MobiVis: A Visualization System for Exploring Mobile Data

Zeqian Shen*   Kwan-Liu Ma†

Visualization & Interface Design Innovation (VIDi)
University of California, Davis

ABSTRACT

The widespread use of mobile devices brings opportunities to capture large-scale, continuous information about human behavior. Mobile data has tremendous value, leading to business opportunities, market strategies, security concerns, etc. Visual analytics systems that support interactive exploration and discovery are needed to extracting insight from the data. However, visual analysis of complex social-spatial-temporal mobile data presents several challenges. We have created MobiVis, a visual analytics tool, which incorporates the idea of presenting social and spatial information in one heterogeneous network. The system supports temporal and semantic filtering through an interactive time chart and ontology graph, respectively, such that data subsets of interest can be isolated for close-up investigation. "*Behavior rings*," a compact radial representation of individual and group behaviors, is introduced to allow easy comparison of behavior patterns. We demonstrate the capability of MobiVis with the results obtained from analyzing the MIT Reality Mining dataset.

Keywords: Mobile Data, Social-Spatial-Temporal Data Visualization, Information Visualization, Visual Analytics

Index Terms: I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction techniques;H.5 [Information Systems]: Information Interfaces and Presentation—User Interfaces

1 INTRODUCTION

The extensive use of mobile devices has impacted our everyday life, from how we communicate socially to how we do our work. In 2004, over 600 million handheld devices were sold, which outnumbers the amount of personal computers sold that year [1]. This number will quickly grow to over a billion. If we collect and analyze data captured by these mobile devices, we can discover communities, understand social behaviors, and infer important connections among events. This sort of information has tremendous value, suggesting business opportunities, market strategies, security concerns, etc. Efforts to collect such data using mobile phones have been ongoing [16, 5], and analysis of the collected data becomes a pressing problem. We are in need of new methods beyond the traditional data mining and statistical techniques.

Visualization has been shown to provide good overviews of large complex data. Visual analytics tools that allow us to "see" the mobile data and support interactive exploration and discovery are needed for extracting insights from the data. However, visual analysis of the mobile data presents several challenges, which demands new approaches to this important problem. First, mobile data is complex since it contains social, spatial and temporal information. All the social and spatial data are time-varying. GPS data provide subjects spatial locations, and we can infer social relations from calling data and Bluetooth proximity information. Next, mobile

---

*e-mail: zqshen@ucdavis.edu

†e-mail: ma@cs.ucdavis.edu

data can be very large-scale. For example, the MIT Reality Mining experiment [5] captures more than 3 million cell phone activities of one hundred subjects over the course of only nine months. Filtering techniques are necessary to support exploring such a large data set. Finally, one of the most important topics for social science study is to classify and compare human behavior patterns. Thus, effective visualization of individual and group behaviors are desired. We have studied how to visualize and analyze large social-spatial-temporal data.

In this paper, we present MobiVis, a visualization system, which supports intuitive exploration of social-spatial-temporal data captured by mobile devices. We introduce the idea of using a heterogeneous network to present both social and spatial information in one single 2D graph visualization. In order to support exploring large-scale mobile data, we have created a visual interface for performing temporal and semantic filtering. We have also developed "*behavior rings*," which enable an analyst to examine and compare individual and group behavior patterns. We have evaluated MobiVis using the MIT Reality Mining data set, and show in this paper that MobiVis can detect findings that conventional tools cannot reveal directly.

2 RELATED WORK

One of the most important experiments collecting mobile data was conducted at University of Helsinki. The Context project studied characterization and analysis of information about users' context and its use in proactive adaptivity [16]. ContextPhone was designed and developed as a software platform consisting of modules to support logging mobile phone usage and communications. It runs on mobile phones using Symbian OS (www.symbian.com) and the Nokia Series 60 Smartphone platform (www.series60.com). Many mobile data collecting projects based on Context have been conducted. Researchers at MIT Media Lab conducted an experiment consisting of one hundred subjects. They captured location, communication, Bluetooth proximity and device usage over the course of nine months [5]. Studies of mobile data that mainly rely on statistical analysis rather than visual analysis have been conducted. For instance, in [4], periodic behavior patterns were analyzed and used to identify social communities. Algorithms that learn routes between important locations and predict the next location when the user is moving were introduced in [12]. A few simple visualizations have been created for the MIT Reality dataset [13, 2]. However, none of them support advanced visual analysis and exploration.

Spatial-temporal data visualization have been widely studied [3, 11]. GeoTime is a system for displaying and tracking events, objects and activities over time and geography within a single, interactive 3-D view [7, 17]. It focused on visualizing the traces of a single event rather than group behaviors. For time-varying social network visualization, SoNIA incorporated innovative graph layout and animation algorithms to enable a temporally coherent animated "movie" of the changing social networks [14]. A similar approach for larger complex networks was introduced in [10].

Different from their work, we introduce a method to put time-varying spatial and social information all together in one visualization. A heterogeneous network consists of different types of entities (e.g., person and location) and relations (e.g., calling and proximity) is used. Visual representation of the heterogeneous net-

work and semantic filtering techniques based on associated ontology graphs [18] are incorporated into our system.

Timelines are good for presenting trends of time-varying data [8]. For example, in [19], univariate time series data are visualized in the form of calendars. Thus, recurrent patterns and trends on multiple time scales can be observed simultaneously. Timelines can also be used for temporal filtering. Timebox is a widget supporting dynamic query on time series data [6]. It allows users to filter data by selecting arbitrary timespans along the timeline. In MobiVis, a 2D time chart equipped with more advanced user interactions is implemented.

