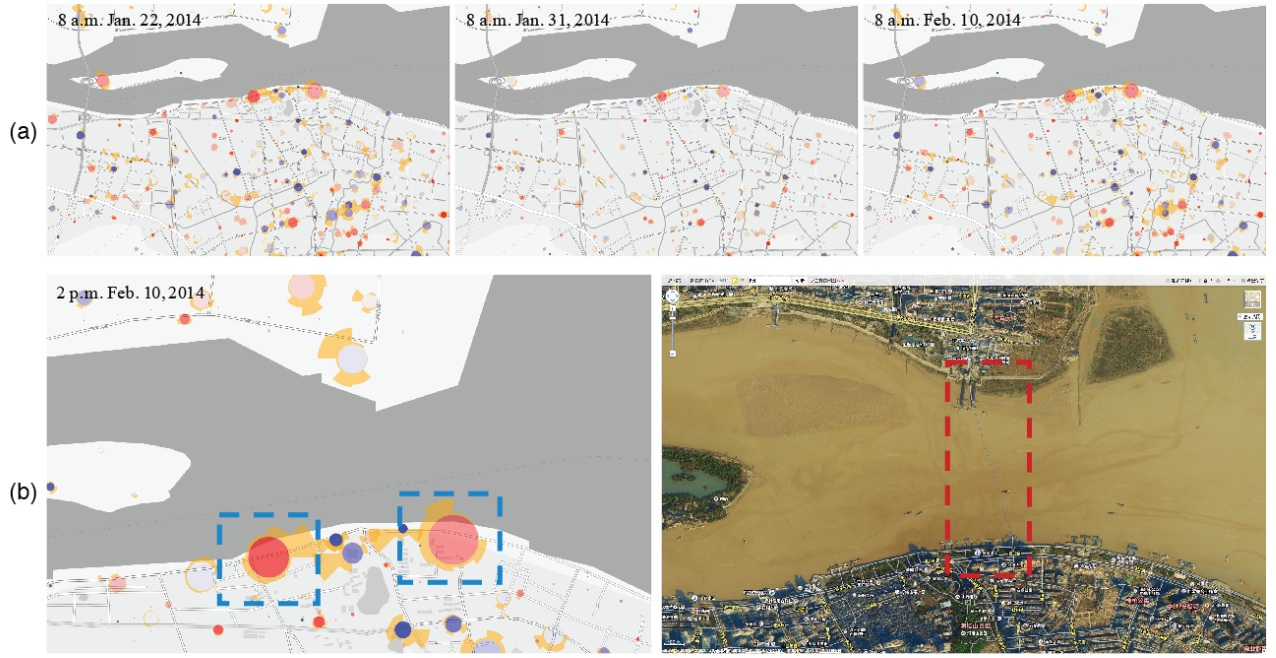Fig. 10. Case 2: Flow links in different levels from (a) to (b) and (c).

Fig. 9. Case 1: (a) The sunburst views for three different days (b) The sunburst view for a ferry area and the corresponding satellite image.
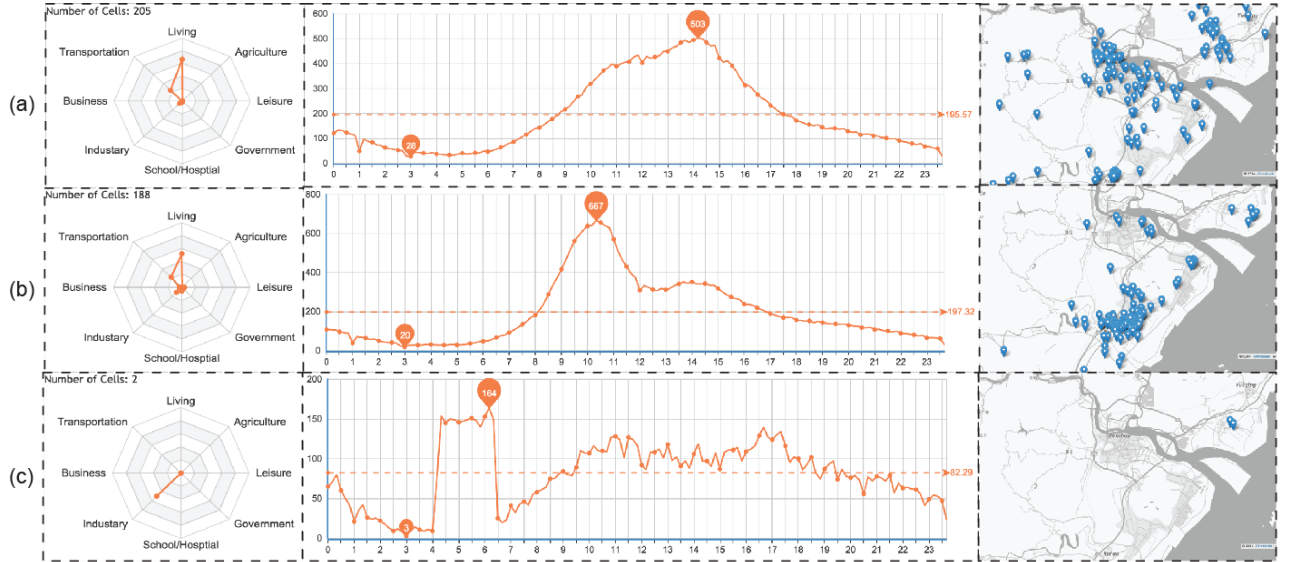

