TrajGraph: A Graph-Based Visual Analytics Approach to Studying Urban Network Centralities Using Taxi Trajectory Data

Xiaoke Huang, Kent State University
Ye Zhao, Kent State University
Jing Yang, UNC Charlotte
Chong Zhang, UNC Charlotte
Chao Ma, Kent State University
Xinyue Ye, Kent State University

Abstract— We propose TrajGraph, a new visual analytics method, for studying urban mobility patterns by integrating graph modeling and visual analysis with taxi trajectory data. A special graph is created to store and manifest real traffic information recorded by taxi trajectories over city streets. It conveys urban transportation dynamics which can be discovered by applying graph analysis algorithms. To support interactive, multiscale visual analytics, a graph partitioning algorithm is applied to create region-level graphs which have smaller size than the original street-level graph. Graph centralities, including Pagerank and betweenness, are computed to characterize the time-varying importance of different urban regions. The centralities are visualized by three coordinated views including a node-link graph view, a map view and a temporal information view. Users can interactively examine the importance of streets to discover and assess city traffic patterns. We have implemented a fully working prototype of this approach and evaluated it using massive taxi trajectories of Shenzhen, China. TrajGraph's capability in revealing the importance of city streets was evaluated by comparing the calculated centralities with the subjective evaluations from a group of drivers in Shenzhen. Feedback from a domain expert was collected. The effectiveness of the visual interface was evaluated through a formal user study. We also present several examples and a case study to demonstrate the usefulness of TrajGraph in urban transportation analysis.

Index Terms—Graph based visual analytics, Centrality, Taxi trajectories, Urban network, Transportation assessment

1 Introduction

Nowadays, large amounts of taxi trajectory data are collected and utilized by transportation administrations, companies, and researchers. The data provides *real* situations from which real traffic flows can be extracted and city-wide transportation patterns can be discovered. In this paper, we propose a new visual analytics approach, TrajGraph, to studying urban dynamic patterns using massive taxi trajectory data, with a specific focus on discovering the importance of different parts of city networks in transportation. Discovering this knowledge, such as the time varying hubs and backbones of road networks, is critical for optimizing urban planning and amending city operations. However, there is a lack of effective approaches to completing this task due to several challenges.

First, most existing methods cannot support effective analysis of the important roles of city roads in real traffic situations. Classic grid-based models manage spatial data in the Euclidean space of a city (e.g., [15]), which does not effectively represent the network structure of roads. Graph methods based on static street networks have been used in geography and transportation to study road network structures and topologies in some city areas (e.g., [5, 27]). However, they cannot reflect the real transportation roles of streets since they do not utilize real traffic data. Although recent researches have been conducted on utilizing urban trajectory data (e.g., [37, 39]), they do not create graphs representing city-wide road networks and do not support interactive analyses of road importance.

- *Xiaoke Huang, Ye Zhao, and Chao Ma are with the Department of Computer Science, Kent State University. Email: zhao@cs.kent.edu; xhuang8,cma1@kent.edu*
- *Jing Yang and Chong Zhang are with the Department of Computer Science, University of North Carolina at Charlotte. E-mail: jyang13,czhang22@uncc.edu*
- *Xinyue Ye is with the Department of Geography, Kent State University. E-mail: xye5@kent.edu*

Manuscript received 31 Mar. 2015; accepted 1 Aug. 2015; date of publication 20 Aug. 2015; date of current version 25 Oct. 2015. For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org. Digital Object Identifier no. 10.1109/TVCG.2015.2467771

Second, domain users need to interactively select city roads and visually analyze their roles. Visualization is desired since it allows the users to incorporate their domain knowledge and human intelligence in the exploratory analysis process. However, the scale and complexity of the data make interactive visualization a challenging task, since street networks in large cities are very big and complex, and the situation is greatly confounded when incorporating dynamic traffic information from massive trajectory data. To facilitate interactive visual exploration, effective and efficient computational methods of road importance should be tightly integrated with intuitive visualizations. Domain users have been impeded due to the absence of such visual analytics systems.

Designed to address these challenges, TrajGraph is a new visual analytics approach which supports domain users, such as city planners and transportation researchers, in finding and comparing the time-varying transportation roles of urban roads utilizing real traffic information. It tightly integrates graph modeling, graph analytics, and visualization techniques. Unlike existing methods, our approach generates a graph from massive taxi trajectory data to represent an urban road network and its traffic information in the whole city. This graph is then aggregated, analyzed, and visualized to allow users to interactively discover the dynamic urban patterns.

In particular, a road segment is mapped to a vertex if taxis pass it, and edges are added between two connected road segments if and only if there are taxis traveling between them. The vertex weight is used to represent traffic information such as the average travel speed to reflect dynamic use of the roads. Therefore, the graph manifests the traffic information in addition to road connectivity in a network. Consequently, TrajGraph provides a means to discover dynamic traffic patterns through graph computing and analysis. To the best of our knowledge, this approach of graph construction and weight definition from taxi trajectories has not been used in previous work.

To address the scalability challenge, a graph partitioning algorithm is applied to aggregate neighboring street vertices into partitions. Thus, a street-level graph is converted into a region-level graph with reduced size. Here the "region" refers to a combination of streets instead of a conventional spatial cell. Users can flexibly select the number of regions to be used in analysis. Therefore, our system enables effective visualizations and interactions over regions and exploratory visual analytics becomes possible. Moreover, graphs can be generated and flexibly studied for different time periods to characterize the temporal variation of importance. For example, users can study the changing traffic patterns by exploring multiple graphs constructed from taxi trajectories in different periods (e.g., every two hours) of a day. Therefore, TrajGraph enhances its usefulness and flexibility to meet the requirement of domain users.

The Pagerank and betweenness centralities are computed for graph vertices, which directly reflect the important roles of city roads in transportation such as hub and backbone regions. Using different definitions of vertex weights, such as street distance, average taxi travel speed, or travel time, the centralities can reveal a variety of critical traffic patterns. More centrality metrics (e.g., closeness) can also be added to our system.