3 MOBIVIS

MobiVis is a visual analytics system designed for exploring and discovery of mobile data. We address the challenges of visualizing complex social-spatial-temporal data in its design and implementation. In this section, we first introduce a methodology to formulate the data into a heterogeneous network. Next, we discuss the interactive time chart and ontology graph, which enable temporal and semantic filtering in MobiVis, respectively. Finally, we introduce behavior rings that can reveal periodical behavior patterns of individuals and groups.

3.1 Data Formulation for Visualization

Mobile data contains both spatial and social information. GPS units on mobile phones record the spatial locations of subjects. Bluetooth proximity relations and calls can infer social relations among the subjects. Usually, spatial information is visualized on a geographical map, and relational information is presented as a social network in a node-link diagram. Therefore, two different mappings of the data exist. In experiments performed by Klein et al. [9], they observed that "switching between completely different visualizations confused the users." A good visualization system should use a minimum number of visualizations and construct in a unified fashion. We therefore decide to integrate both spatial and social views in a single visualization. This would also help find the hidden correlation between social and spatial information.

The social information including phone calls and proximity can be formulated as an undirected graph, where each vertex denotes a person, and edges denote calls or proximity relations between persons. Therefore, the edges are time-varying. The spatial information in mobile data can be defined as $SP = \{(p_i, l_j, t) : p_i \in P, l_j \in L\}$, which denotes that person p_i is at location l_j at time t. Since it can be thought of as time-varying relations between persons and locations, we can also formulate the spatial information as an undirected graph. Both persons P and locations L are converted into vertices. The "locate at" relations, SP, are converted into time-varying edges. Then, we can integrate the social graph and spatial graph into one graph, which is a time-varying heterogeneous graph.

An example is illustrated in Fig. 1 to demonstrate this idea. The spatial view on the top [1] shows that person A moves along the red line from the CS Department to the Library, and arrives at Train Station. Person B moves along the blue line from the Univ. Center to the Library, and arrives at the Central Park. The social network view at the bottom shows the relations among A, B and the other two persons. We integrate both views into the heterogeneous network on the right by converting persons and locations to vertices. In the network, location vertices are drawn as triangles, and an edge connecting a person and a location denotes that the person has visited the location. Note that the edges in the heterogeneous network are time-varying. We draw all of them in the figure for demonstration purpose. The mobile data often contains other types of information, such as subjects' occupations, ages and personalized phone settings. These information can also be integrated into

---

[1]The map of University of California, Davis is retrieved from maps.google.com
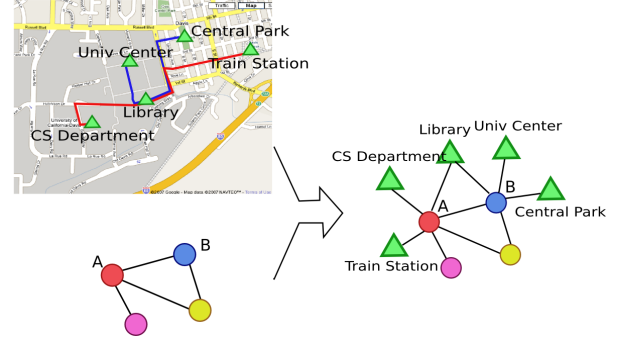

Figure 1: Using heterogeneous network to present both spatial and social information in mobile data. The spatial view at the top and the social network view at the bottom are integrated into one single heterogeneous network on the right.

the heterogeneous graph using the similar formulation method (for more details, please read our previous paper [18]).

Visualization of this time-varying heterogeneous graph is used as the centerpiece in MobiVis. Therefore, we want to introduce formal definition of time-varying heterogeneous graphs. The graph is defined as $G = (V, E, vt, et)$. The nodes are static, while edges are dynamic. $V = \{v_1, v_2, ..., v_n\}$ denotes the vertex set and $E = \{(v_i, v_j, time_k) : v_i, v_j \in V\}$ denotes the edge set, where $time_k(day, hour)$ denotes the valid timespan of the edge. The associated ontology graph is defined as $OG = (T_V, T_E)$. $T_V = \{t_1, t_2, ..., t_m\}$ is a set of vertex types and $T_E = \{(t_i, t_j) : t_i, t_j \in T_V\}$ is a set of edge types. vt denotes a mapping from V to T_V that associates a vertex to its type. If v is a vertex in the graph, $vt(v)$ denotes the type for vertex v. Similarly, et denotes a mapping from E to T_E that associates an edge with its type. If e is an edge in the graph, $et(e)$ denotes the type for edge e. It is important to note that a heterogeneous graph cannot have vertices and edges with types that are not presented in its associated ontology graph. In other words, T_V and T_E are, respectively, supersets of the vertex and edge types that occur in graph G. Both time-varying heterogeneous graphs and the associated ontology graphs discussed in this paper are undirected.