Fig. 11. Case 3: Three typical temporal flow patterns. (a) A general rise-and-fall pattern. (b) A pattern distributed in a small town. (c) An outlier in which the human activity is raised from 4 a.m to 6 a.m.

6.4 Case 3: Analyzing Temporal Patterns

Further we utilize the views for temporal pattern analysis to investigate human activities in the temporal space. By common sense it is expected that the human activity should be raised in the morning, remain stable on the daytime and reduce in the evening. In this section we study if it is the case. After performing clustering algorithm on the flow sequences on Jan. 30, 2014 with $k = 50$, the cell towers are divided into 50 clusters. We choose three typical clusters in the result, as illustrated in Figure 11. The first cluster (Figure 11 (a)) shows a general pattern in most cell towers which depicts the

abovementioned rise-and-fall pattern in the line chart. In the radar chart it can be found that most of the cell towers lie in living districts and transportation districts. In Figure 11 (b), a cluster of cell towers that locate in a small town is presented. In the corresponding line chart, the flow drops down deeply after 12 o'clock. Furthermore, the cluster with two outliers shown in Figure 11 (c) indicates that the flow grows up quickly at 4:00 am and drop down at 6:00 am, which might be a local event near the cell towers. We infer that the event may be related to industrial production because in the radar chart both of the cell towers are located in an industrial area.
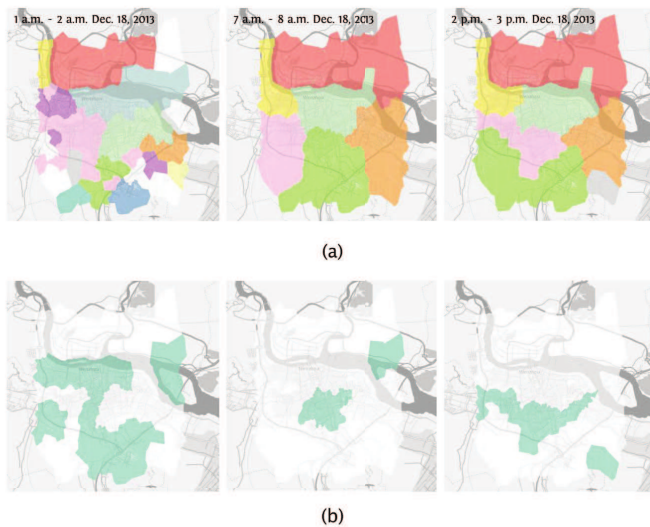
Fig. 12. Case 4: Community structures in different times: (a) The communities of cell towers in three specified hours. (b) The corresponding difference maps.

6.5 Case 4: Analyzing the Relation between Social Ties and Mobility

In this case, we study the community structure of the phone call network and how crowd flow influences the spatial distribution of the CDRs. We take the CDR and trajectory data of Dec. 18, 2013 as the test data. We aggregate the CDRs in each hour and compute the communities. Figure 12 (a) displays communities of cell towers in three separated hours and the difference maps from the maps in previous hours. From the perspective of community numbers, in the time interval of 1 a.m. ~ 2 a.m., the fragments of communities are significantly more than other two hours.

7 DISCUSSION AND CONCLUSION

In this paper, we propose an Eulerian approach for studying urban crowd flow. The multiple visualization views support the visual exploration the flow volume and direction for each cell. Meanwhile, a visual-enhanced analysis method is dedicated to discover the temporal patterns and correlations of cells. We study the relations between the human social relations and their flow for the purpose of connecting the social information and physical movements.

One promising extension of our work is to integrate four analysis steps in one view. Currently our system provides an isolated view for each method. We also expect to integrate network flow mining approach with our system.

