# RelationLines: Visual Reasoning of Egocentric Relations from Heterogeneous Urban Data

WEI CHEN, Zhejiang University, State Key Lab of CAD&CG, China
JING XIA, Zhejiang University, State Key Lab of CAD&CG and Alibaba Group, China
XUMENG WANG, Zhejiang University, State Key Lab of CAD&CG, China
YI WANG, Zhejiang University, State Key Lab of CAD&CG, China
JUN CHEN, Zhejiang University, State Key Lab of CAD&CG, China
TIANLONG GU*, Guilin University of Electronic Technology, China

The increased accessibility of urban sensor data and the popularity of social network applications is enabling the discovery of crowd mobility and personal communication patterns. However, studying the egocentric relationships of an individual (i.e., the egocentric relations) can be very challenging because available data may refer to direct contacts, such as phone calls between individuals, or indirect contacts, such as paired location presence. In this paper, we develop methods to integrate three facets extracted from heterogeneous urban data (timelines, calls and locations) through a progressive visual reasoning and inspection scheme. Our approach uses a detect-and-filter scheme, such that, prior to visual refinement and analysis, a coarse detection is performed to extract the target individual and construct the timeline of the target. It then detects spatio-temporal co-occurrences or call-based contacts to develop the egocentric network of the individual. The filtering stage is enhanced with a line-based visual reasoning interface that facilitates flexible and comprehensive investigation of egocentric relationships and connections in terms of time, space and social networks. The integrated system, RelationLines, is demonstrated using a dataset that contains taxi GPS data, cell-base mobility data, mobile calling data, microblog data and POI data of a city with millions of citizens. We examine the effectiveness and efficiency of our system by three case studies and user review.

CCS Concepts: • **Human-centered computing** → **Visualization systems and tools**; • **Information systems** → *Database management system engines*;

Additional Key Words and Phrases: Location-based; egocentric relations; visual reasoning; heterogeneous urban data; timeline

This research has been supported by National 973 Program of China (2015CB352503), National Natural Science Foundation of China (61772456, U1609217, 61761136020, U1501252, U1711263), Alibaba-Zhejiang University Joint Institute of Frontier Technologies.
Authors' addresses: W. Chen, J. Xia, X. Wang, Y. Wang, J. Chen Zhejiang University, Hangzhou, Zhejiang, China; T. Gu (corresponding author), Guilin University of Electronic Technology, Gulin, Guangzhou, China.
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2018 Copyright held by the owner/author(s). Publication rights licensed to Association for Computing Machinery.
2157-6904/2018/3-ART39 $15.00
https://doi.org/0000001.0000001

**ACM Reference Format:**
Wei Chen, Jing Xia, Xumeng Wang, Yi Wang, Jun Chen, and Tianlong Gu*. 2018. RelationLines: Visual Reasoning of Egocentric Relations from Heterogeneous Urban Data. *ACM Trans. Intell. Syst. Technol.* 9, 4, Article 39 (March 2018), 23 pages.
https://doi.org/0000001.0000001

## 1 INTRODUCTION

The demographic-based social strategy [28] can be extensively studied by focusing on the micro-level social network structures, such as the egocentric network, social ties, social triads, and communities [5, 23, 33]. In particular, analyzing egocentric networks requires answering questions such as, what are the network properties of a central user or what are the demographic distributions of a central user's friends? However, few public data can answer these question directly.

Fortunately, heterogeneous urban datasets, such as taxi GPS data [6, 8, 20, 39], cell-based mobility data [41], social media data [10, 22, 32, 37, 42–45], and social network data, provide an enormous opportunity for discovering egocentric network, which can be mined to extract the interdependency [48] among individuals who share similar spatial-temporal locations or trajectories. For instance, cell-based mobility data can be used to discover the egocentric relations between mobile users. Taxi data combined with mobile data or social network data can identify relationships between taxi drivers and passengers, and call data provides evidences of direct contact between individuals. Egocentric relations, denoting both direct connections, such as phone calls, and indirect connections, such as co-occurrences, present how persons connect, as well as the degree of connection to each other both socially and geographically.

Despite many pioneering works on social network analysis [5, 23, 33] and urban data analysis [48], mining egocentric networks from urban data is still a challenging task. We identify two main factors that influence the efficiency of analyzing egocentric network.

First, managing and utilizing heterogeneous urban data requires appropriate data management and representation schemes for efficient egocentric network discovery. One essential task for extracting network is to match trajectories in multi-scale non-uniform spatio-temporal datasets. For instance, cell-based mobility data and taxi GPS data are quite different in terms of their temporal and spatial granularities and accuracy. The cell-based mobility data is inaccurate in terms of locations and temporally nonuniform because a record is triggered only when the mobile phone holder changes its location. In contrast, taxi GPS data is precise and temporally dense. As such, studying the associations among such varied data sources is nontrivial.

Second, the reasoning process is full of uncertainty. Human mobility contains various and vague patterns and locations. Cluster analysis approaches have limited ability to provide detailed explanation and description of individualities. Blind classifications may not only neglect important hints, but also magnify the uncertainty, so we decide to start from decoding them based on individual analysis. However, even focusing on the individual egocentric network, this process is still challenging.

On the one hand, while machine intelligence is good at extracting information, like similarity, domain knowledge is essential for deep insight onto the data [26]. For example, inferring home and work places has to consider job types and work patterns. A naive implementation of identifying work and home, where home is extracted based on their locations at 2 a.m., may mis-classify an individual's location. Thus, discovering egocentric network demands the integration of human intelligence like expertise, judgment, and experience to deduct location-related semantics. On the other hand, occasional geographical co-occurrences cause interferences to constructing egocentric network. For instance, residents living in a district may visit a near shopping mall, which lead to a number of occasional geographical co-occurrences. It is challenging to infer the interpersonal relations by solely considering this clue. Yet, if two of them have positional co-occurrences on all weekdays, they are likely to be colleagues. As such, the concrete discovery of the interpersonal relations demands an iterative process of interactively detect potential sets of objects (persons) and progressively filter the candidates by leveraging situational context and posing additional querying conditions.

In this paper, we propose a progressive detect-and-filter scheme for visually reasoning with egocentric network from urban data. Our approach first identifies an individual of interest by specifying query conditions, such as geographical constraints, taxi trajectories or geotags in microblog posts. These queries are used to automatically generate an individual's temporal history and analysts can examine the places an individual has visited as well as their trajectories. The potential social ties of the individual of interest are constructed by interactively manipulating direct and indirect relationships extracted from the phone records or the geographical co-occurrences (the trajectories). By further leveraging the timeline, locations and relations, the egocentric relationships can be progressively refined and explored with a suite of visual interactions including tagging, sorting, filtering and discovering. Our web-based interface, RelationLines, affords a flexible reasoning process to conduct the detect-and-filter scheme, and analyze and verify the closeness between an individual and associated location-based connections. Our contributions include:

- A progressive detect-and-filter scheme that integrates space, timeline and relationships for studying egocentric networks;
- A flexible visual reasoning approach that supports the construction, refinement and verification of egocentric relationships through a customizable rule-based interface.

## 2 RELATED WORK

### 2.1 Visual analysis of urban data

There have been a vast amount of works dedicated to urban computing [12, 46, 47]. Though effective, they often fail to incorporate human intelligence into the analysis process in a progressive fashion. Visual analysis, an ever growing research field, has been demonstrating its superiority in many situations. Human mobility patterns [40] and traffic flow are two popular topics.

The analysis result of human mobility patterns is significant to urban planners, sociologists, etc. Landesberger et al. [34] present a technique, which could extract mobility patterns from adjustable temporal and spatial granularities. Furthermore, the similarity of mobility patterns can be observed based on both MDS layout and the difference view.

From a microscopic perspective, TripVista [15] seeks to explore traffic data to discover traffic patterns and abnormal behaviors. It has the ability to effectively help users analyze the traffic flow at road intersections. For an analyst who is not familiar with the city, the transiting from macro perspective to micro perspective also need assistance. Wang et al. [35] introduce a dynamic road-based trajectory query model, enabling users to query trajectories through brushing on the road and applying spatial and temporal filter. The related interaction design can speed up the process of locating traffic problems. A series of researchers turn their attentions to semantic trajectories [1]. Being similar with map applications, semantic query is friendly to a majority of users. In addition, if the analysis target is individual, we should pay attention to privacy preservation [36].

### 2.2 Egocentric network analysis from urban data

Urban data has been widely used for analyzing social networks [16, 31], with the most common types of data for this purpose being mobile phone records and online social network data. Mining useful and uncertain information from heterogeneous networks can not be completed independently by either machines or humans. Instead, the entire process must utilize iterative interactions between users and machines. Conventional interactions include adding, moving and removing nodes, highlighting related links, filtering based on node degrees and finding the paths connecting two nodes.

To find small groups among relevant persons, we need to detect the subset of a relation network instead of simplifying or removing redundant parts of a larger relationship network [30]. Renoust et al. [29] achieves data abstraction by using a substrate network and a catalyst network to facilitate subsequent interactions. With Refinery [17], users can upvote or downvote items during their evaluation and choose relevant items to focus on. Apolo [7] enables ranking local subsets of nodes in meaningful ways by a rank-in-place feature, enabling effective uncertainty exploration and interaction. Crnovrsanin et al. [9] develop a network visualization system based on the records of user interactions and node importance.