3.1.1 MIT Reality Mining Dataset

In this paper, a mobile dataset collected by MIT Media Lab in the Reality Mining experiment [13, 5] is used. In the experiment, Nokia 6600 smart phones, pre-installed with logging software, were distributed to one hundred subjects, of which seventy-five are either students or faculty in MIT Media Laboratory, and twenty-five are incoming students at MIT Sloan business school. The experiment was run over the course of the 2004-2005 academic year. Over 500,000 hours of continuous data on daily human behaviors were captured. Moreover, subjects were asked to take surveys regarding their social activities and interactions with others. Part of the MIT Reality Mining dataset is publicly available in the form of MySQL database. The available dataset is somewhat noisy. After the cleaning process, we are able to obtain a dataset that contains social activities of 83 anonymous subjects from August 2004 to March 2005. The dataset contains personal information obtained from the user survey, Bluetooth proximity relations, phone calls, and locations.

The ontology of the derived heterogeneous network is shown in Fig. 2. There are 20 types of nodes in the network, including person, location, and 18 survey questions. There are two types of edges between person nodes: phone calls and proximity relations. In the dataset, a person's current location is indicated by the celltower to which his/her mobile phone connects. Geographic information of celltower locations are not available. Fortunately, in the experiment,
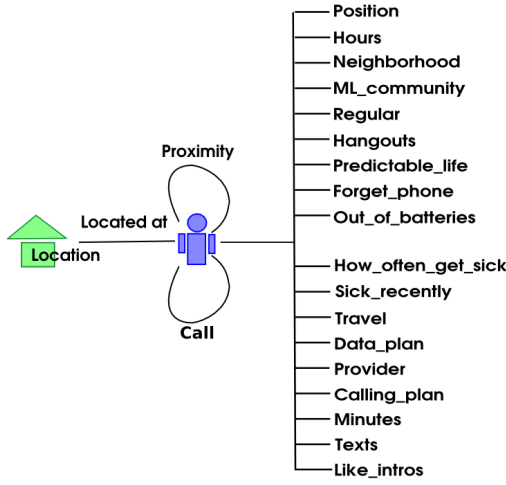
Figure 2: Ontology graph of the mobile social network derived from MIT Reality Mining dataset. The data is formulated as a time-varying heterogeneous network. There are 20 node types, including person, location, and 18 survey questions. There are 21 edge types. Two types of edges exist between two person nodes: call and proximity. Nodes are static, while edges associated to proximity relations, calls, and locations are time-varying.

subjects are asked to name their current location when a previously unseen celltower is encountered. We believe, the personalized cell-tower names (e.g., home, media lab, sloan, and parents) are more valuable than geographical locations for social behavior analysis. The problem is that there are too many distinct celltower names. It is not appropriate to map all of them as location nodes. We decide to put the names into several major categories, such as home, work, travel, etc. For example, "Jon home" and "home pearl st" are both considered "home". The classification is done automatically based on manually picked keywords. Those names that cannot be classified are put into the category "others." Each category is mapped to a location node. Edges associated to proximity relations, calls, and locations are time-varying. Besides person and location, each survey question is considered as a node type and answers as nodes of that type. An edge between such an answer node and a person node indicates that the subject gave such an answer for the particular survey question.

3.2 Filtering Techniques

The continuous human behaviors captured by mobile devices can be very large. In MIT Reality Mining dataset, for instance, there are thousands of Bluetooth encounters during a regular weekday. The visual analytics tool should allow users to filter the sheer number of data and isolate subsets of interest. In MobiVis, we incorporate two filtering methods: semantic filtering using ontology graph and temporal filtering through a time chart. Their design and implementation are discussed below.

3.2.1 Semantic Filtering Using Ontology Graph

The heterogeneous network derived from the mobile data can be too large to visualize with limited screen space and resolution. In MobiVis, we use the semantic information that resides on the nodes and edges to filter the data and find subgraphs of interest. The semantic filtering technique introduced in [18] is used. An ontology graph derived from the heterogeneous network is drawn as a semantic overview (See Fig. 2). It contains all node types and possible relations between them. The system allows users to interactively select node types in the ontology graph. As soon as node types TS are chosen, a subgraph including only nodes with selected

types will be derived, i.e., $G[TS] = (VS_{TS}, ES_{TS}, vt, et)$ of a selected set of node types $TS \subseteq T_V$, where $VS = \{v \in V : vt(v) \in TS\}$ and $ES = \{(v_i, v_j) \in E : vt(v_i), vt(v_j) \in TS\}$. The subgraph is drawn in the main visualization window of MobiVis. Fig. 3 shows the sub-graph consisting of persons, positions and hangout places obtained by selecting these node types in the ontology graph. Linlog [15], a force-directed graph layout algorithm, is used, because it is general enough to work with many types of networks, relatively easy to implement, and adaptable to satisfy different requirements. The visualization shows that there are five major groups of people: Sloan students, Media Lab graduates, student, new graduates and senior graduates. Node sizes are determined by their degrees. The three most popular hangout places are gym, restaurant/bar and friend's place.

3.2.2 Temporal Filtering Using Interactive Timechart

For mobile data, presenting and exploring the temporal information are critical. Human behaviors in mobile data exhibit repetitive patterns and trends. Therefore, the visual representation should make the repetitive patterns more salient and allow users to select recurrent timespans. Traditional 1D timeline is not effective in this case. In MobiVis, we choose 2D time charts with both axes denoting different time scales, and colors of blocks in the chart denoting time-varying values (See Fig. 5). In this example, the vertical and horizontal axes denote time and date, respectively. It is ideal for observing daily patterns. The vertical lines separate weeks, and horizontal lines denote hours. The color of a block denotes the location of subject 57 at the time indicated by its coordinates. Red denotes work, blue is for home, and green is for entertainment places. We can see that subject 57 usually goes to work around noon, and comes back home around midnight. Users can change time scales of the axes to fit their tasks. For example, using days of a week for vertical axis, and weeks for horizontal axis, makes weekly patterns clearer.