REFERENCES

[1] G. Andrienko, N. Andrienko, C. Hurter, S. Rinzivillo, and S. Wrobel, "Scalable analysis of movement data for extracting and exploring significant places," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 7, pp. 1078–1094, 2013.

[2] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: Concepts, methodologies, and applications," *ACM Transaction on Intelligent Systems and Technology*, vol. 5, no. 3, October 2014.

[3] G. Andrienko, N. Andrienko, J. Dykes, S. I. Fabrikant, and M. Wachowicz, "Geovisualization of dynamics, movement and change: Key issues and developing approaches in visualization research," *Information Visualization*, vol. 7, no. 3, pp. 173–180, Jun. 2008. [Online]. Available: http://dx.doi.org/10.1057/ivs.2008.23

[4] W. Zeng, C.-W. Fu, S. M. Arisona, A. Erath, and H. Qu, "Visualizing mobility of public transportation system," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, 2014.

[5] Z. Wang, M. Lu, X. Yuan, J. Zhang, and H. v. d. Wetering, "Visual traffic jam analysis based on trajectory data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2159–2168, 2013.

[6] N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva, "Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2149–2158, 2013.

[7] W. Tang, H. Zhuang, and J. Tang, "Learning to infer social ties in large networks," in *Machine Learning and Knowledge Discovery in Databases*. Springer, 2011, pp. 381–397.

[8] D. Lian, C. Zhao, X. Xie, G. Sun, E. Chen, and Y. Rui, "Geomf: joint geographical modeling and matrix factorization for point-of-interest recommendation," in *Proceedings of ACM SIGKDD*. ACM, 2014, pp. 831–840.

[9] J. Chae, D. Thom, H. Bosch, Y. Jang, R. Maciejewski, D. S. Ebert, and T. Ertl, "Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition," in *IEEE Symposium on Visual Analytics Science and Technology*, 2012, pp. 143–152.

[10] Z. Liao, Y. Yu, and B. Chen, "Anomaly detection in GPS data based on visual analytics," *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*, pp. 51–58, Sep. 2010.

[11] J. Pu, S. Liu, Y. Ding, H. Qu, and L. Ni, "T-watcher: A new visual analytic system for effective traffic surveillance," in *IEEE 14th International Conference on Mobile Data Management (MDM)*, vol. 1, 2013, pp. 127–136.

[12] Y. Zhang, "User mobility from the view of cellular data networks," in *INFOCOM, 2014 Proceedings IEEE*. IEEE, 2014, pp. 1348–1356.

[13] J. L. Toole, C. Herrera-Yaque, C. M. Schneider, and M. C. Gonzalez, "Coupling human mobility and social ties," *Journal of The Royal Society Interface*, vol. 12, no. 105, pp. 20 141 128–20 141 128, Feb. 2015.

[14] S. Gao, Y. Liu, Y. Wang, and X. Ma, "Discovering Spatial Interaction Communities from Mobile Phone Data," *Transactions in GIS*, vol. 17, no. 3, pp. 463–481, May 2013.

[15] H. Doraiswamy, N. Ferreira, T. Damoulas, J. Freire, and C. T. Silva, "Using topological analysis to support event-guided exploration in urban data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 2634–2643, 2014.

[16] J. Zhang, Y. E, J. Ma, Y. Zhao, B. Xu, L. Sun, J. Chen, and X. Yuan, "Visual analysis of public utility service problems in a metropolis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, 2014.

[17] D. Guo and X. Zhu, "Origin-destination flow data smoothing and mapping," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, 2014.

[18] Z. Wang, T. Ye, M. Lu, X. Yuan, H. Qu, J. Yuan, and Q. Wu, "Visual exploration of sparse traffic trajectory data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 1813–1822, 2014.

[19] Y. Zheng and X. Zhou, *Computing with spatial trajectories*. Springer Science & Business Media, 2011.

[20] R. Scheepens, N. Willems, H. van de Wetering, and J. J. van Wijk, "Interactive visualization of multivariate trajectory data with density maps," in *IEEE Pacific Visualization Symposium*, 2011, pp. 147–154.

[21] D. Chu, D. A. Sheets, Y. Zhao, Y. Wu, J. Yang, M. Zheng, and G. Chen, "Visualizing hidden themes of taxi movement with semantic transformation," in *IEEE Pacific Visualization Symposium*, 2014, pp. 137–144.

[22] H. Liu, Y. Gao, L. Lu, S. Liu, H. Qu, and L. M. Ni, "Visual analysis of route diversity," in *IEEE Conference on Visual Analytics Science and Technology*, 2011, pp. 171–180.