### 2.3 Management of urban data

Urban data includes distinct datasets, like high-dimensional data and spatio-temporal data. Trajectory data is common spatio-temporal data that is widely applied in a variety of analysis scenarios. There are numerous schemes for indexing trajectories, including 3D R-tree, STR-tree (Spatio-Temporal R-Tree), TB-Tree (Trajectory-Bundle tree), k-d tree, etc. [11, 27]. Zheng et al. [48] classify related index approaches into three categories: augmented R-tree, multiversion R-tree and grid based index. Spatial-temporal data or other multidimensional data can be queried faster at the cost of increasing storage or reducing functionality. Many studies have been performed on managing spatio-temporal data. STC (space-time cube) [8, 13] aggregates data according to 2D position and 1D time respectively. STC representations can not only be employed on different topics, such as action data [4] and eye-tracking data [18], but further extnesions to STC representations, such as the space-time-attribute cube [14] and the nanocube [19], have also been developed. Nanocube indexes by calculating counts for aggregated keys. The main advantage of Nanocube is that Nanocube can offer efficient storage and particular querying. Unfortunately, querying down to any individual record remains a challenging problem.

## 3 DETECT-AND-FILTER SCHEME AND TASKS

Social networks based on dyadic relationships are fundamentally important for understanding human sociality [25]. The focus of our work is to discover the dyadic relationships of a central target to other persons. The collection of the detected dyadic relationships naturally composes an egocentric network.

To achieve this goal, we design a detect-and-filter visual reasoning scheme (see Figure 1) that leverages an integrated visual representations for understanding, querying, and inferring the data. Four tasks are accordingly identified (Section 3.2).

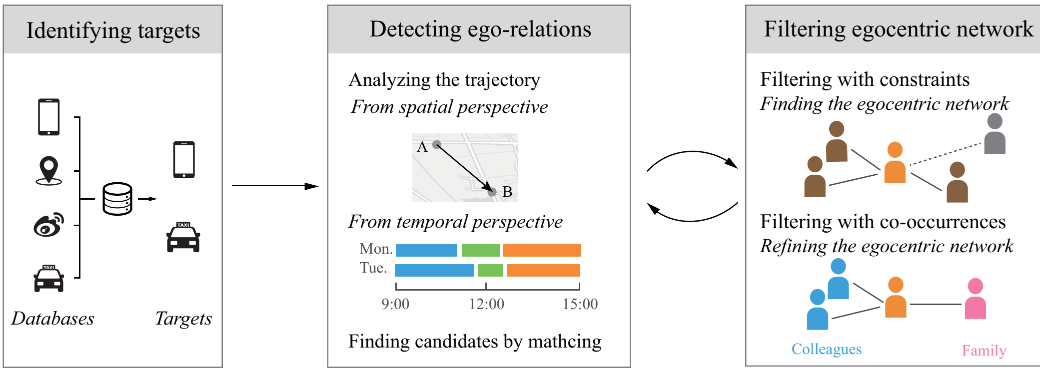



Fig. 1. The detect-and-filter scheme of RelationLines. The scheme starts by identifying targets. It then constructs the trace timeline of the target both on the map and with a timeline chart. The egocentric network of the target can be iteratively constructed, filtered, and refined with the additions of trajectory-based constraints and the co-occurrence information.

### 3.1 The Detect-and-Filter Scheme

A careful investigation of one's interdependency with others depends on an iterative discovery and refinement process concerning the locations, trajectories and phone calls, etc. The refinement and confirmation of the closeness, the potential that they know each other, or find evidences of their mutual understanding, have to be conducted by analysts.

Our detect-and-filter scheme starts from identifying a target (a mobile phone or a taxi) by specifying spot-based constraints derived from heterogeneous data. At the detecting stage, analysts extract candidates that meet some requirements, which could be staying at a spot or following a sequence of meaningful trajectory points with the target in a specific time interval. To define appropriate constraints, analysts need infer the living pattern of a target by observing the trajectory of the target and leveraging other data, like POIs and microblog posts. The candidates found in the detecting stage may have distinct relationships with the target. In the filtering stage, we need to further explore the egocentric relations of the target by progressively refining constraints. Analysts can iteratively modulate new filtering constraints and refine the egocentric relations. For each new metric, individuals of top qualified relations are extracted and kept in the egocentric network. Finally, analysts can study and refine the constructed ego-network by filtering according to the co-occurrences.

### 3.2 Tasks

**Task1 Identifying an individual -** Center to our approach is the detection of the egocentric network of an individual. Analyst starts the analysis from a person. The system needs to provide means to allow for specifying the querying conditions to find a specific person from the input data.

**Task2 Studying the mobility behavior of an individual -** To detect the co-occurrence of pairs of persons, or understand the life style of one person, we need to study the moving behavior of an individual by comprehensively considering the location, time and POIs in the data. These attributes of persons can be used as evidences for building the egocentric network.

**Task3 Constructing an egocentric network -** The most important task is to construct an egocentric network after the target and its mobility behavior are known.

Generally speaking, the construction of the network seeks to generate three kinds of objects: the persons that have social ties with the target (nodes), the tie types between pairs of persons (edges), and the tie strength (the weights of edges).

**Task4 Studying and refining an egocentric network -** The egocentric network in Task3 is constructed by generated constraints induced from Task2. The detected social ties, the ego-network and egocentric relations can be further studied and refined.

# 4 DATA ORGANIZATION

## 4.1 Data

The underlying data includes a collection of taxi trajectories, cell-based station data, mobile calling data, POI data and microblog data of a large city.

**Cell-based mobility data -** The cell-based mobility data contains 600 million records of more than 5 million citizens over 9303 cell stations from 20 Jan 2014 to 28 Jan 2014. The total size of the data is 34 gigabytes. The cell ID and the phone number are recorded whenever a mobile phone enters or leaves the range of a cell station. However, the records are not geographically accurate. First, a phone is not necessarily connected to its nearest cell station, because the connection has multiple conditions: cell intensity, signal priority (4G signal for 4G mobiles), and the current capacity of the cell stations. Second, the locations of stationary cell stations are not the exact locations of a mobile phone. Also, cell stations in downtown regions are much more intensively distributed than in remote regions or mountain regions. In addition to the inaccuracy issue, mobile phone users may shut their phones or leave their phones somewhere for a long time.

**Mobile calling data -** The mobile calling dataset contains 550 million calling records from 15 Dec 2013 to 22 Dec 2013. Each record contains a caller ID and a receiver ID (if the receiver is in the city), the caller's and the receiver's cell stations, a time and a duration of the call being made. This dataset is the only one that confirms the connections between two persons. In our experiments, it serves as a ground truth of the connections between two persons, because it has no temporal overlap with other datasets.

**Taxi GPS data -** The taxi GPS data consists of 3610 taxis whose locations are precisely recorded every 20 seconds from 20 Jan 2014 to 28 Jan 2014.

**POI data -** Each POI is labeled with a type by the mobile service provider, such as shopping malls, hospitals, governments and educational institutes, etc. We collected the accurate location of all POIs in the city, and aggregated the POIs by mobile cell stations.

**Microblog data -** The microblog dataset contains about 93491 geo-tagged microblog posts whose time stamps fall between 14 Jan 2014 and 28 Feb 2014. It contains 38000 accounts, of which 1347 accounts made more than 10 geo-tagged posts, and 316 accounts made more than 20 geo-tagged posts.

## 4.2 Spatial regularization with quad tree

Note that different sources of data (e.g., mobile calling data, taxi GPS data) have diverse temporal granularities, while sharing common geography. We propose to reformulate the spatial information of each data record into a uniform geographical structure to favor uniform spot-based matching and trajectory matching. A quad structure is applied to reformulate the spatial information of all records in the data. The quad tree recursively partitions the underlying space (a rectangular area with a diagonal length of 35.57 kilometers) into multi-level hierarchical grids. A geographical location on the map is encoded with a 14-digit code, whose header encodes its hierarchy, so that different locations in the data can be matched at any level between 1 (diagonal length: 17.78 kilometers) to 14 (diagonal length: 2.17 meters).

The quad tree structure has two advantages in trajectory matching. First, the quad tree structure enables fast matching because the storage of different location headers can be avoided. Second, the quad tree structure can adapt to different levels of matching. Actually, matching at different levels plays a significant role in utilizing heterogeneous data. Analysts are capable to identify interesting pairs from massive objects, like the passengers in the same cars, by matching at the highest level. In our case, the locations recorded by cell stations are less accurate than GPS-based coordinates of taxi, and thus analysts must match at different levels to find appropriate candidates. Furthermore, matching level may have to be changed even employing single dataset. For example, downtown has a high population density and a high cell-based station density, which means that the locations stand more accurately. In other words, for the trajectories in downtown, the level can be higher, while the chosen level in urban area can be relatively low.

## 5 MATCHING APPROACH

To measure similarities among different objects, we propose a matching approach based on quad tree.

### 5.1 Basic trace matching constraints

Following the concept of three fundamental movement sets proposed in [2], we clarify the basic operations for matching two traces from urban data. **A spot-based constraint**, denoted as $c(t, s_c)$, contains a time point $t$ and a geographical position $s_c$. Figure 2 shows three spot-based constraints $c1$ and $c2$. To describe an atomic trace of a person that is extracted from the cell-based mobility data, we define **a cell-based interval** $r(t_s; t_e; s_r)$, which contains a start time $t_s$, an end time $t_e$ and a geographical position $s_r$. It indicates that the individual stays in position $s_r$ from $t_s$ to $t_e$. Examples in Figure 2 include $r(T1, T3, L1)$ and $r(T2, T4, L3)$, etc. Thus, **a trajectory**, which is composed of ordered cell-based interval records, can be denoted as

$$R = \{r_1, r_2, ..., r_n\}, \tag{1}$$

where $t_{s_i} < t_{s_j}$ for $i < j$. Figure 2 shows two trajectories, one in red and the other in blue.