The time chart also enables users to select more than continuous timespans. Drawing a rectangle on the time chart, users can specify an advanced time window, which is defined as $TW(start\_day, start\_hour, end\_day, end\_hour)$. A moment is in the timewindow (i.e., $time(day, hour) \in TW$) if $day \in [start\_day, end\_day)$ and $hour \in [start\_hour, end\_hour)$. The time chart is linked with the heterogeneous graph visualization, which shows an induced graph contains aggregated activities within the selected timewindow, i.e., $G[TW] = (V, ES_{TW}, vt, et)$, where $ES_{TW} = \{(v_i, v_j, time_k) \in E : time_k \in TW\}$. Such time windows are very useful in time-varying data analysis. For instance, users can investigate recent night life of a subject by selecting activities between 9pm and midnight for the last three weeks (See Fig. 5). To isolate the same activities, in 1D timeline, users need to select 9pm-midnight timespans for everyday of the last three weeks.

The time chart has two modes. One mode shows the occurrence of time-varying neighbors of a selected node as the example above, which shows the location neighbors of person node 57. Another mode shows the occurrence of the selected node. The occurrence of a node is defined as the sum of valid time-varying edges connecting to it at the moment. In other words, it is how many other nodes that have the selected node as neighbors. Fig. 4 shows the occurrence of location node "work," i.e., the number of subjects at work, over time. This example reveals clear patterns at different timescales. On a daily basis, the most common working hours are $11am - 7pm$. On a yearly basis, there is a large break near the end of December because of the Christmas holiday season.

3.3 Behavior Rings

The time chart can reveal the behaviors of a single subject or group, but does not support comparison. For example, we want to compare subjects' weekly calling patterns. One way is to display multiple time charts side by side for this purpose but there is a limit on

Figure 3: Network with person, position and hangout places. It shows subjects' occupations and usual hangout places from the user survey. There are five major groups of people: Sloan students, Media Lab graduates, students, new graduates, and senior graduates. The three most popular hangout places are gyms, restaurants/bars and friend's places.
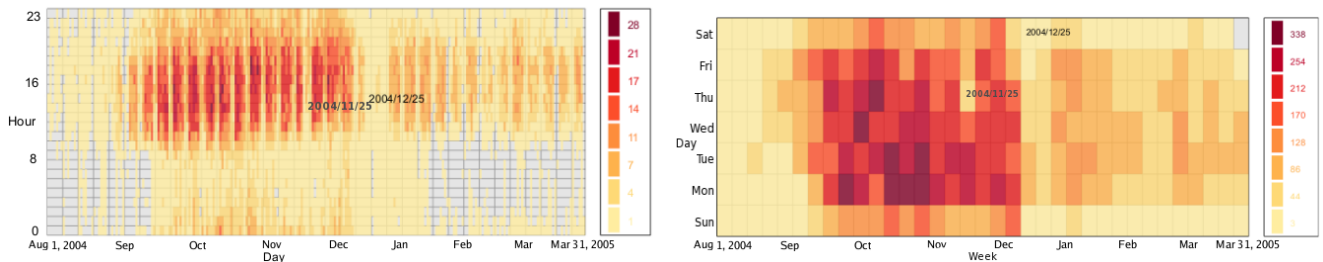

(a) In the daily view, the vertical axis denotes hours in a day. The most common working hours are from 11am to 7pm. The periodical vertical breaks indicate weekends.

(b) In the weekly view, the vertical axis denotes days in a week. The weekly working pattern is revealed.

Figure 4: Time chart whowing the number of subjects at work. Repetitive patterns can be observed in views of different timescales. In both views, there are two large breaks: Thanksgiving and Christmas holidays.
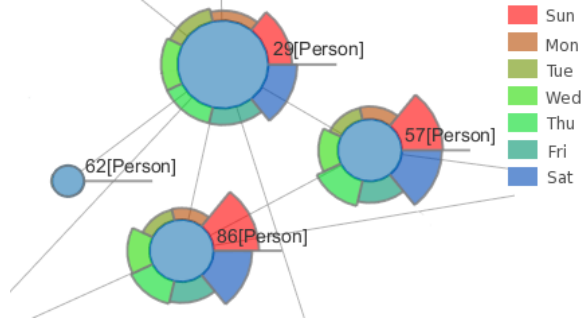
(a) A daily ring of calls illustrates the frequency of calls every week day. Subject 29, 57, and 86 made calls more frequently. In addition, this example reveals that a subject often made more calls during weekends than weekdays.
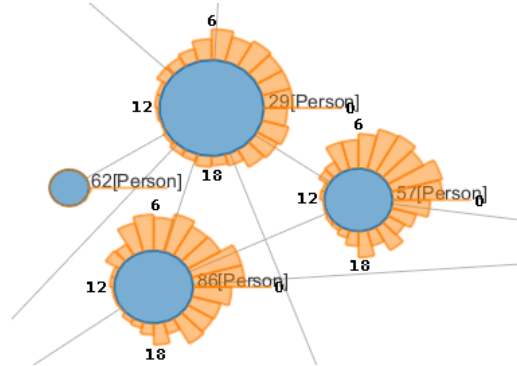
(b) An hourly ring of calls shows the frequency of calls every hour. Subject 29, 57, and 86 made more calls during night than daytime.