A node-link graph view visually presents the centralities. On this view users can easily select regions of interest according to their importance. A map view coordinated with the node-link graph view then allows users to visually examine and compare their importance and traffic information (e.g., traffic flow, speed) in geographic context. Furthermore, on a temporal information view, the temporal changes of the importance and traffic information are depicted and can be well compared among regions. The temporal changes can also be depicted as rose charts on the map. The three views effectively support knowledge exploration related to urban region importance.

TrajGraph was evaluated from multiple aspects. First, we evaluated whether the importance of regions computed from centralities matched the opinions of local residents. The evaluation was conducted with a group of drivers in Shenzhen, the city where the trajectory data came from. Second, we collected expert feedback from a geographer and urban planner on the usability of the system. Third, we conducted a user study to test whether the coordinated node-link graph view improved user performance in visual exploration on the map view. We also present a case study and report the implementation and computation performance of the system. These studies indicate that TrajGraph is efficient and effective in utilizing taxi trajectory data to gain transportation knowledge.

2 Related Work

Urban modeling and planning: Urban researchers have developed strategies that improve safety, mobility and sustainability in transportation systems [25, 34]. Conventional platforms for urban transportation forecasting, planning and analysis (e.g., [6, 8]) have enhanced human ability to simulate, maintain, and operate transportation infrastructure. Many studies have been aimed to discover the relationship between travel behavior and street network structure (see a literature review in [23]), in which human travel data was recently used in statistical regression analyses. Here the travel data was studied in aggregate levels, such as the congestion indexes (e.g., Travel Time Index) of U.S. metropolitan areas [23]. These tools and approaches are not developed utilizing real trajectory data.

Graph based urban study: In geography, Rozenblat and Melancon [28] developed a project of modeling and analyzing multilevel geographical networks produced in complex dynamic systems, such as a hierarchy of cities generated according to their control of multinational firms. Their work utilized centralities and node-link visualizations in modeling such networks. However, they did not specifically use real-world population trajectory data to construct graphs representing city roads/regions and traffic information. Using graph-based methods, Rosvall et al. [27] and Porta et al. [26] studied and compared city networks. Meanwhile, Buhl et al. [5] analyzed street network topology to study evolution and functional properties. Crucitti et al. [10] studied graph centralities of different world cities to compare planned and self-organized cities. However, these methods used a small part of a city to create a graph directly from street geometries. They identified structural or topological features of cities, but did not study transportation and traffic patterns, since no real traffic information was involved. In contrast, we utilize taxi trajectory data to create urban graphs and study dynamic transportation patterns through interactive visualization.

Taxi trajectory data mining and visualization: Taxi trajectory data has been used in mining population behavior and traffic patterns [38]. Yuan et al. [37] developed the T-Drive system to recommend optimal driving directions by computing the fastest paths on a time-dependent landmark graph. Their application required fast routing so a small set of landmarks (e.g., frequently visited roads) were used to simplify city networks. Our method has a different goal of visually discovering importance of city streets, so we compute centralities over all streets and aggregate regions through graph partitions. Zheng et al. [39] divided a city into spatial regions and then computed traffic transitions between each pair of regions. Flaw region pairs having heavy traffic beyond the designed capacity were modeled into graphs which were mined for frequent sub-graph patterns to identify flawed city planning. Wang et al. [33] explored traffic data recorded on sparsely distributed cells in a city. Inter-cell correlations were studied in a similar way as a simple graph, which was visualized to display the correlations for visual exploration. In these methods, the graphs were designed for reflecting relations of selected salient regions which might not be geographical neighbors. Users cannot interactively study the important roles of urban roads arbitrarily. In contrast, our graph is constructed to reflect real street connections and their traffic information. It supports exploratory visual analysis on streets. Andrienko et al. [1] transformed GPS-tracked car trajectories into aggregated flows between cellular areas to depict important moving patterns over a city. They grouped points of trajectories into clusters enclosed by convex polygons (e.g., Voronoi) to find salient areas. The connection between these areas formed a graph where the traffic flow over their links was visualized. Our method is different by constructing graphs of streets, instead of clustered areas from GPS points, and applying centrality computation over the graphs.

A large number of approaches have been proposed to visually explore movement data [2]. Many of them are focused on trajectory origins and destinations (i.e., OD data), such as Flowstrates [3], OD maps [36], visual queries [12], and visual analysis of human mobility [18]. Other work visualizes trajectories using various visual metaphors (e.g., lines/curves, heatmaps, and time rings) and interactions, such as TripVista [14], FromDaDy [17], vessel movement [35], route diversity [21], taxi topics [7], and more [9, 13, 30, 32, 31].

Graph centrality and visualization: For studying social networks, Perer and Shneiderman [24] applied centralities to enable users to rank nodes using ordered lists and highlight important nodes in node-link diagrams. Their work showed how the centralities can help analyzers understand complex networks more effectively. It inspired our work where centralities are used to help domain users explore urban transportation data.

3 Requirement Analysis and TrajGraph Overview

In this paper, we use the taxi trajectory data of Shenzhen, China as an example dataset. Shenzhen is a big city in southern China. It has about fifteen million residents in a condensed area. Taxis are a major instrument of resident transportation. The trajectory data was acquired from 15,206 taxis. Each day a taxi trajectory records around 3k sample points with time, GPS location, and speed at an interval of 20 seconds. There are nearly 60 million sample points for each day. Our aim is to use this data for visual analytics of the traffic-related roles of different city streets, in order to provide support for city planners and transportation analyzers. This aim involves several design requirements for a visual analytics approach:

- *R1: Traffic information representation*
  The approach needs to model real traffic information over the city network from the trajectories. The importance of city roads, varying at different time periods, should be computed at user desired spatial and temporal scales. Meanwhile, the connectivity of the roads needs to be maintained correctly for interactive user analyses over geographical spaces;
- *R2: System scalability*
  The approach needs to handle a large number of city streets and massive trajectory data. The raw data should be optimally aggregated so that (1) the visual representations are convenient for domain users