A trajectory $R$ satisfies a spot-based constraint $c$ if there exists one cell-based interval record $r$ in $R$ that satisfies the constraint. And a cell-based interval record $r$ satisfies the constraint $c$ if $t_s < t < t_e$ and $s_c$ and $s_r$ are in the same neighborhood as being structured in Section 4.2. In Figure 2; $c1$ is satisfied by cell-based interval $r(T1, T3, L1)$ and $r(T2, T4, L3)$; and $c3$ is satisfied by no one. The **degree of matching** ($DoM$) between two cell-based intervals $r_i$ and $r_j$ can be defined as the duration of the overlapped time if the two positions are in the same local region, as denoted in Equation 2. The $DoM$ of the trajectory in red and the trajectory in blue is the duration indicated by the gray areas.

$$DoM(r_i, r_j) = \begin{cases} 0, & \text{if } s_{r_i} \text{ and } s_{r_j} \text{ are not in the same local region,} \\ 0, & \text{if } r_i \text{ and } r_j \text{ have no temporal overlap,} \\ \min(t_{e_i}, t_{e_j}) - \max(t_{s_i}, t_{s_j}), & \text{otherwise.} \end{cases} \tag{2}$$

The **matching of two trajectories** $R1$ and $R2$ is denoted as the sum of pairwise matching of the segments:

$$DoM(R1, R2) = \sum_{r_i \in R1 \text{ and } r_j \in R2} DoM(r_i, r_j). \tag{3}$$

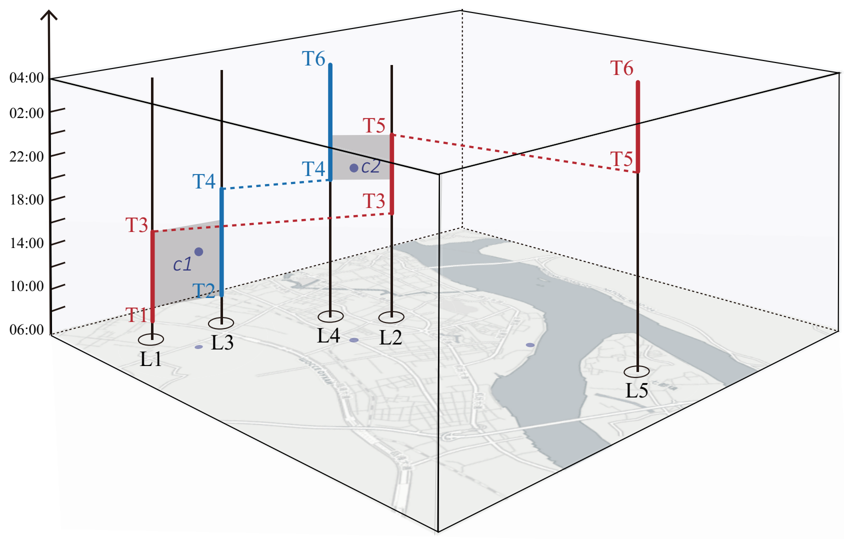



Fig. 2. Illustrating the basic concepts of spatio-temporal matching.

## 5.2 Spatial-temporal matching

Two fundamental data matching operations are designed to support comprehensive data query. The first one performs the matching on the basis of spots and time points, called spot-based matching. The other, trajectory-based matching, seeks to match trajectories, which can be used to reveal the time duration of geographical co-occurrence of two persons.

*5.2.1 Spot-based matching.* The spot-based matching is designed to find persons whose traces satisfy a spot-based constraint $c(t, s_c)$ given by the analysts. The geographical position $s_c$ is encoded with a quad tree node. The constraints are interactively specified by either sketching on the map or from selected microblog posts.

*5.2.2 Trajectory-based matching.* Trajectory-based matching is performed by matching a given trajectory-based constraint, namely a trajectory, from a group of trajectories, described as $DoM(R1, R2)$. The matching is performed one-by-one for each candidate with either way described below:

**Partial matching** Partial matching corresponds to the degree of matching between two trajectories, with each pair of the segments $r_i, r_j$ matched by Equation 2.

**Binary matching** A binary matching judges how many cell-based intervals perfectly match for two trajectories. It counts the number of matched intervals in the trajectories. The $DoM(r_i, r_j)$ of the cell-based intervals is either 1 if they perfectly match or 0 otherwise.

The partial trajectory-based matching is preferred when we handle stationary trajectories, while the binary matching is preferred for moving trajectories. For example, the binary matching between two individuals (see Figure 2) results in 2, while the partial matching under the same conditions yields $((T3 - T2) + (T5 - T4))/(T6 - T1)$.

## 6 INTERFACE

RelationLines is composed of five views: a target view (Figure 3) that identifies the target individual, a timeline view (Figure 5) that depicts the mobility behaviors, an egocentric relation view (Figure 6) that summarizes relations based on the direct and indirect contacts between the individual and others, a co-occurrence timeline view (Figure 7) that verifies relations among contacts and a map view (Figure 4) that verifies all graphical information. Each view is corresponding to a task respectively, except the map view, which may facilitate analysts in all tasks. Discovering egocentric relations of an individual can be progressively performed by exploring and manipulating these views.

### 6.1 The target view (Task1)

The target view is used to identify a target (a taxi or a mobile phone) with our data by heterogeneous constraints: spot-based constraints, taxi IDs and microblog posts [1]. For instance, the analysts can search a mobile phone user by specifying spot-based constraints with the help of the map and other views. They can also retrieve a taxi by its ID or find a taxi that follows a specific route to locate a taxi. Furthermore, they can link a microblog account with a mobile phone user if the user made many geo-tagged microblog posts. Note that our system displays the strings like "id100001" instead of the real nicknames because of the privacy preservation. For the same reason, we employ the day of the week to instead of the specific date in the system.

Figure 3 illustrates the process of searching and analyzing potential candidates. In particular, the spot-based constraints are represented by two geographical icons on the map view with associated temporal information. Both constraints are listed in the target view for selection and modification. Two matched candidates are found after the analysts perform a spot-based matching. The analysts can further evaluate the results on the map view or the timeline view.

### 6.2 The map view and the timeline view (Task2)

The map view shows the geographical scene, and serves as a positional reference for multiple purposes. For instance, markers can be interactively placed on the map to specify the spatial constraints. The data that contains positional information like the microblog posts and trajectories can also be shown on the map (see Figure 3 (b)).

We additionally design a **radial timeline chart** (see Figure 4) that emphasizes on the spatial information of the traces in the period for individual traces. For each location visited by the target, a radial chart (see Figure 4 (a)) is generated and placed on the location in the map view. The chart contains multiple nested rings, each of which encodes one day (24 hours) clockwise. If an individual visits a spot during a certain period uniquely, the corresponding arc of a ring in the associated radial chart is set to be the color related to its location. We cluster all locations visited by the target based on latitude and longitude via K-means. The number of the clusters is adjustable. Then, locations in the same cluster are

[1]We do not allow for identifying from the mobile calling data because it has no temporal overlap with other sources in our test data.

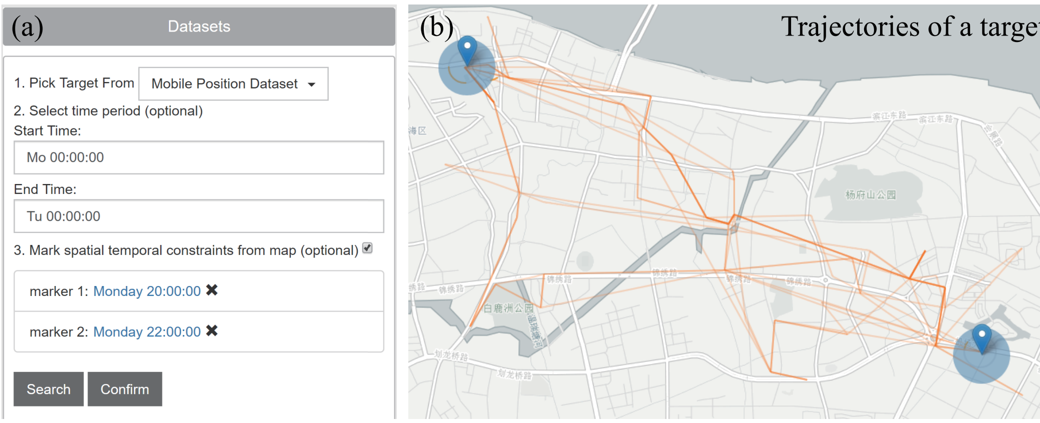


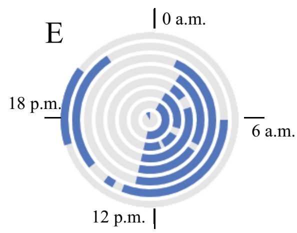

Fig. 3. A target is identified by spot-based constraints. Both the found candidates and spot-based constraints are listed in the target view (a). The analysts can visually analyze the traces of candidates and the constraints on the map view (b) and the timeline view.

mapped to the colors belonging to the same part of the color wheel so that similar locations have similar colors. In this way, the collection of all radial charts is a reformulation of the trace timeline on the basis of the map. The ring number varies with the focused time duration: only one ring is used when the analysts select one day in the timeline view. Weekly patterns can be revealed with the help of the radial timeline charts. Figure 4 (a) depicts the timelines of the target at locations E, A and C in Figure 4 (b). From the visualization, one can easily infer that the target stays at E for nearly half a day in six days, goes to A in four mornings, and visits C in most afternoons. Figure 4 (b) discloses all routing behaviors. One of them is from E $\rightarrow$ A $\rightarrow$ C $\rightarrow$ E (highlighted with arrows). Furthermore, the analysts can compare traces of multiple persons on the map.