Figure 6: In a behavior ring, occurrences of selected activities are arranged radially around a subject in counter-clockwise order. The rings provide an abstraction of subjects' behaviors and allow users to compare across different individuals in the network view. In this example, daily and hourly rings of phone calls for four subjects are illustrated. More phone calls are made after work, since only phone calls among participants, who are colleagues, are counted.
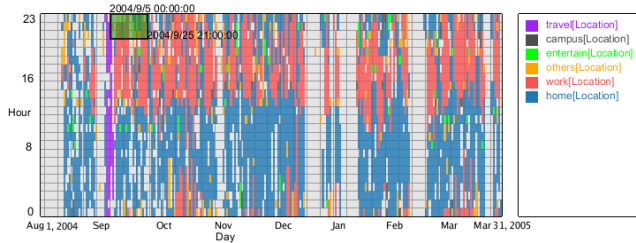

Figure 5: Time chart showing the locations of subject 57 over time. Subject 57 usually goes to work around noon, and returns home around midnight. Drawing a time window in the time chart, the activities between 9pm and midnight and within the three-week span from 2004/9/5 to 2004/9/25 are selected.

how many charts can be simultaneously displayed and how much information can be shown on each chart. An alternative visual representation that is more compact and integrable with other representations is needed. We have developed a radial-layout design resembling Florence Nightingale's coxcomb. We call it *behavior rings*. Like coxcombs, we use radially lay-out, pie-shaped wedges to represent time-varying information, such as some particular activity including phone calls, proximity relations, and locations accumulated over a selected period. The whole period (from August 2004 to March 2005) is used by default. The size of the wedge can represent the accumulated occurrence of the activity during a user-specified recurrent time, e.g., every Saturday. Color or texture may be used to represent other information. Unlike coxcombs, we reserve a large portion of the inner region to display additional information. This additional information could be nested behavior rings or other visual or textual representations. Furthermore, we use behavior rings to represent nodes of a social network. This is an interesting and important option. Fig. 6(a) shows an example of daily behavior rings presenting accumulated occurrences of phone calls of every day in a week from August 2004 to March 2005. Each ring contains 7 wedges, which denote 7 days in a week. In the visualization, the wedges are arranged counter-clockwise, and the starting point is marked by a longer spike. Subjects 29, 57, and 86 made calls more often than the other two subjects. In addition, we

can also see that these three subjects made more calls during weekends than weekdays. We further investigate these subjects' hourly calling patterns during the same period by enabling the hourly behavior ring. In an hourly ring, wedges denoting every hour in a day are arranged counter-closewise as in Fig. 6(b). Subjects 29, 57, and 86 made more phone calls during the night than the daytime. Both daily and hourly calling patterns show that these three subjects call more often after work. The reason is that only calls among participating subjects, who are colleagues, are counted in the dataset. Therefore, they usually do not need to call each other when they are at work.

A network of behavior rings display much richer information about the the involved subjects. Together with cluster operations, the user can interactively combine nodes to see the behavior of a large group. In the network view, users are allowed to circle a region and all person nodes within the region are selected. Then, a macro ring, which shows the aggregated activities of the selected person nodes, is drawn. In Fig. 7, we compare the weekly behavior patterns of subjects with different positions by enabling group behavior rings of proximity relations. Subjects are grouped according to their academic positions. Behavior rings of every group are visualized. Groups including "New Grad", "Student", and "Media Lab Grad" gather together very often, while subjects belonging to Sloan business school do not. Senior graduate students are often alone. In addition, we can find that persons are closer to each other during weekdays when they are at work.

Behavior rings provide abstractions of subjects' dynamic behaviors in the network view and enable users to find subjects with interesting behaviors for further investigation.

4 RESULTS

Organizational behavior patterns in MIT Reality Mining data are analyzed to demonstrate the capability of MobiVis for visual analysis of mobile data. Moreover, examples of using MobiVis to find and resolve inconsistencies and uncertainties in data are presented in this section.

4.1 Inferring Friendship Network From Observed Behaviors

One of the key issues in social network analysis is friendship. Social scientists usually obtain the friendship network by user survey.
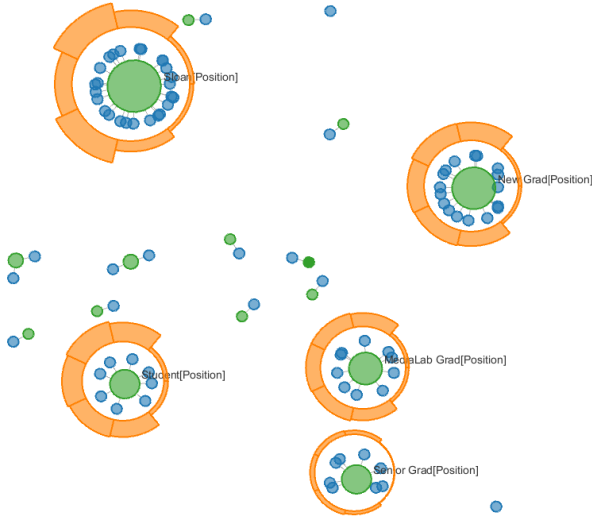
Figure 7: Comparison of group behaviors using behavior rings. Persons are divided into groups by their academic positions. Rings of proximity activities are enabled. The visualization reveals that persons in group "Student", "Media Lab Grad", and "New Grad" are often close to someone else. People are closer to each other during weekdays.