[23] C. Tominski, H. Schumann, G. Andrienko, and N. Andrienko, "Stacking-based visualization of trajectory attribute data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2565–2574, 2012.

[24] W. Zeng, C.-W. Fu, S. M. Arisona, and H. Qu, "Visualizing interchange patterns in massive movement data," in *Computer Graphics Forum*, vol. 32, no. 3pt3. Wiley Online Library, 2013, pp. 271–280.

[25] "Traffic," http://en.wikipedia.org/wiki/Traffic_(disambiguation).

[26] W. Chen, F. Guo, and F.-Y. Wang, "A survey of traffic data visualization," *Intelligent Transportation Systems, IEEE Transactions on*, vol. PP, no. 99, pp. 1–15, 2015.

[27] H. Piringer, M. Buchetics, and R. Benedik, "Alvis: Situation awareness in the surveillance of road tunnels," in *IEEE Conference on Visual Analytics Science and Technology*, 2012, pp. 153–162.

[28] H. Guo, Z. Wang, B. Yu, H. Zhao, and X. Yuan, "Tripvista: Triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection," in *IEEE Pacific Visualization Symposium*, 2011, pp. 163–170.

[29] G. Andrienko, N. Andrienko, S. Rinzivillo, M. Nanni, D. Pedreschi, and F. Giannotti, "Interactive visual clustering of large collections of trajectories," in *IEEE Symposium on Visual Analytics Science and Technology*, 2009, pp. 3–10.

[30] F. Wang, W. Chen, F. Wu, Y. Zhao, H. Hong, T. Gu, L. Wang, R. Liang, and H. Bao, "A visual reasoning approach for data-driven transport assessment on urban road," in *IEEE Conference on Visual Analytics Science and Technology*, 2014.

[31] H. Zang and J. C. Bolot, "Mining call and mobility data to improve paging efficiency in cellular networks," in *MobiCom '07: Proceedings of the 13th annual ACM international conference on Mobile computing and networking*. New York, New York, USA: ACM Request Permissions, Sep. 2007, p. 123.

[32] C. Song, Z. Qu, N. Blumm, and A. L. Barabasi, "Limits of Predictability in Human Mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, Feb. 2010.

[33] J. Pu, S. Liu, P. Xu, H. Qu, and L. M. Ni, "MViewer: mobile phone spatiotemporal data viewer," *Frontiers of Computer Science*, vol. 8, no. 2, pp. 298–315, Jan. 2014.

[34] "3gpp technical specification ts 23.003 numbering, addressing and identification." [Online]. Available: http://www.3gpp.org/DynaReport/23003.htm

[35] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in *Visual Languages, 1996. Proceedings., IEEE Symposium on*. IEEE, 1996, pp. 336–343.

[36] "Cs 294-7: Handoff strategies," 1996.

[37] H. Xiong, D. Zhang, D. Zhang, and V. Gauthier, "Predicting Mobile Phone User Locations by Exploiting Collective Behavioral Patterns," in *2012 IEEE 9th Int'l Conference on Ubiquitous Intelligence & Computing / 9th Int'l Conference on Autonomic & Trusted Computing (UIC/ATC)*. IEEE, 2012, pp. 164–171.

[38] R. C. Browning, E. A. Baker, J. A. Herron, and R. Kram, "Effects of obesity and sex on the energetic cost and preferred speed of walking," *Journal of Applied Physiology*, vol. 100, no. 2, pp. 390–398, 2006.

Chen Li is an undergraduate student in College of Computer Science and Technology at Zhejiang University, China. His research interests include visual analytics and data mining.

Zhendong Cao is an undergraduate student in College of Computer Science and Technology at Zhejiang University, China. His research areas of interest include Information Visualization and Visual Analytics.

Fei Wang is a Ph.D. student in State Key Lab of CAD&CG at Zhejiang University. His current research focuses on Mobile Data Visualization and Analysis.

Yuxin Ma is a Ph.D. student in State Key Lab of CAD&CG, Zhejiang University. His current research focuses on information visualization and visual analytics.

Tao Lin is currently an undergraduate student in College of Computer Science and Technology at Zhejiang University, China. His research interests include information visualization and visual analytics. More information can be obtained from nblintao.github.io.

Dr. Wei Chen is a professor in State Key Lab of CAD&CG, Zhejiang University. He has published more than 60 papers in international journal and conferences. He served as Steering committee of IEEE Pacific Visualization, Conference chair of IEEE Pacific Visualization 2015, Paper Co-chair of IEEE Pacific Visualization 2013. For more information, please refer to http://www.cad.zju.edu.cn/home/chenwei/.