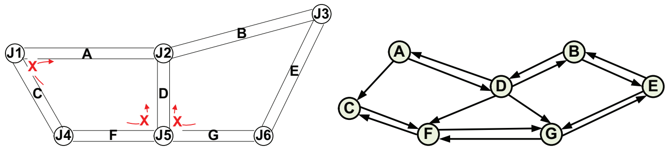
to perform visual reasoning and exploration without overwhelming amounts of data and cluttering, and (2) the computation speed is fast enough to enable interactive visual analysis;

- *R3: Meaningful and intuitive visualization*
  The importance computed by the approach should carry clear meaning in urban transportation to domain users. The visualization should be intuitive so that the users can easily identify interesting patterns;
- *R4: Interactive visual analytics capability*
  Users should be allowed to interactively explore the importance and traffic information in the whole urban domain or in an area of interest within different time periods. They should be able to study different types of traffic-related roles within the same visual analytics framework. In addition, users should be facilitated to compare multiple regions and investigate the time varying patterns when conducting analysis tasks.

To address these needs, we propose TrajGraph, a visual analytics approach that integrates graph based traffic data modeling, automatic graph analyses, coordinated graph, map, and temporal data visualizations, as well as interactions. It provides the following solutions to the aforementioned requirements:

- *R1:* It constructs a special graph directly from massive taxi trajectories to represent not only the complex and unstructured urban network, but also the real traffic information over the network. For example, the raw GPS samples (60 million points) in Shenzhen is converted into a graph of 37k vertices and 1,500k edges.
- *R2:* It applies graph partitioning to create a region-level graph with a small number of vertices and edges. Through this aggregation, users can interactively explore the city-wide traffic information at desired region levels instead of at the street level, which has too many vertices and edges to examine. For example, by creating a graph of 100 vertices and 628 edges from the aforementioned raw graph, users are enabled to interactively and visually examine 100 regions instead of 37k streets. The levels of aggregation in space and time can be flexibly controlled by users. Users can select and compare multiple regions of interest. Users can further study the regions at the street level for more details.
- *R3:* It computes the graph centralities, Pagerank and betweenness, from different types of traffic information. They reveal meaningful roles (Table 1) of city streets, such as hubs and backbones, at a given time period. The centralities are visualized by salient visual metaphors so that users can immediately identify and compare the roles. More graph measures can be further examined to reflect domain knowledge without changing the visual analytics platform.
- *R4:* The regions, streets, and their centralities and traffic information can be interactively explored in a map view. A node-link graph view is provided as an interactive control panel for the map view, from which users can visually examine the centralities and select regions of interest with ease. Users can also examine the temporal changes of centralities by line charts on a temporal information view. The three components address the requirements of user exploration. Meanwhile, users can select an area to compute a local graph and analyze the roles of streets in this area. Moreover, users can interactively change the number of partitions and adjust parameters in graph construction and partition generation to achieve finer or coarser granularities in both time and space.

A fully working prototype of TrajGraph has been implemented. With this prototype, domain users such as city planners and traffic analyzers can investigate a city in a global view to find interesting regions. They can then interactively examine the traffic patterns in more details over multiple views. This enables the preferred multiscale information seeking mantra [29]: "overview first, zoom/filter, details on demand" to facilitate visual exploration in urban transportation. To our best knowledge, there is no similar system that can provide such interactive visual analysis capabilities by utilizing taxi data on discovering the transportation roles of city networks.



(a) A street network (b) A corresponding graph

Fig. 1. *Using graph to represent a street network in TrajGraph.*

4 TrajGraph Model

4.1 Transforming Trajectories to a Graph

TrajGraph constructs a graph, G_T, to represent a road network where taxis travel on in a given time period T. A vertex in G_T represents a road segment in a city. If a taxi travels from road segment A to its connected neighbor B, a directed edge $\overline{AB}$ is added to G_T. G_T can thus represent a road network from a large set of taxis traversing city streets. Those streets where no taxis travel on are not considered significant for the study of street importance. In addition, multiple graphs created for different time periods can represent the varying traffic information. This approach enables users to analyze the temporal changes of transportation functions over multiple G_Ts such as by computing G_T for every two hours.

TrajGraph uses a *dual graph* representation which maps streets to vertices and their connections to edges. Fig. 1b is an example graph with vertices (A to G). It represents a street network shown in Fig. 1a with six junctions (J1 to J6). When taxis travel over the streets, several turns over the junctions are disallowed (shown in red arrows), such as from C to A. Thus the directed edge $\overline{CA}$ is not included in the graph while $\overline{AC}$ is. Consequently, this graph representation makes it easy to reflect road network complexity and real traffic constraints. In contrast, using a *primal graph* mapping streets to edges and intersections to vertices, the constraints need to be modeled by introducing extra nodes and computation to the graph [22].

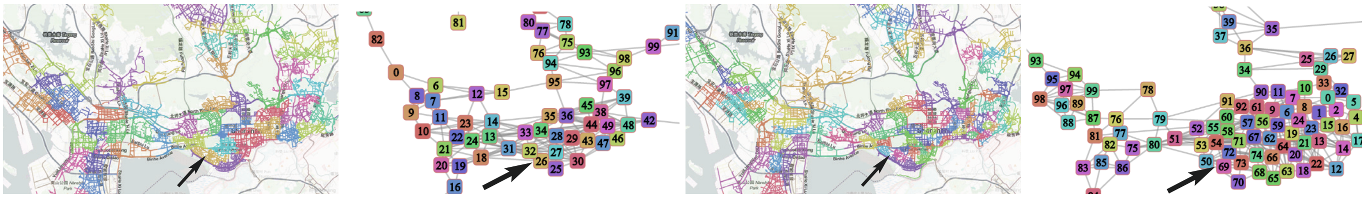
4.1.1 Graph Generation

In generating G_T, the input is the GPS sample points of all taxi trajectories during T. The output is G_T whose vertices represent street segments and whose edges represent the linkage between two physically connected segments (e.g., A and B). Here, the edge is added when there exists at least one taxi travels from A to B. The computational steps are:

1. for each trajectory T_j do
2. for each GPS sample point Pt of T_j in T do
   {
   read next sample point on T_j
3. nextPt = Pt.nextPoint(); if nextPt is null, go to 1;
4. cID = Pt.roadID; nID = nextPt.roadID;
5. if (cID == nID), go to 2;
   remove incorrect consecutive samples
6. if (Distance(Pt, nextPt)$>\xi_d$), go to 2;
7. if (TimeDifference(Pt, nextPt)$>\xi_t$), go to 2;
   update graph
8. add a new vertex N_{cID} to G_T when it is not in G_T;
9. add a new vertex N_{nID} to G_T when it is not in G_T;
10. add a directed edge $E_{cID,nID}$ to G_T when it is not in G_T;
11. update the weight of N_{cID} with proper methods (Sec. 4.1.2);
   }

The weight of vertices is updated in Step 11 according to different graph types which are discussed in Sec. 4.1.2.

In the raw taxi trajectory dataset, information would be wrongly recorded if the device was misoperated or missed signals. This can lead to erroneous time stamps and locations of sample points. We need to keep only valid consecutive sample points on a trajectory so that the graph can correctly reflect connectivity between neighboring road segments. For example, if a taxi travels from street A to street



(a) (b) (c) (d)

Fig. 2. *Creating region level graph of ShenZhen by graph partitioning: (a)(b) without traffic information; (c)(d) with traffic information. More nodes (i.e., regions) are generated in the arrowed downtown areas in (c)(d) than in (a)(b). Colors are selected to show different regions on the map.*

B then to Street C, ideally graph edges $E_{A,B}$ and $E_{B,C}$ should be created. However, the GPS sample on B may be missing, so that $E_{A,C}$ is connected although they are not physically linked. Here we need to avoid connecting $E_{A,C}$ in G_T. Two thresholds, time difference ξ_t and spatial distance ξ_d, are introduced in Step 6 and 7, respectively. We have tested different values of ξ_t and ξ_d. If they are set too small, neighboring road segments are not connected by edges, while using larger values of them will connect many faraway road segments. In our experiments, we found setting $\xi_t = 90$ seconds (the normal device sampling interval is 20 seconds) and $\xi_d = 900$ meters (roughly the distance a car travels with a low city speed 36km/h in 90 seconds) can create satisfying noise removal results. These values are then used in the examples of this paper.

4.1.2 Graph Types

Different types of graphs can be created by defining different vertex weights in Step 11 of the aforementioned algorithm. They include but are not limited to:

- **VW1**: Define vertex weight as the length of streets;
  This setting is used when streets' lengths are to be used in analysis. G_T then represents the urban network's original geometric structure.
- **VW2**: Define vertex weight as the number of taxis;
  The weight of a vertex reflects the number of taxis passing this street in both directions during T. G_T stores the taxi flow information of all streets.
- **VW3**: Define vertex weight as the average travel time;
  Average travel time on a street can be calculated using the street length and the average travel speed during T. G_T stores moving time information over streets during T.
- **VW4**: Define vertex weight as the average speed;
  Average speed computed from taxis reflects the traffic situation of a street during T. G_T then stores traffic movement information during T.

Applying graph algorithms using different settings can identify different transportation patterns (See Sec. 4.3). We use 1 as a uniform edge weight, which gives clear meaning of graph centralities than using both vertex and edge weights.

4.2 Creating a Multilevel Graph

Generating a graph by assigning each street segment to a vertex leads to a large graph. This is because the city scale is large and a long road is divided into multiple segments to be accurately identified in GIS systems. For example, Shenzhen's street-level graph has nearly 37,634 vertices and 1,512,691 edges. Although this scale is not considered large in graph data mining, it is not tractable in a visual analytical system to compute betweenness and closeness centralities, which requires computing the shortest path between each pair of vertices. Moreover, it is not feasible for users to explore the importance of such a large number of streets in a map view or a graph view due to the clutter problem.

In an effective reasoning process, users typically study multiple regions of a city for their roles and performance in transportation. Then they explore regions of interest to study detail information of their local streets. To support such a multilevel visual analytics process, we generate a multilevel graph. A street-level graph can be simplified by aggregating vertices into groups and replacing each group as a new upper level vertex. The edges between the groups are combined following specific rules. Therefore, a region-level graph is created, where a "region" is defined by a connected cluster of street segments. This is different from a spatial region in the city. The region in our graph is computed naturally from trajectory data and reflects traffic patterns over streets. This simplification process can be repeated to construct a hierarchical graph if necessary. In particular, a graph partitioning algorithm is applied to implement the process.

4.2.1 Graph Partitioning

Graph partitioning is used to divide a graph into several chunks while satisfying certain constraints and objectives [19]. The most common constraint is to produce partitions having similar chunk sizes, while the most common objective is to minimize the number of edges between the divided chunks. We use a multilevel k-way approach in the well-known METIS partition [19]. Here k is a given number of partitions to achieve. For TrajGraph, we set the aim to equalizing the sums of vertex weights among all chunks, while minimizing the sum of edge weights between these chunks.

Fig. 2 illustrates two different ways to partitioning a street-level graph of Shenzhen, where the colors are selected to show different regions on the map. The first one does not include the real traffic information sampled by taxis, and the second one introduces such real traffic information. For the first one, a region-level graph is created by setting $k = 100$ while using **VW1**. Fig. 2a displays the regions (i.e., aggregated streets) in different colors over the city map. Fig. 2b is the corresponding node-link graph view. Here the sizes of the regions are optimized by equalizing the sum of street lengths in each region. In general a node covers a larger area in suburbs than a node in downtown areas. Therefore, the graph shows the structural features of the city due to the geographical street distribution. However, it does not consider traffic information. Many small streets in a condensed downtown area suffering heavy traffic flows may be grouped into one region node. In such a case, it is preferred that they can be further studied in finer scales, i.e., by dividing them into more regions.