In addition to the self-reported information, we can also infer the friendship network from observed human behaviors. For example, persons who often hang out together during weekends are likely to be friends. In Fig. 8, examples are presented to show how MobiVis can help isolate important social behaviors and infer the friendship network.

First, we try to infer the friendship network from phone calls. The assumption is that friends often call each other. We select all person nodes and phone calls among them. Positions are also added into the visualization (See Fig. 8(a)). Each edge denotes the calling relation between two persons, and its width denotes the total duration of calls between them. There are two closely connected friendship groups. The group at the bottom left consists of students from Sloan business school, and the one at top right is from Media Lab. Comparing the connection density of two groups, we can conclude that students from Sloan are more likely to make friends with each other. Inside the group of Media Lab, senior graduate students have less friends.

Besides phone calls, friends also tend to hang out with each other in their spare time. We choose to examine the proximity relations during Saturday nights, which are defined as periods between 11pm on Saturday and 3am Sunday morning. All Saturday nights are selected on the time chart. Person nodes, position nodes, and proximity edges are added into the visualization (See Fig. 8(b)). The friendship network presented here is very similar to the one inferred from phone calls. People from Sloan rarely get together with those from Media Lab on Saturday nights.

From the study above, we can draw the conclusion that in the MIT Reality Mining experiment, subjects are more likely to make friends with their colleagues under the assumption that phones and proximity relations on Saturday nights can infer friendships.

4.2 Comparing Group Behavior Patterns

We are also interested in comparing the daily behavior patterns of different groups in the friendship network. First, we enable the behavior rings of proximity relations for each group (See Fig. 9(a)). Each ring has twelve wedges, which denote the activities during every hour in a day. The size of each wedge indicates how close a

group of subjects are to others. In the behavior ring of Sloan students at bottom left, the wedge size increases at 9am, while in those rings of Media Lab students, the wedge increases around 12pm. Thus, the group of Sloan students gather together two hours earlier than those groups of Media Lab students. Because subjects are closer to each other during work hours, their closeness can suggests whether they are at work or not. Therefore, we further study the daily working pattern by enabling the behavior rings of the time for which subjects stay at location "work." In Fig. 9(b), the behavior ring of Sloan students shows that they go to work around 9am. In the rings of Media Lab students, the wedges become significant around 10am. Therefore, we can draw the conclusion that Sloan students go to work earlier than those from the Media Lab. In addition, senior graduate students spend the longest time at work, because all the wedges in their behavior ring are of considerable size.

4.3 Resolve Errors and Uncertainties in the Data

The locations of a subject in the dataset are derived from user-defined celltower names. We use a fairly simple approach to classify the names based on manually selected keywords, which left many names unrecognizable. The analysis of individual behaviors using MobiVis can actually help us gain a better understanding of the celltower names and improve our classifier. Fig. 10 shows location occurrence of subject 24. We can see that most of time the subject is at location "others," which means the celltower name cannot be determined using our keywords. Moving the mouse over those blocks, we can see the actual celltower name: "Burton connor." We observe that subject 24 can be seen at this location every night. Therefore, it is most likely his/her home. Searching "Burton connor" online, we find that it is an undergraduate dorm at MIT.

Errors in the dataset can be revealed during the interactive exploration using MobiVis. Fig. 11 shows locations of subject 92. We can see that from July 2004 to October 2004, the subject appears to have stayed at home all the time, which is abnormal. We checked the original database and found the following record:

```
starttime            endtime            location
2004-07-10 17:57:22  2004-10-07 17:57:54  Home
```

indicates that subject 29 stays at home from July 10th to Oct. 7th, which cannot be true. We believe this is an error in the log file. The ending date of the timespan is probably 2004-07-10. The month and day got switched for some reason.

5 CONCLUSION AND FUTURE WORK

In this paper we present MobiVis, a visual analysis system for exploring and understanding social-spatial-temporal mobile data. In MobiVis, both spatial and social data are transformed into relational data and presented as a heterogeneous network. In order to handle the sheer number of observed activities, an interactive time chart and an ontology graph are incorporated for temporal and semantic filtering, respectively. The time chart provides an overview of time-varying activities in the network, reveals repetitive behavior patterns, and enables filtering based on advanced time windows. The ontology graph is used as a guide for exploration based on semantic information of nodes and links in the network. The resulting networks are intuitively displayed and respond to user interaction. We developed behavior rings to allow users to easily compare behavior patterns of individuals and groups in the network view. Visual analytics systems such as MobiVis are expected to enable important discoveries in data exploration of emerging mobile data.