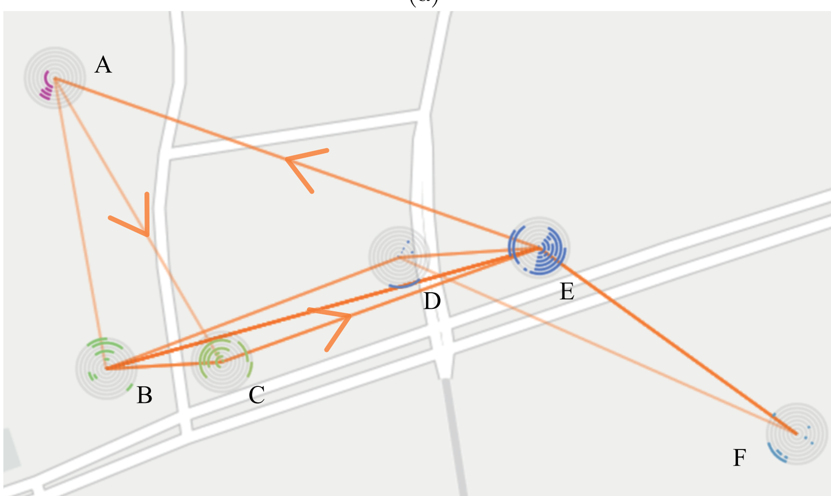
Once an individual is selected, its trace in a long period is plotted in the timeline view (Figure 5). The visual form used in the timeline view is an enhanced bar chart, called the **trace timeline chart**. The $x$ and $y$ axe of the trace timeline chart span 24-hours and multiple days, respectively. A trace of a target is represented with a set of bars along the $x$ axis. The trace timeline charts for the station-based mobility data and the taxi data are different, as shown in Figure 5.

**Station-based mobility data -** For the mobility data, the bars on the timeline represent the places the individual stayed longer than 30 minutes. The bars indicate the frequent places the individual stays, which may be home, work place, hospitals or shopping malls. The blank space between two bars (Figure 5 (a)) indicates a transition route from one place (the left bar) to another (the right bar). The colors of the bars is depending on the locations of the individual stays. We employ the same color mapping approach as the radial charts, thus analysts can find related locations on the map in accordance with the color.

The analysts can highlight important places such as places the individual stayed longer than 60 minutes or places the individual visited more than 3 times. Frequent visited places are easily revealed by the colors of bars. For example, Figure 5 (a) reveals that the individual stayed in the same places (in green) for 4 days. The analysts further




(a)




(b)

Fig. 4. The trace of an individual on the map. (a) The radial glyph depicts weekly moving behaviors. (b) A routing behavior.

infer that these places may be the work place because they appear commonly in the afternoon.

**Taxi GPS data -** For the taxi data, the bars represent the duration when the target (a taxi) rides passengers (Figure 5 (b)). Each bar indicates a short-term relation between the taxi driver and the passengers. Unlike the trace timeline chart used for the mobility data, the bars for taxi data is encoded in the same color because taxis visit multiple places during their rides. Figure 5 (b) depicts the riding pattern of a taxi.

The analysts can directly select one day on the trace timeline chart, and study the trajectories on the map view. The places and their connecting routes are simultaneously shown in the map view. The analysts can also brush a set of bars or the connecting intervals to view their corresponding traces on the map. In addition, the traces selected from the trace timeline chart of the mobility data and that of the taxi data can be matched, and the resultant trajectory intervals indicate potential relations between two objects in reality. The top matched results are automatically listed in the egocentric relation view, each of which can be examined individually on the interface.

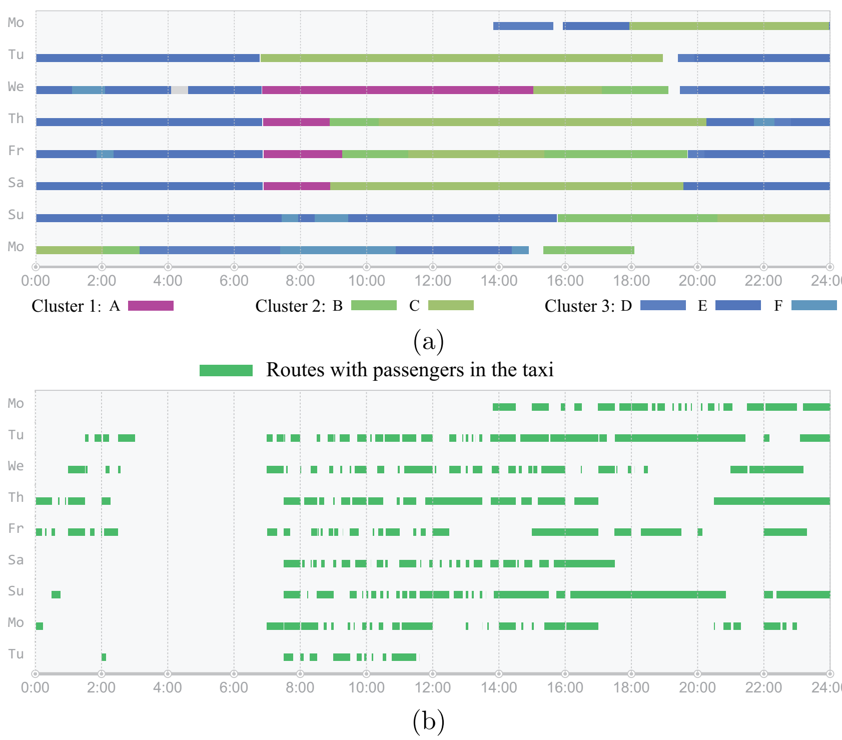



Fig. 5. Trace timeline charts for a mobile phone user (a) and a taxi (b). The color in (a) is randomly assigned to different types of locations. The bars in blue implies that the user may stay at the work place in the afternoon. The rectangles in green are trajectory traces selected by the analysts to filter the co-occurrence individuals during those time spans. In the meantime, a taxi riding pattern may be disclosed from (b).

### 6.3 The egocentric relation view (Task3)

Individuals of top qualified relations are summarized in the egocentric relation view (see Figure 6). The egocentric relations are visualized with a bar chart, called *the matched timeline list*. This chart (Figure 6) displays top-30 qualified individuals (as rows) and all employed constraints (as columns). Both the indirect relations such as co-occurrence and direct relations such as mobile calls are maintained as relations. In particular, the direct relation is named as "call", and represented with a bar, whose color is orange, and whose size is user-adjustable. Other indirect relations are sequentially made by the analysts, e.g., "M1", "M2", and "M3" in Figure 6. Each one is represented with a bar, whose color is automatically assigned as a graduated gray. Besides, we normalize these individuals' degrees of match to $[0, 1]$ as the length of bars.

The analysts can sort the individuals according to relations by one constraint or all constraints. Figure 6 illustrates a list after it sorted by all constraints. They can also add a trajectory-based constraint by brushing a time span on the trace timeline chart, delete a useless matching constraint, or re-weight the matching constraint by dragging the constraint headers. The refinement and investigation can be progressively performed by adding new constraints. In order to improve the efficiency, we set an amount limit as 100 and a matching threshold as 0.8. For each constraints, we keep a group of individuals, whose degree of match is top 100 or exceeds the threshold. After analysts adjust the constraints, system returns the top-30 individuals from the individuals belonging to all the groups based on all constraints automatically.

Whenever some relations of interest are detected, the analysts can make annotations to related individuals beside the small icon of the candidate. In addition, the analysts can interactively change the size of each bar, so that the corresponding weight is adjusted.

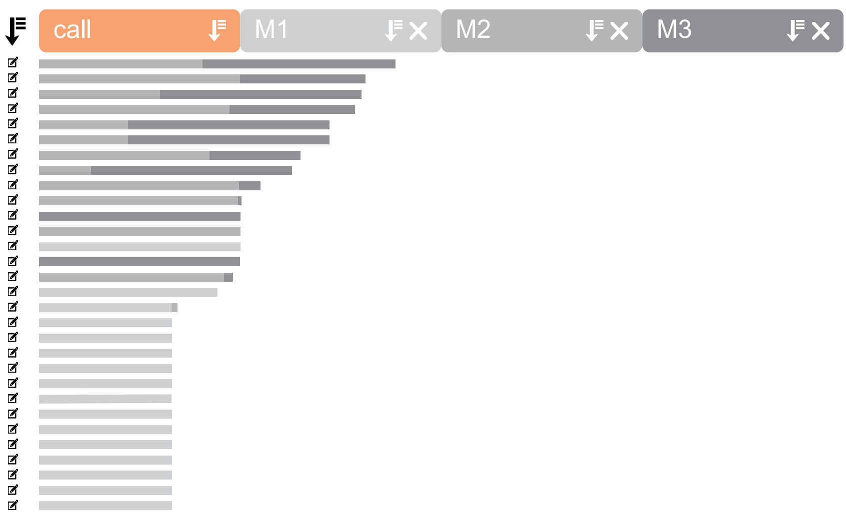



Fig. 6. The matched timeline list for analyzing ego-centric relations of the underlying target shows the top-30 of the detected candidates from previous ego-relation analysis process. The list can be sorted with respect to specified constraints.