Therefore, we further apply the partitioning in the second way where **VW2** is used in weights. **VW2** introduces the traffic flow information and then the created regions cover a large area with calm traffic. In downtown areas, heavily used streets are grouped into fine regions. Fig. 2c-d display the aggregated streets and the node-link graph generated using this method. Compared to Fig. 2a-b, the new results create more nodes (i.e., regions) in the downtown areas of Shenzhen. For example, in the arrowed area there are more refined regions. Therefore, users can study these regions in more detail.

k is a heuristic number which users can choose to study the city transportation in different granularities. A very large k may introduce too many small regions that are hard for interactive visual exploration, while a very small k cannot reflect enough details of regional traffic information. Our system allows users to change k on the fly. In the following sections, we use a medium number $k = 100$ in the discussion of centrality computing and visual exploration.

4.2.2 Partition Refinement

The quality of graph partitioning results is highly affected by noise in taxi trajectory data. We have discussed the graph generation method which handles sample errors (Sec. 4.1.1). However, after graph generation there may still exist incorrect edges between road segments

(a) Original partition (b) Refined partition

Fig. 3. *Partition refinement to overcome problems caused by erroneous data.*

Table 1. *Using centralities on different graph types.*

Type	Reflected feature	High Pagerank refers to	High betweenness refers to
VW1	road structure and topology	hubs from road structures	backbones from road structures
VW2	real traffic flow	hubs used by taxis	roads less used by taxis
VW3	real travel time	congestion roads	backbones on fast paths
VW4	real vehicle speed	fluent traffic roads	backbones on congested paths

that are not physical neighbors. For example, Fig. 3a shows one aggregated region of streets with separated segments. We propose a refinement approach on the partitions to overcome the problem. DBSCAN is a popular clustering method based on density reachability [11]. We utilize it on each partition to remove noisy segments. Starting from a center street segment in a partition, DBSCAN performs a ε-neighborhood search, where a given minimum number n is used to imply a dense region of segments. DBSCAN returns density-reachable clusters of the street segments where the largest one is used as the refined result to replace the original partition. The leftover segments are discarded as they represent isolated small segments from inaccurate raw data. Fig. 3b shows the refined partition. The DBSCAN parameters can be adjusted for different levels of clustering. A smaller n or larger ε may not effectively remove the separated segments, while a larger n or smaller ε can lead to the discard of too many street segments. Here we set $\varepsilon = \xi_d = 900$ meters which is consistent with the threshold used in graph generation. We also set $n = 5$ which can provide a good refinement result in our tests.

4.3 Revealing Transportation Patterns by Graph Centralities

To investigate the roles of streets in transportation, centralities are calculated for graph vertices. In particular, we compute Pagerank [20] and betweenness, which are widely used indicators in graph analysis, to characterize *hub* and *backbone* streets/regions. Pagerank originally determines the importance of a web page in internet. Using it in TrajGraph, the importance of a street segment is scored according to the concept that links to high scoring street segments increase the score more than links to low scoring street segments. The streets with high Pagerank scores are preferred hub streets. Betweenness centrality defines that a vertex is important if it lies on many of the shortest paths between two vertices. It can measure whether a street/region is a backbone in the urban network.

Fig. 4 shows two examples of how the centralities reveal transportation patterns of Shenzhen. The region-level graph with 100 vertices, G_{1D}, is created from one day's trajectory data. Fig. 4a shows several regions in Shenzhen's downtown area. They are colored by the Pagerank score of each region vertex. Here the vertex weight is set by **VW2** which is the amount of taxis passing a region. Red color is used for high Pagerank scores and blue is used for low scores, while yellow and green are in between, according to a divergence color spectrum from ColorBrewer [16]. A high score infers a hub region that is highly used by taxis to reach other locations. Fig. 4a reveals three regions with high scores: A (including Shenzhen Library), B (including a very large Caitian shopping center), and C (including the famous Huaqiang North Commercial District). They are the most prominent city centers of Shenzhen. Interestingly, another region (D), which is partly surrounded by them, does not have a comparable high Pagerank. D includes the major intersections connecting to highways in the north, but it is not as popular as A,B,C to be used by taxis. In Fig. 4b, we compute the betweenness with **VW3**, where the average travel time over streets is used to compute the shortest paths. Therefore, the betweenness of G_{1D} quantifies to what extent a region acting as a necessary "backbone" on the *fastest* pathes among regions. A high score reflects that a region is highly preferred to be used as a fast access path to many city areas. It is not surprising that another region (E) has the largest betweenness in Fig. 4b, since it includes a major highway from the downtown area to many northern suburban areas through a tunnel. A, B, and C do not have very high betweenness values since taxis can travel to other regions easily in the downtown area without using them.

In TrajGraph, the default setting is using **VW2** for Pagerank and **VW3** for betweenness, which generate the most frequently used and meaningful importance for city roads. We use this setting in the examples of the following sections. Indeed, users can also use the centralities over different graph weight types which can manifest various transportation features, as illustrated in Table 1. Moreover, other graph measures such as closeness can also be applied. In general, these parameters can be steered to support different analysis tasks without much change to the visual analytics framework.