MobiVis are suitable for both expert and novice users. The next step is to incorporate more advanced data analysis methods, so that the system is capable of performing more sophisticated analytic tasks. For example, clustering the subjects based on their eigenbehaviors [4] can help identify common behavior patterns, determine
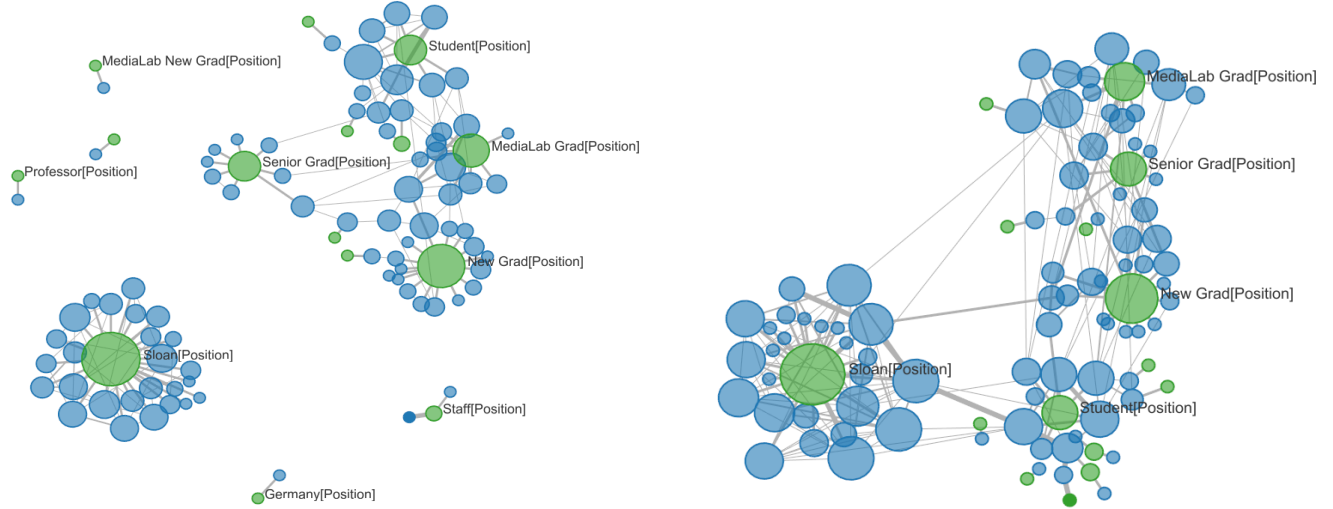
(a) Calling network during the whole period. There are two closely connected groups. The one at the bottom left consists of subjects from Sloan business school, and the other one indicates Media Lab. Senior graduate students are less connected than others.

(b) Network of proximity relations on Saturday nights. There are also two clusters: Sloan business school and Media Lab.

Figure 8: Inferring friendship network from observed behaviors. Phone calls and proximity relations on Saturday nights indicate that subjects are more likely to call or hangout with colleagues, respectively. With the assumption, that phone calls and gathering during weekend night can infer friendships, we can conclude that subjects are more likely to make friends with their colleagues in the MIT Reality Mining experiment.
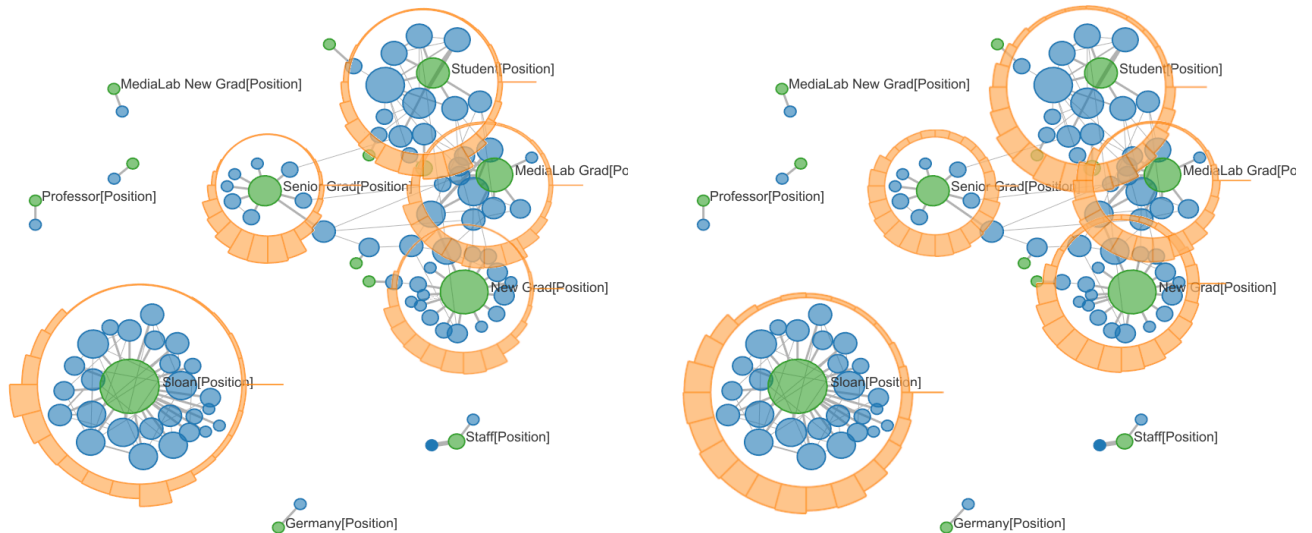
(a) Comparison of daily proximity patterns of groups in the friendship network. The group of Sloan students at bottom left gather earlier than those groups of Media Lab students.

(b) Comparison of daily working patterns of groups in the friendship network. The total time of groups of persons staying at location "work" are illustrated by behavior rings. Students of Sloan business school go to work earlier than those from Media Lab. Senior graduate students spend the longest time at work.