### 6.4 The co-occurrence relation view (Task4)

The co-occurrence timeline chart (see Figure 7) shows the matched trajectories among individuals during the entire timeline chart. The initial chart shows the timelines of matched individuals in parallel, and visualizes the co-occurrence (see Figure 7 (a)) that shows the timelines of matched individuals in parallel. For each trace of the target, it counts the number of occurrences and encodes it with a sequential orange color. The analysts are able to detect the time spans indicated the co-occurrences among individuals which have not been selected (brushed) previously.

As indicated in Figure 5 (a), the brushed time spans are roughly from 16:20 to 17:40 on the first Tuesday, from 16:40 to 17:40 on Wednesday, from 17:10 to 17:50 on Thursday and from 16:50 to 18:50 on Friday. Nevertheless, Figure 7 (a) indicates a large co-occurrence in the morning. When the analysts select a cell-based interval before dawn (from 1:00 to 2:50), individuals with co-occurrence during that period are highlighted in brown (Figure 7 (b)). The analysts can easily notice that each of them has a long or short co-occurrence with the target . Among them, the Candidate2 has almost a full co-occurrence with the target. Thus the analysts make an assumption that it is the target's family member. The time span around 9 a.m. can also be used to filter the co-occurrence relations in the trace timeline chart of the target.

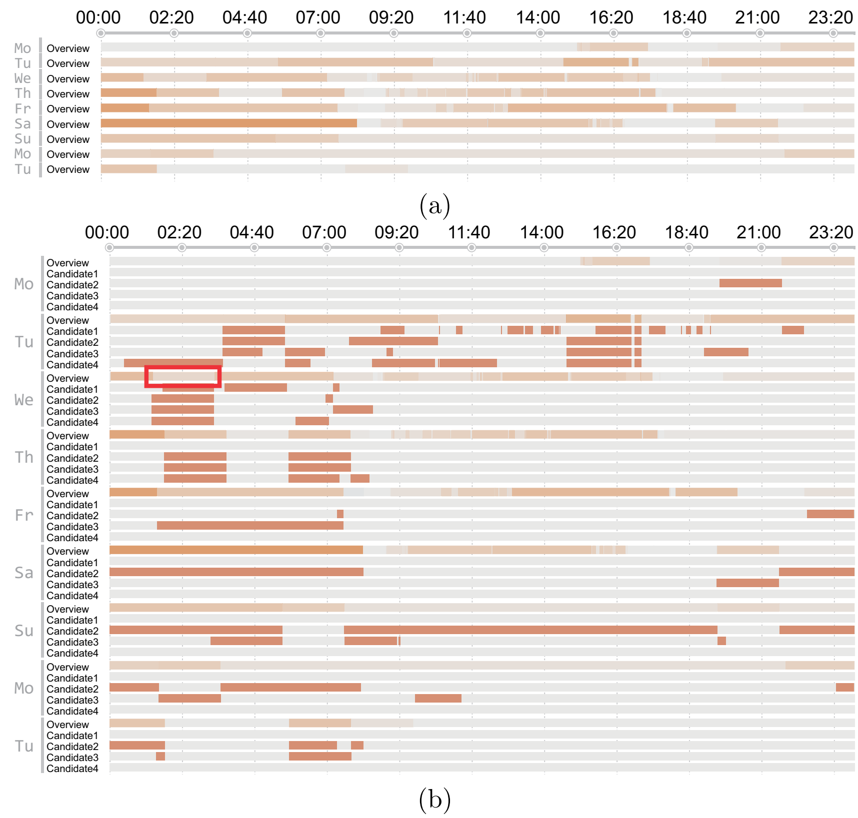



Fig. 7. The co-occurrence timeline chart of top-30 related individuals. (a) The co-occurrence between the selected individuals and the target is displayed. (b) The analysts select a cell-based interval before dawn (with a orange rectangle) to filter the individuals.

## 7 EVALUATION

In this section, we describe three case studies and a user review.

### 7.1 The egocentric relations of a microblog user

The first case study detected an one-to-one relationship by inferring the information extracted from microblog posts. In order to construct egocentric network reasonably, we first excluded inactive users (meaningful posts less than 10 during the nine days). Then, we found a female user (P1) who exhibits an interesting trace (Task1). She has 7 posts which contain geo-tags and time stamps. The geo-tags of these posts refer to two places, namely, A and B in Figure 8 (b). The posts in A talk about the coming birth of her baby ( the birthday) and the good wishes for the born baby (a week later the birth). The posts in B are made one day before the birth and four days after the birth. The POI data indicates that B is the hospital (see Figure 8 (b)), which is consistent with the posts made in B. To retrieve all traces, we first perform a spot-based matching at the 10th level. The result misses some trajectory points, which means that the matching level should be adjusted. We change the level to be 6, and then identify the mobile ID of P1 (Figure 8 (a)) as well as all traces (Figure 8 (c)). The trace timeline chart depicts her traces sequentially: staying in A for one and half day, transferring between A and B for several times, staying in B for days, and finally coming back to A (Task2). There are no trace record during Thursday to Friday, probably because the phone is off. The spots in blue and pink on the trace timeline chart refer to A and B respectively. The parts in blue has a large coverage of the timeline, implying that A is her home. We further seek to analyze her social ties. With the assump-

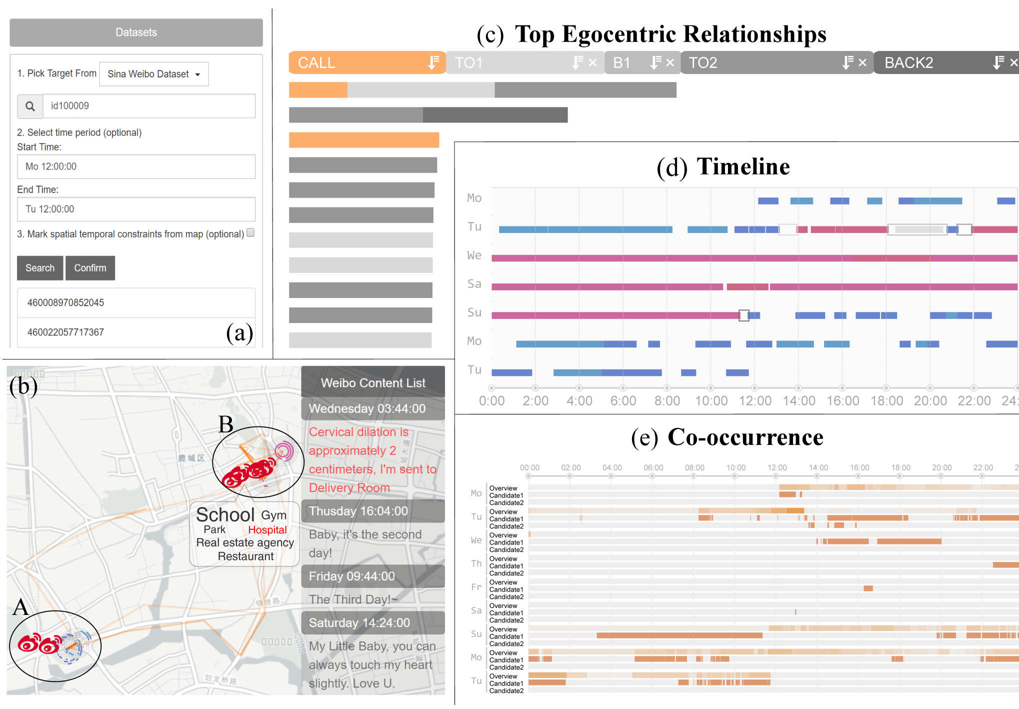



Fig. 8. The interface of RelationLines. (a) The target view which is used to retrieve a microblog user. (b) The map view with geo-tagged posts that indicate the target gave birth to a baby. Furthermore, her posts are all concentrated in location A and B. (c) The egocentric relations of the target. (d) The trace timeline chart of the target which depicts the transitions between her home and the hospital. (e) The co-occurrence timeline chart of the detected egocentric network. For more details on how RelationLines helps discover the egocentric relations, please refer to Section 6.1.

tion that a pregnant woman is usually accompanied with her family member, we match her trace moves to A (home) from B (hospital) with the trajectories in the cell-based mobility data, yielding a list of candidates (Figure 8 (d)). There is one candidate (denoted as P2) who has made phone calls with P1 according to the mobile phone calling data (Task3). In addition, their traces have many co-occurrences, implying that this person is her family member, probably her husband (Figure 8 (e)). Finally, we analyze the social ties of other candidates in Figure 8(c) and (e), and find two other people who stay in the same place as P2 (Task4). We infer that they are probably family members or close friends of P1 and P2, and come to take care of P1.