5 Visualization

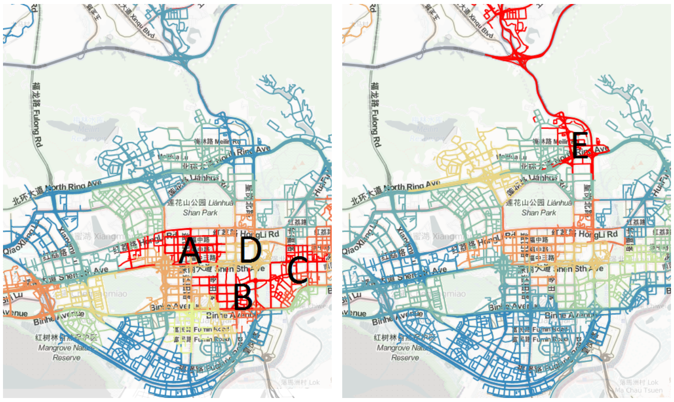
5.1 Visualization Interface

An interactive visualization interface is provided in our prototype so that domain users can conduct exploratory analysis. Fig. 5 shows the three coordinated views of the interface: (1)(2) the node-link graph view; (3)(5) the temporal information view; and (4)(6) the map view. When users click the mouse on a region in one of the first two views, the visual representations of this region in all the views will be highlighted. A set of buttons and list boxes are provided for the selection of time periods, centralities, and traffic information (e.g., travel speed or time). Users can load files that represent different numbers of partitions, which can be pre-generated or created on the fly during analysis.

Node-link graph view: To allow users to explore important regions effectively, a node-link diagram (Fig. 5(1)(2)) visualizes the region-level graph with centrality or traffic information. The position of vertices is the geographical center of a region, so the view reflects the geographical distribution of the urban network. Therefore, users can intuitively associate the vertices in this view with the regions on the map. The graph is a directed graph. In order to reduce clutter, only one edge between neighboring nodes is shown to reflect their linkage. The IDs of the region vertices are shown as the labels. To reduce clutter in dense downtown areas, a force-directed method is applied to avoid node overlapping while keeping the relative positions of the vertices as much as possible.

The normalized scores of the centrality under investigation are visually presented to users by the colors of the vertices. The same divergence color spectrum from ColorBrewer [16] used in Fig. 4 is applied, where vertices with high centralities are shown in red to yellow and vertices with low centralities are shown in green to blue. This color scale is chosen as the default color scheme since vertices with high centralities and low centralities stand out with it, which are often the most interesting regions in an urban network. Other color scales, such as monochrome schemes and other divergence spectrums, are also provided in the prototype and users can interactively change the color scheme according to their tasks and preference.

A key feature of our visual analytics system is that the graph view serves as a control panel for the other views. Users can browse the graph view for vertices with interesting centralities and click (select) them to investigate the regions in the map view. They can also compare the temporal patterns of selected regions in the temporal information view. Zooming and panning are provided in this view for interactive



(a) Pagerank using **VW2** (b) Betweenness using **VW3**

Fig. 4. *Urban network centralities shown on a part of ShenZhen, China.*

exploration. A user study (Sec. 6.3) showed that this view promotes a more effective visual analytics process than only using the map view.

Temporal information view: Once vertices are selected from the node-link graph view, the temporal information of the selected regions is visualized in the temporal information view (Fig. 5(3)(5)). A line chart is used to display temporal changes of the centrality scores and general traffic information (average speed, average travel time, and traffic flow). It is effective for comparing the changing values of one region or among multiple regions. The polylines representing the regions have the same color as the corresponding vertices. The detail information is also visualized in a text box.

Map view: The map view is the major instrument to visualize roads and regions with traffic information and centralities (Fig. 5(4)(6)). Users can choose from a variety of background maps such as the topographical, satellite, and transportation maps, to provide visual cues and geographical context. Corresponding to the selection in the node-link graph view, the streets in the selected regions are drawn over the background map. They can be colored to represent the centralities or to show traffic information such as average speed, average travel time, and flow. Moreover, a *rose chart* can be turned on to visualize the temporal changes of the information on a region. Using each arc of the rose chart to represent a two-hour window, one day is shown in 12 arcs in the circle. The chart helps users examine the time varying information and find a time period of interest to conduct further study.

5.2 Multiscale Visualization

The transportation patterns can be examined at different levels of detail in two ways. First, starting from a city-wide graph where the vertices represent regions, users can select a region from the node-link view to examine its details. In the map view, users can study the centralities and traffic information at the street level, which is shown in Fig. 5(6). This visual analytics process is further discussed in Sec. 8. Second, users can select an arbitrary area on the map (currently rectangle selection is supported). Then a local graph is constructed from the trajectories, which only includes the streets inside the selected area. Users can then perform visual analyses to investigate the centralities of these streets, which tell the local importance of these streets in the area. Note that the centralities are not the same as those computed from the city-wide graph. For example, a street in the area may have a small betweenness in the whole city, but locally, it has a high betweenness indicating that it is a backbone street inside the area. Sec. 5.4 provides an example of exploring a local area. These functions help users flexibly conduct visual exploration tasks.

5.3 Example: Investigating Time Varying Centralities

It is of great importance to compare and analyze the temporal differences and changes of centralities in a region and among different regions. The centralities can be computed for a small (e.g., two hours) time window. Then, each day a region has 12 Pagerank or betweenness scores. Fig. 6b-c shows the temporal changes of the centralities of three neighboring regions, ID90, ID91, and ID92 in Fig. 6a, where the *rose charts* show their Pagerank values at different time periods. It demonstrates that ID91 has the highest Pagerank and the lowest betweenness most of the time. It indicates that ID91 is a popular region visited very often by taxi drivers, but it is not a backbone region used to reach many other regions. ID90 generally has a higher betweenness that is drastically changing over time. A very high betweenness value at 16:00 shows the region is used as a fast bypass route (to reach north suburbs through the tunnel). But its score drops at 18:00 which means it loses its function. A possible reason may be that the experienced taxi drivers avoid using the streets in this region during the rush hours. That may also be the reason of why both betweenness and Pagerank of ID92 increase at 18:00. This region neighboring to ID90 is used as an alternative route to avoid ID90. The size of the time window can be changed to a longer (half day) or shorter (one hour or half hour) period for different analysis granularities.

5.4 Example: Exploring Centralities in a Local Area

In this example, users select an area of interest and analyze the centralities computed from the corresponding local graph (Fig. 7a). Fig. 7b-c show the streets colored by their Pagerank and betweenness, respectively. Moreover, Fig. 7d displays the streets colored by the density of taxi flows where red refers to highly occupied streets. A (i.e., Node 6) includes the most important roads of this local area since it has high scores in both Pagerank and betweenness. It shows that this road is a fast path and used by many drivers to reach other streets. In comparison, B has a high Pagerank but its betweenness is low. It indicates B is used by many taxis but it is not the fastest path possibly due to traffic jams. This is confirmed by Fig. 7d which reveals that B has a large traffic flow.