Figure 9: Comparison of behavior patterns of groups in the friendship network derived by phone calls. Group behavior rings are added to Fig. 8(a). For each group, its behavior ring illustrates the total frequency of a certain activity during every hour of a day. Both visualizations reveal that students of Sloan business school have different working hours from those of Media Lab.
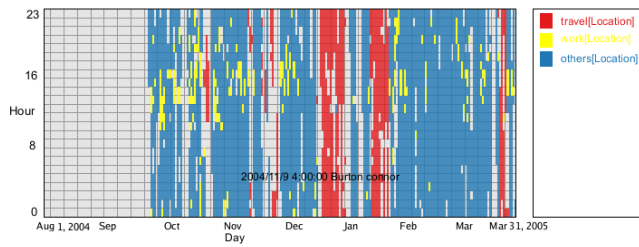
Figure 10: Location occurrence of subject 24. The subject stays at a location with a unrecognizable name, "Burton connor," every night. Based on the the pattern of staying time, it is inferred to be home for the subject. Investigation results of "Burton connor" confirm that it is an undergraduate dorm in MIT.
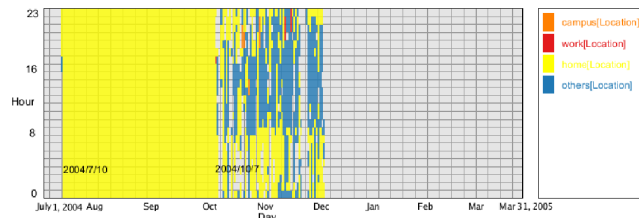

Figure 11: Location occurrence of subject 92. Visualization shows an abnormal activity, i.e., the subject stays at home all the time from July to October. Checking the original data reveals an error in the dataset.

behavioral similarity between both individuals and groups, and enable accurate classification of group affiliations.

For mobile data that contain activities over a very long period of time, the scalability of the time chart could be an issue. The visibility of behavior patterns relies on the number of pixels per cell entry in the chart. A zoomable interface for the time chart can solve the problem.

In our study of the MIT Reality Mining dataset, we came across a large number of errors and uncertainties in the data. For example, the survey answers entered by subjects contain typos, and user defined celltower names are too abstract to understand. Other errors may be caused by crashes of the capturing software running on mobile phones. These errors are surely common in mobile experiments. MobiVis can manifest errors and uncertainties in the raw data set throughout the analysis process. It would be useful to incorporate supervised machine-learning methods into the system. Therefore, when an uncertainty is resolved manually by users, similar ones will be automatically identified and resolved.

ACKNOWLEDGEMENTS

REFERENCES

[1] http://news.bbc.co.uk/1/hi/business/4257739.stm.

[2] http://reality.media.mit.edu/viz.php.

[3] P. Compieta, S. D. Martino, M. Bertolotto, F. Ferrucci, and T. Kechadi. Exploratory spatio-temporal data mining and visualization. *J. Vis. Lang. Comput.*, 18(3):255–279, 2007.

[4] N. Eagle and A. Pentland. Eigenbehaviors: Identifying structure in routine. In *Proc. Roy. Soc. A (in submission)*, 2006.

[5] N. Eagle and A. Pentland. Reality mining: sensing complex social systems. *Personal Ubiquitous Comput.*, 10(4):255–268, 2006.

[6] H. Hochheiser and B. Shneiderman. Dynamic query tools for time series data sets: timebox widgets for interactive exploration. *Information Visualization*, 3(1):1–18, 2004.

[7] T. Kapler and W. Wright. Geotime information visualization. In *INFOVIS '04: Proceedings of the IEEE Symposium on Information Visualization (INFOVIS'04)*, pages 25–32, Washington, DC, USA, 2004. IEEE Computer Society.

[8] G. M. Karam. Visualization using timelines. In *ISSTA '94: Proceedings of the 1994 ACM SIGSOFT international symposium on Software testing and analysis*, pages 125–137, New York, NY, USA, 1994. ACM.

[9] P. Klein, F. Mller, H. Reiterer, and M. Eibl. Visual information retrieval with the supertable + scatterplot. In *IV*, pages 70–75, 2002.

[10] G. Kumar and M. Garland. Visual exploration of complex time-varying graphs. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):805–812, 2006.

[11] M.-P. Kwan and J. Lee. *Geovisualization of human activity patterns using 3D GIS: A time-geographic approach*, chapter 3, pages 48–66. Oxford University Press, 2004.

[12] K. Laasonen. Clustering and prediction of mobile user routes from cellular data. In *PKDD*, pages 569–576, 2005.

[13] M. J. Lambert. Visualizing and analyzing human-centered data streams. Master's thesis, Massachusetts Institute of Technology, USA, 2005.

[14] J. Moody, D. McFarland, and S. Bender-deMoll. Dynamic network visualization. *American Journal of Sociology*, 110(4):1206–1241, 2005.

[15] A. Noack. Energy-based clustering of graphs with nonuniform degrees. In *Proc. of Graph Drawing '05*, pages 309–320, 2005.

[16] M. Raento, A. Oulasvirta, R. Petit, and H. Toivonen. Contextphone: A prototyping platform for context-aware mobile applications. *IEEE Pervasive Computing*, 4(2):51–59, 2005.

[17] R. H. Ryan Eccles, Thomas Kapler and W. Wright. Stories in geotime. In *VAST '07: Proceedings of the IEEE Symposium on VAST (VAST '07)*, 2007.

[18] Z. Shen, K.-L. Ma, and T. Eliassi-Rad. Visual analysis of large heterogeneous social networks by semantic and structural abstraction. *IEEE Transactions on Visualization and Computer Graphics*, 12(6):1427–1439, 2006.

[19] J. J. van Wijk and E. R. V. Selow. Cluster and calendar based visualization of time series data. In *INFOVIS '99: Proceedings of the 1999 IEEE Symposium on Information Visualization*, page 4, Washington, DC, USA, 1999. IEEE Computer Society.