### 7.2 Behaviors of taxi drivers

The goal of the second case study is examine the relationship between taxi and individuals. We randomly select a taxi, C04937, from the taxi GPS data (Task1). By studying the trace timeline chart, we find that C04937 is fully occupied everyday (the highlighted green in Figure 9 (b)), except a 2-hour gap around 7 a.m. (Task2). It is common that in the city two taxi drivers will share a taxi and change shifts. We infer that the shift time is around 7 a.m.. To identify the driver of the taxi, we select the trajectories around 7 a.m. as trajectory-based constraints. The top-30 list indicates 2 matched mobile phones: P1 and P2, and the former matches better than the latter. However, when we compare two individuals with the taxi's trajectories, we find that P1 has co-occurrences with the taxi day and night in all days while "P2" only has co-occurrences at night. Then we adjust our assumption that P2 is the night shift driver and P1 might be a mobile phone of the taxi driver which is fixed in the taxi. Observing the trajectory of P2 further confirms that the shifts all took place around 7 a.m.. To identify the second driver, we select trajectories in the day time and refine the egocentric relations. P1 remains top in the top-30 list. Besides, we find that another individual P3 matches the taxi trajectories. Comparing its trajectories with the taxi's on the timeline, we realize that this individual and P2 match the taxi trajectory all day from the first Monday to Saturday. Nevertheless, the matched driver in the rest days is still missing. We select several trajectories in the daytime after Sunday and further refine the egocentric relations. P4 matches the trajectory-based constraints and indicates a high co-occurrence in the daytime after Sunday.

Now that we have found three drivers and a mobile phone related to the taxi, it is convenient to find the locations where they shift duties. The locations are A for morning shift (see Figure 9 (c)) and B for night shift (see Figure 9 (d)) . We also find out that A is very close to the destination of the last ride, and the same situation happens for B too. It is common because taxi drivers tend to give ride to a passenger whose destination is near the shifting location

### 7.3 Behaviors of a commuter

In the final case study, we attempt to construct the egocentric network based on the detailed commuting traces of an individual. The common traces shared by different persons may have a large number of co-occurrences, which contain much more uncertainty. By leveraging the result of spot-based matching, we find an office employee who stays at his/her in his company every afternoon (Task1). This person (supposed to be a male) has a very regular life style, as shown in the trace timeline chart and map (Figure 10 (a) and (c)). He moves from Region C (colored in light blue in (c)) to Region A (colored in pink in (c)) in the morning or in the afternoon almost everyday, implying a commuting trace from home to work (Task2). By brushing the areas in green and pink, we detect both possible family members or neighbors and workmates. The direct contact relation (phone calls) confirms his family relation with Candidate1, who shares a long home-related trace with him. Another call relation confirms his workmate relation with another individual, who shares a long work-related trace with him (Task3).

By comparing his trajectories with his family member Candidate1, we discover region B. It indicates that they walk around region B for some time in the morning, and then stay at a spot for an hour at noon. Our initial assumption is that they go for shopping in B and had lunch in a restaurant. We then brush B to search other possible family members who join the shopping, However, to our surprise, the matched individuals (Figure 10 (d)) mostly are

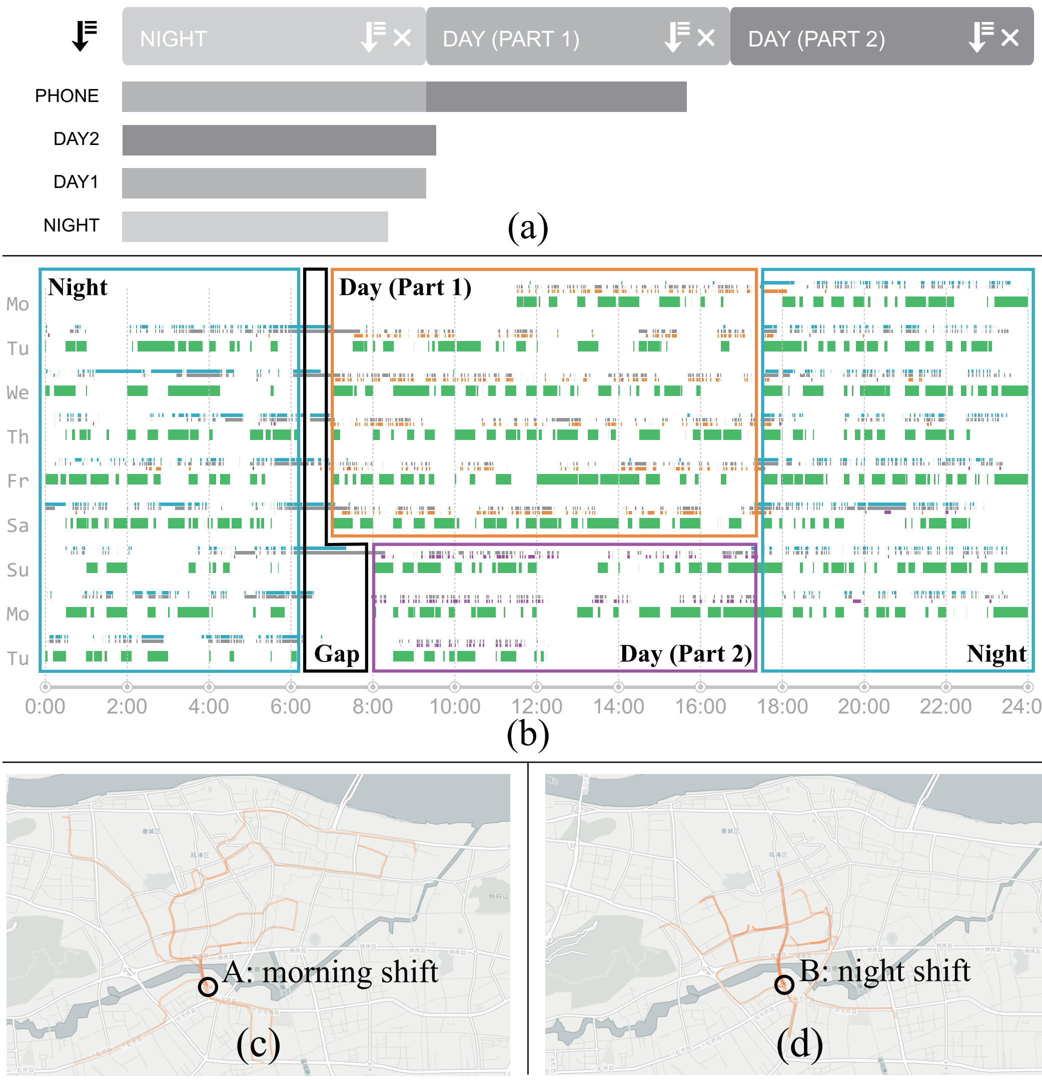



Fig. 9. Behaviors of taxi drivers. (a) The matched results of traces of different taxis. (b) The co-occurrences between the four IDs and the taxi. (c) The shifting location around 7:00 a.m.. (d) The shifting location around 17:30 p.m..

of work-related co-occurrence. Then our guess becomes whether the place is a bus transit location. However, when we search the bus exchanging from his home to work, region B is not on the commuting route. Besides, according to the co-occurrence timeline chart in (d), they move in the region B together for almost 2 hours. And some of them also go there consecutively on days before the second Monday. We then make another assumption that maybe it is the fieldwork place of the individual. And maybe the fieldwork is an open activity so that his family member joins him on that day.

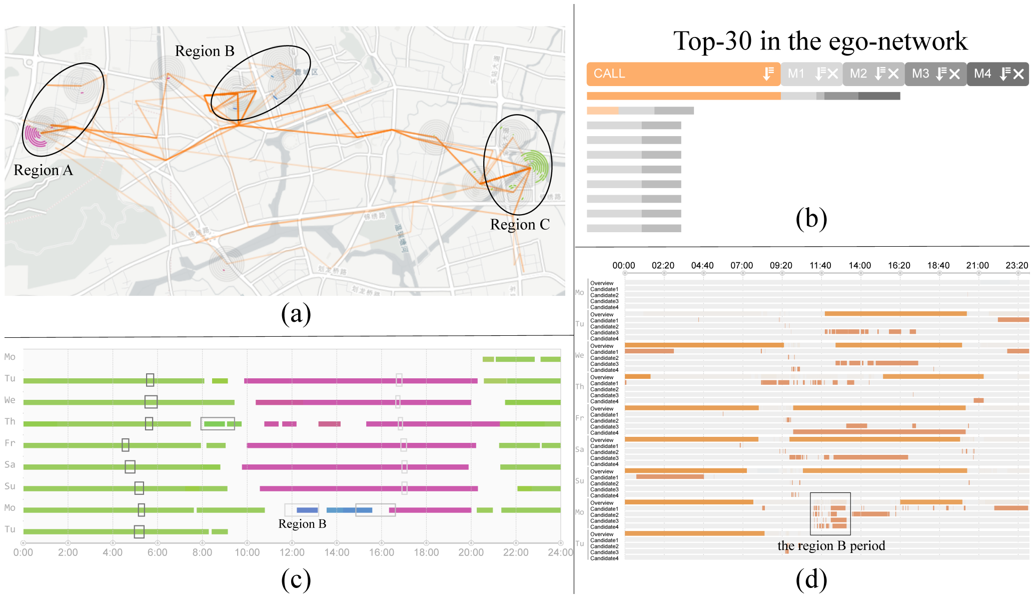



Fig. 10. Analyzing the moving behavior of a commuter. (a) The routings on the map. (b) The constructed egocentric relations based on the constraints derived from the brushed areas (unfilled rectangles) in (c). (c) The trace timeline chart, together with the trace (encoded as the thin grey lines on the top of the timeline bars) of "Candidate4". (d) The clustered pattern of the possible workmates of the commuter.