6 Evaluation

6.1 Analytical Utility: Study with Drivers of Shenzhen

We conducted a study with a group of drivers of Shenzhen to evaluate the assumption in our approach that Pagerank and betweenness centralities reflect the importance of city streets. Drivers have the most direct observations of real traffic situations. Their opinions are close to the ground truth which is hard to be discovered using other means. We recruited 20 volunteer residents of Shenzhen with different backgrounds and occupations, most of whom drive their cars everyday in Shenzhen. Their driving experience ranged from 1 to 15 years. Fifteen of them claimed that they were *very familiar* with Shenzhen's roads and another five claimed *familiar*.

We conducted the study with the participants one by one using an image based comparison approach. Fig. 8(a) shows an example image where two regions, A and B, in Shenzhen are circled on the map in red and blue. They are neighboring regions and connected vertices in our graph. During 6-8am A had higher Pagerank than B. In the study each participant was asked to look at the image and answer "*Which region is more like a hub region during 6-8am? A hub region is used more often by drivers to communicate between different city areas.*". The participant selected A or B to answer the question. Two pairs of such regions during 6-8am and two pairs of regions during 4-6pm were chosen, so that 4 Pagerank questions were answered by each participant. In a similar process, two pairs of regions during 6-8am and two pairs during 4-6pm were used to create 4 betweenness questions. Each participant was asked to choose "*Which region is more like a backbone region during 6-8am? A backbone region contains important roads which may cause big traffic problem if broken.*". Each participant answered 8 questions in total. The 8 pairs of regions were selected randomly from Shenzhen's graph by a nonresident person.

We compared the answers of all questions to the computed centralities. For example, if a participant selected A to answer the Pagerank question for Fig. 8(a), then we said the answer agreed with our Pagerank, otherwise it was a miss. The result is shown in Fig. 8(b). The total agreement rate of all questions was 63.75% (66.25% for Pagerank and 61.25% for betweenness). Fig. 8(b) also shows the rates for the drivers with different lengths of driving experience (1-6 years and 7+ years). There were no significant differences between the groups.

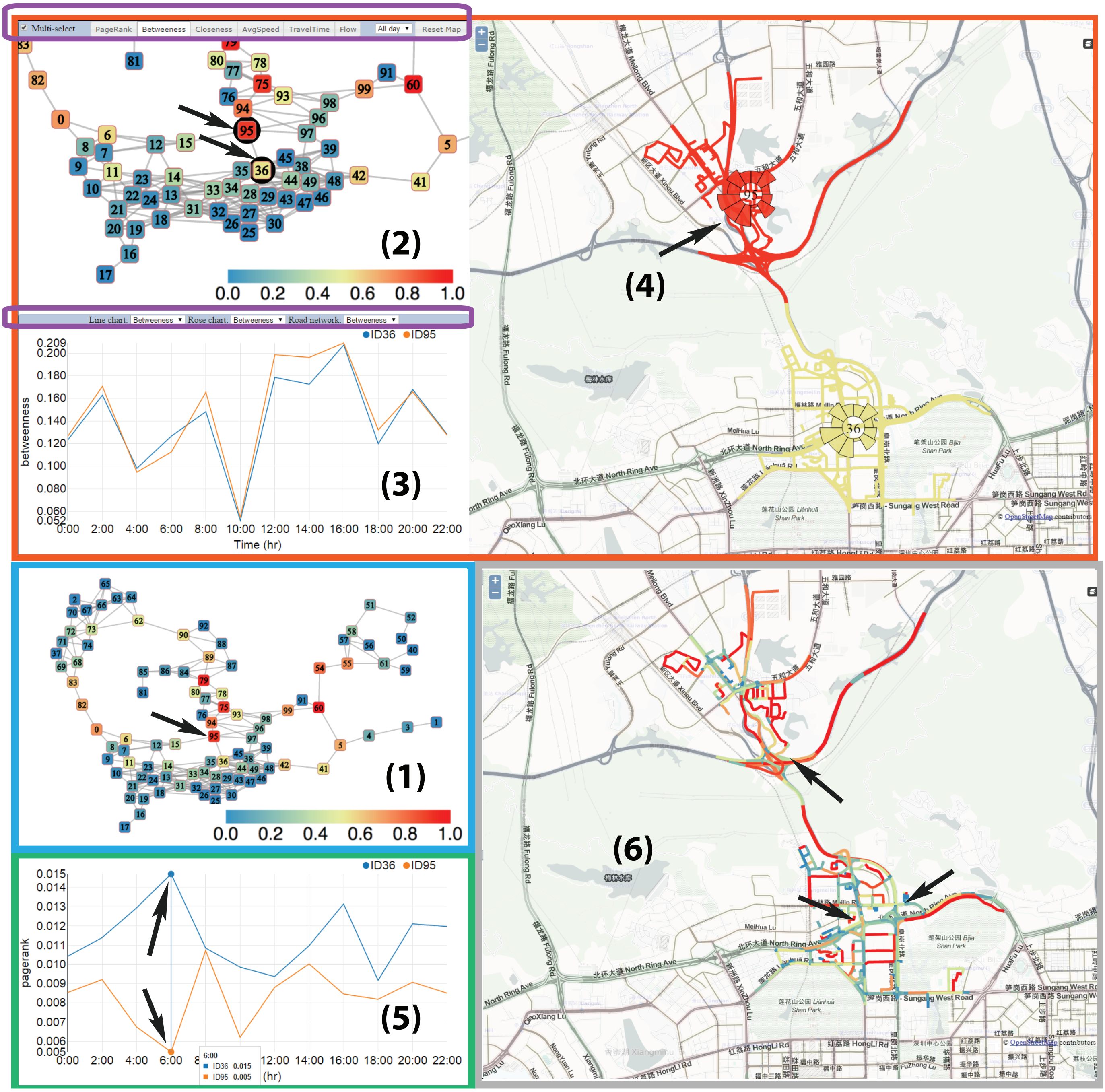
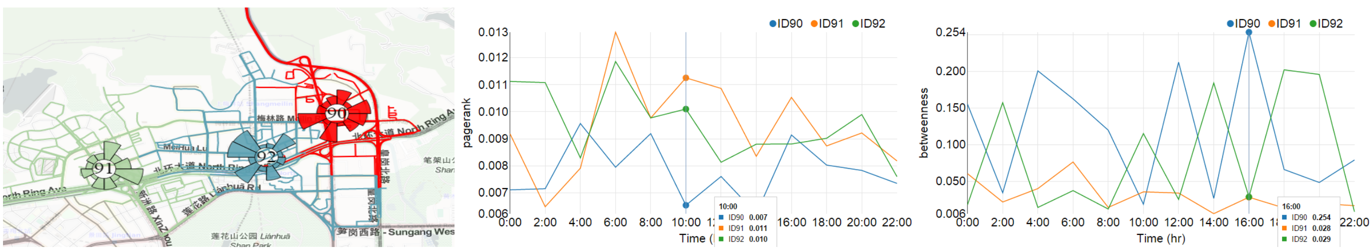
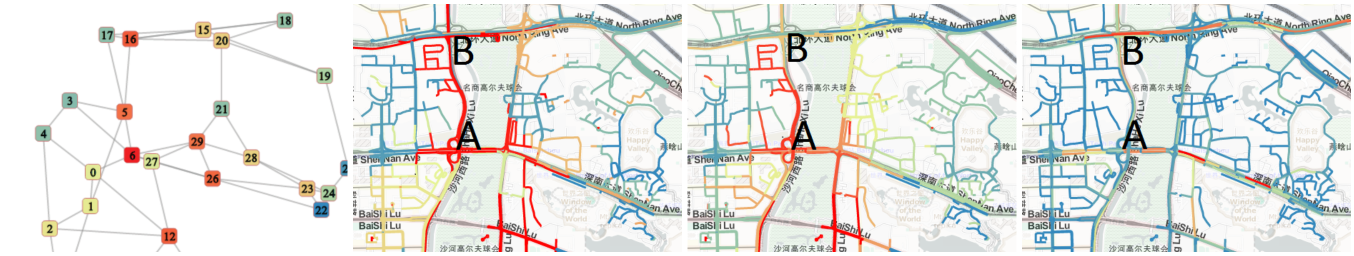