## 7.4 Expert review

We asked two experts to evaluate our system. The first one is a retired policeman who used to be responsible for investigating cases. And the second one is a director of the Public Security Bureau Command Center, working with the direction of security by video surveillance to investigate clues to the case. Both of them have been working for decades. We described our system in detail and presented the three case studies mentioned above.

The first expert told us that behavior trajectory data is the main data they employed in their work. He affirmed our thought that behaviors and spatio-temporal stays reflect the relationship between individuals. Furthermore, "detailed features such as activity area during specific time," he said, "can leverage to construct social network with different relationships, like colleagues, friends and families." He finally drew a conclusion that it makes sense to combine our system with public security operations.

The second expert gave strongly positive comments to our system. He introduced that exploring personal relationship networks by trajectories is of great significance for the public security organizations. "The primary principle of police investigation is to find out the cross space-time connections between people, connections between people and the scene of the case," he said "and to restore the activity trajectories of all relevant personnel in the case, which plays an important role in investigation of the case, delimiting the scope of suspects, and obtaining evidence." Then, we asked him what is the benefit of our research on his work. "The case investigation starts with proving the activity trajectories of persons involved in the case. If we can quickly access to personal relationship networks and activity trajectories of suspects, victims and related witnesses by technical means, we can increase the efficiency of investigation and reduce the judicial costs." replied he. His answer proved the practical value of our system.

## 8 PERFORMANCE AND SCALABILITY

In this section, we discuss about system performance, scalability and ego-network reliability.

### 8.1 Performance

To accelerate the query process of 600 million cell-based mobility data with many strategies, we use MonetDB, a column-oriented database, which achieves fast query by column-based conditions. Multiple tables are created for data with different quad-tree encoding headers, which indicate their corresponding geographical regions. The smaller a table is, the faster the data can be queried in one table. Because a small table can only accommodate data of a small region, queries in a large region would result in manual manipulations of cross-table query and records merging, which is time-consuming. Thus we partition a table into four tables recursively if a region contains more than millions records. The partitioning results in 74 tables and the lengths of the corresponding quad-tree encoding headers vary between three to five digits. A five-digit region is employed to cover the neighborhood of the individual and avoids cross-table query.

Column-oriented databases are inefficient in querying rows of complete records. It takes 3.3 seconds to get the one-week trajectories of an individual. To tackle this problem, we build a MySql database for the data with the same configuration, except that MySql makes multiple partitions in one table instead of multiple tables. Given an ID, MySql can return the one-week trajectories in tens of milliseconds. The MonetDB database is built on a workstation with 256G memory and 24 Xeon E7540@2.00GHz processors. The MySql database is built on a desktop with 8G memory and 4 Core i7 4770@3.4GHz processors.

Mysql and MonetDB provide fast query for interactive visual query. ID-oriented queries are conducted in MySql and can be fulfilled within milliseconds. Spot-based matching and trajectory-based matching are conducted in MonetDB. Table 1 summarizes the average query time in MonetDB with respect to the number of cell-based intervals. Retrieving a cell-based interval with 20 points can be accomplished in 4 seconds, which is practical for an interactive system.

| The number of cell-based intervals | 2 | 5 | 10 | 20 | 30 | 40 | 50 | 100 |
|---|---|---|---|---|---|---|---|---|
| The average query time (seconds) | 0.35 | 1.67 | 3.34 | 5.92 | 9.39 | 11.88 | 12.82 | 30.75 |

Table 1. The performance statistics of MonetDB query.

### 8.2 Scalability

In addition to the cell-based mobility data employed in our study, trajectory data can be extracted from distinct datasets, like mobile phone GPS data and sparse transportation data [38]. Different trajectory data may differ in spatial granularity and temporal granularity. Any trajectory data can be employed in our approach after certain preprocessing.

Our data organization approach is suitable to spatio-temporal data in varied spatial granularities by defining the level amount of the quad tree, namely, the accuracy. More hierarchies can preserve more accuracy in the price of extra storage. Simultaneously, analysts should take the area of efficient regions into consideration. For example, each record in our cell-based mobility data indicates that an individual stays in a region instead of a specific point. A wide effective region may also reduce the need for accuracy. Analysts can modify the number of matched digits in accordance with that area. Relaxing matching accuracy is necessary for a wide area to hit the “target”, but it may also increase the amount of “candidates”. However, it is impossible to completely eliminate the error caused by data inaccuracy. In such a case, combining with other information, like direct communications, plays a significant role.

Adapting to trajectory data with different temporal granularities makes little influence on data storage. When matching trajectories, we provide two options to handle different temporal granularities: partial matching and binary matching (in Section 5.2). Partial matching is effective for the trajectories that can only be split into long intervals, such as the trajectories recorded by cell stations. On the contrary, the trajectory data, like mobile phone GPS data, is more suitable for the other technique.

### 8.3 Trajectory matching approach

There are other trajectory matching approaches [3, 24]. Compared with our solution, they seek to calculate (both temporal and spatial) similarity in a more concrete yet complex way. However, our target is studying the relationship between two individuals based on location co-occurrences. We match a pair of trajectories with a unique consideration: whether two individuals are close adequately to present oc-currences or not. For example, the husband may have a rest in the cafe and wait for his wife, who shops in the clothes store next to the cafe. Then, they have dinner together in a restaurant nearby. The distance between the couple varies over time, but the entire process indicates the same thing: they go out together. There is no need to split this process, and thus we simplify the computation of the similarity of trajectory into dyadic matching (keep in touch or not). On the other hand, our target is heterogeneous urban data, which means that the interpersonal distance may be effected by the accuracy of the recorded locations. An appropriate granularity concerning the spatial similarity is demanded for optimal comparison and discovery tasks.

### 8.4 Reliability

There are several situations that our approach can not yield a reliable ego-network. It is challenging to analyze individuals whose mobility data is incomplete or who do not move. This is because our visual reasoning process proceeds on the basis of the individual’s trace patterns. Ego-relations of individuals who have frequently visited places are often more predictable. And individuals without stable visited places are unpredictable because the analysts cannot identify their behaviors at those places. For instance, in Case 3, we detect the places visited by the individual and his “friends”, because these places are neither home or work places. . Nevertheless, without direct evidence, not conclusions but only assumptions can be made. On the other hand, the microblog data in Case 1 provides a direct evidence that the pregnant woman gave birth to her baby in the hospital. To extend the relations to the top-30 individuals, the co-occurrence timeline chart is employed to analyze the interpersonal relations among ego-related individuals. It helps enhance the relation closeness and detect communities based on the trajectory-based constraints of all time spans.

## 9 CONCLUSION

In this paper, we propose a detect-and-filter scheme for visual reasoning of egocentric relations from urban data. The coarse detection is performed to extract a target individual from heterogeneous data constraints and construct the timeline of an individual. It detects spatio-temporal co-occurrences with spot-based matching, trajectory-based matching and call-based matching to develop the egocentric network of the target. The filtering stage is enhanced with a line-based visual reasoning interface that facilitates flexible and comprehensive investigation of egocentric relationships and connections in terms of time, space and social networks. We demonstrate the efficiency of RelationLines, with the experiments over a dataset that contains taxi GPS data, mobile cell station data, mobile calling data, microblog data and POI data of a city with millions of citizens.

In the future, we plan to employ interactive neural network methods [21] to enhance the intelligence of relation mining. With adequate knowledge, appropriate cluster analysis will be more adapt to massive heterogeneous urban data.

## REFERENCES


[1] Shamal Al-Dohuki, Yingyu Wu, Farah Kamw, Jing Yang, Xin Li, Ye Zhao, Xinyue Ye, Wei Chen, Chao Ma, and Fei Wang. 2017. SemanticTraj: A new approach to interacting with massive taxi trajectories. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 11–20.

[2] Gennady Andrienko, Natalia Andrienko, Peter Bak, Daniel Keim, and Stefan Wrobel. 2013. *Visual analytics of movement*. Springer Science & Business Media.

[3] Gennady Andrienko, Natalia Andrienko, and Stefan Wrobel. 2007. Visual analytics tools for analysis of movement data. *ACM SIGKDD Explorations Newsletter* 9, 2 (2007), 38–46.

[4] Moshe Blank, Lena Gorelick, Eli Shechtman, Michal Irani, and Ronen Basri. 2005. Actions as space-time shapes. In *Proceedings of the 10th IEEE International Conference on Computer Vision*, Vol. 2. 1395–1402.

[5] Stephen P Borgatti, Ajay Mehra, Daniel J Brass, and Giuseppe Labianca. 2009. Network analysis in the social sciences. *Science* 323, 5916 (2009), 892–895.

[6] Nan Cao, Chaoguang Lin, Qiuhan Zhu, Yu-Ru Lin, Xian Teng, and Xidao Wen. 2018. Voila: Visual Anomaly Detection and Monitoring with Streaming Spatiotemporal Data. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2018), 23–33.

[7] Duen Horng Chau, Aniket Kittur, Jason I Hong, and Christos Faloutsos. 2011. Apolo: making sense of large network data by combining rich user interaction and machine learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 167–176.

[8] Wei Chen, Zhaosong Huang, Feiran Wu, Minfeng Zhu, Huihua Guan, and Ross Maciejewski. 2018. VAUD: A Visual Analysis Approach for Exploring Spatio-Temporal Urban Data. *IEEE Transactions on Visualization and Computer Graphics* (2018).

[9] Tarik Crnovrsanin, Isaac Liao, Yingcai Wu, and Kwan-Liu Ma. 2011. Visual recommendations for network navigation. *Computer Graphics Forum* 30, 3 (2011), 1081–1090.

[10] W. Cui, S. Liu, L. Tan, C. Shi, Y. Song, Z. Gao, H. Qu, and X. Tong. 2011. TextFlow: towards better understanding of evolving topics in text. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2412–2421.

[11] N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva. 2013. Visual exploration of big spatio-temporal urban data: a study of New York City taxi trips. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2149–2158.

[12] Anna Fredrikson, Chris North, Catherine Plaisant, and Ben Shneiderman. 1999. Temporal, geographical and categorical aggregations viewed through coordinated displays: a case study with highway incident data. In *Proceedings of the 8th ACM internation conference on Information and knowledge management*. 26–34.

[13] Peter Gatalsky, Natalia Andrienko, and Gennady Andrienko. 2004. Interactive analysis of event data using space-time cube. In *Proceedings of the 8th International Conference on Information Visualisation*. 145–152.

[14] Diansheng Guo, Jin Chen, Alan M MacEachren, and Ke Liao. 2006. A visualization system for space-time and multivariate patterns (vis-stamp). *IEEE Transactions on Visualization and Computer Graphics* 12, 6 (2006), 1461–1474.

[15] Hanqi Guo, Zuchao Wang, Bowen Yu, Huijing Zhao, and Xiaoru Yuan. 2011. TripVista: Triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection. In *Proceedings of the IEEE Pacific on Visualization Symposium*. 163–170.

[16] Yu-Yueh Huang, Cheng-Te Li, and Shyh-Kang Jeng. 2015. Mining location-based social networks for criminal activity prediction. In *Proceedings of the 24th Wireless and Optical Communication Conference*. IEEE, 185–189.

[17] S Kairam, NH Riche, S Drucker, R Fernandez, and J Heer. 2015. Refinery: Visual exploration of large, heterogeneous networks through associative browsing. *Computer Graphics Forum* 34, 3 (2015), 301–310.

[18] Kuno Kurzhals and Daniel Weiskopf. 2013. Space-time visual analytics of eye-tracking data for dynamic stimuli. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2129–2138.

[19] Lauro Lins, James T Klosowski, and Carlos Scheidegger. 2013. Nanocubes for real-time exploration of spatiotemporal datasets. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2456–2465.

[20] Dongyu Liu, Di Weng, Yuhong Li, Jie Bao, Yu Zheng, Huamin Qu, and Yingcai Wu. 2017. Smartadp: Visual analytics of large-scale taxi trajectories for selecting billboard locations. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 1–10.

[21] Mengchen Liu, Jiaxin Shi, Zhen Li, Chongxuan Li, Jun Zhu, and Shixia Liu. 2017. Towards better analysis of deep convolutional neural networks. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 91–100.

[22] S. Liu, J. Yin, X. Wang, W. Cui, K. Cao, and J. Pei. 2016. Online Visual Analytics of Text Streams. *IEEE Transactions on Visualization and Computer Graphics* 22, 11 (2016), 2451–2466.

[23] Julian Mcauley and Jure Leskovec. 2014. Discovering social circles in ego networks. *ACM Transactions on Knowledge Discovery from Data* 8, 1 (2014), 4.

[24] Harvey J Miller. 2005. A measurement theory for time geography. *Geographical analysis* 37, 1 (2005), 17–45.

[25] Vasyl Palchykov, Kimmo Kaski, Janos Kertész, Albert-László Barabási, and Robin IM Dunbar. 2012. Sex differences in intimate relationships. *Scientific reports* 2 (2012).

[26] Heiko Paulheim, Petar Ristoski, Evgeny Mitichkin, and Christian Bizer. 2014. Data mining with background knowledge from the web. *RapidMiner World* (2014), 1–14.

[27] Dieter Pfoser, Christian S Jensen, Yannis Theodoridis, et al. 2000. Novel approaches to the indexing of moving object trajectories. In *Proceedings of the Conference on Very Large Data Bases*. 395–406.

[28] Mikolaj Jan Piskorski. 2011. Social strategies that work. *Harvard Business Review* 89, 11 (2011), 116–122.

[29] Benjamin Renoust, Guy Melancon, and Tamara Munzner. 2015. Detangler: Visual Analytics for Multiplex Networks. *Computer Graphics Forum* 34, 3 (2015), 321–330.

[30] Céline Rozenblat and Guy Melançon. 2013. *Methods for multilevel analysis and visualisation of geographical networks*. Springer.

[31] Yang Song, Zheng Hu, Xiaoming Leng, Hui Tian, Kun Yang, and Xin Ke. 2015. Friendship influence on mobile behavior of location based social network users. *Journal of Communications and Networks* 17, 2 (2015), 126–132.

[32] Guodao Sun, Tan Tang, Tai-Quan Peng, Ronghua Liang, and Yingcai Wu. 2017. SocialWave: Visual Analysis of Spatio-temporal Diffusion of Information on Social Media. *ACM Transactions on Intelligent Systems and Technology (TIST)* 9, 2 (2017), 15.

[33] Johan Ugander, Lars Backstrom, Cameron Marlow, and Jon Kleinberg. 2012. Structural diversity in social contagion. *Proceedings of the National Academy of Sciences* 109, 16 (2012), 5962–5966.

[34] Landesberger T Von, F Brodkorb, P Roskosch, N Andrienko, G Andrienko, and A Kerren. 2016. MobilityGraphs: Visual Analysis of Mass Mobility Dynamics via Spatio-Temporal Graphs and Clustering. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 11–20.

[35] Fei Wang, Wei Chen, Feiran Wu, Ye Zhao, Han Hong, Tianyu Gu, Long Wang, Ronghua Liang, and Hujun Bao. 2014. A visual reasoning approach for data-driven transport assessment on urban roads. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*. 103–112.

[36] Xumeng Wang, Jia-Kai Chou, Wei Chen, Huihua Guan, Wenlong Chen, Tianyi Lao, and Kwan-Liu Ma. 2017. A Utility-aware Visual Approach for Anonymizing Multi-attribute Tabular Data. *IEEE Transactions on Visualization and Computer Graphics* (2017).

[37] X. Wang, S. Liu, J. Liu, J. Chen, J. J. H. Zhu, and B. Guo. 2016. TopicPanorama: A Full Picture of Relevant Topics. *IEEE Transactions on Visualization and Computer Graphics* 22, 12 (2016), 2508–2521.

[38] Zuchao Wang, Tangzhi Ye, Min Lu, Xiaoru Yuan, Huamin Qu, Jacky Yuan, and Qianliang Wu. 2014. Visual exploration of sparse traffic trajectory data. *IEEE Transactions on Visualization and Computer*

*Graphics* 20, 12 (2014), 1813–1822.
[39] Di Weng, Heming Zhu, Jie Bao, Yu Zheng, and Yingcai Wu. [n. d.]. HomeFinder Revisited: Finding Ideal Homes with Reachability-Centric Multi-Criteria Decision Making. ([n. d.]). http://zjuvis.org/files/reach.pdf.
[40] Wenchao Wu, Jiayi Xu, Haipeng Zeng, Yixian Zheng, Huamin Qu, Bing Ni, Mingxuan Yuan, and Lionel M Ni. 2016. TelCoVis: Visual Exploration of Co-occurrence in Urban Human Mobility Based on Telco Data. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 935–944.
[41] Wenchao Wu, Yixian Zheng, Nan Cao, Haipeng Zeng, Bing Ni, Huamin Qu, and Lionel M Ni. 2017. MobiSeg: Interactive region segmentation using heterogeneous mobility data. In *Pacific Visualization Symposium*. 91–100.
[42] Yingcai Wu, Nan Cao, David Gotz, Yap-Peng Tan, and Daniel A Keim. 2016. A survey on visual analytics of social media data. *IEEE Transactions on Multimedia* 18, 11 (2016), 2135–2148.
[43] Yingcai Wu, Zhutian Chen, Guodao Sun, Xiao Xie, Nan Cao, Shixia Liu, and Weiwei Cui. 2017. StreamExplorer: A Multi-Stage System for Visually Exploring Events in Social Streams. *IEEE Transactions on Visualization and Computer Graphics* (2017).
[44] Yingcai Wu, Shixia Liu, Kai Yan, Mengchen Liu, and Fangzhao Wu. 2014. Opinionflow: Visual analysis of opinion diffusion on social media. *IEEE transactions on visualization and computer graphics* 20, 12 (2014), 1763–1772.
[45] Panpan Xu, Yingcai Wu, Enxun Wei, Tai-Quan Peng, Shixia Liu, Jonathan JH Zhu, and Huamin Qu. 2013. Visual analysis of topic competition on social media. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2012–2021.
[46] Yu Zheng, Licia Capra, Ouri Wolfson, and Hai Yang. 2014. Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology* 5, 3 (2014), 38.
[47] Yu Zheng, Yanchi Liu, Jing Yuan, and Xing Xie. 2011. Urban computing with taxicabs. In *Proceedings of the 13th International Conference on Ubiquitous Computing*. ACM, 89–98.
[48] Yu Zheng and Xiaofang Zhou. 2011. *Computing with spatial trajectories*. Springer Science & Business Media.

Received August 2017; revised February 2018; accepted March 2018