Fig. 5. *A case study (see details in Sec. 8). The interface includes three coordinated views: (1)(2) a node-link graph view where users can zoom and select interesting nodes; (3)(5) a temporal information view where users can examine the temporal changes of centralities and traffic information of selected regions; and (4)(6) a map view with rose charts where users can study centralities and traffic information on the city map. A set of buttons and list boxes (shown in purple rectangles) are provided for the selection of time period, centralities, and traffic information.*



(a) Three neighboring regions (b) Temporal Pagerank (c) Temporal betweenness

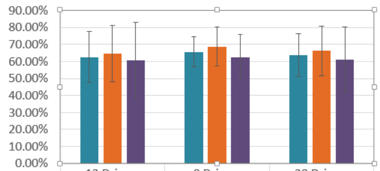
Fig. 6. *Temporal changes of centralities of three regions.*



(a) A local graph (b) Pagerank of (a) (c) Betweeneness of (a) (d) Taxi flow of (a)

Fig. 7. *Computing and visualizing temporal centralities of a local selected region of Shenzhen.*

(a)



	12 Drivers 7+ Years	8 Drivers 1~6 Years	20 Drivers All Drivers
Total	62.50%	65.63%	63.75%
Pagerank	64.58%	68.75%	66.25%
Betweenness	60.42%	62.50%	61.25%

(b)

Fig. 8. *Study with Shenzhen drivers. (a) Two neighboring regions with Pagerank A > Pagerank B. (b) The agreement rate of our computation with the subjective evaluation of the drivers.*

In this study, the agreement of our computation with the drivers' *subjective evaluation* of the city's traffic features was more than 60%. It indicated that the metrics of Pagerank and betweenness can provide meaningful knowledge of urban transportation features. The rates were not extremely high possibly because the drivers may not be familiar with all regions of the city. Feedback from the drivers was that they sometimes were confused about why some streets form one region while others form another. This is because the automatically generated regions may not match people's mental image of city networks and functional sections. This question deserves future work together with urban transportation researchers.

6.2 Domain Expert Testimonial Feedback

To learn whether TrajGraph is valuable to domain users, we interviewed an urban geographer with a PhD degree, who is a former urban transportation planner. We sent him a written document of our approach and he read it. We then met him in person and discussed the model, the centralities, and the visualization process. We installed our prototype on his personal computer, and taught him how to use it. He spent two hours using the system and provided us a written document of his feedback.

First, he evaluated our approach of studying city structures with trajectories: "Policy-makers and transportation scientists realize that many cities have been characterized by a polycentric urban development model and the associated traffic jam. However, very few tools exist for visual inspection of the pulse of a city at various spatial and temporal scales. Taxi trajectory data involves interdependent entities connected and intersected to form network topologies. Urban traffic flow can be viewed as transportation demands aggregately distributed in street networks. Hierarchy is an important property of street network which suggests that only a small number of streets are prominent. In other words, a small percent of top streets accommodate about a majority percent of traffic flow. However, it is a data-intensive computation to identify such structure and importance. At the same time, the visualization component is usually abandoned due to the data size burn. This new tool addresses an emerging need to provide non-technical users to evaluate city's hierarchy and traffic patterns."

Then he stated "This platform develops a powerful visualization tool to address such complexity. Using taxi trajectories helps understand the urban spatial structure and can shed lights on the mechanism of urban development. The interface is easy to navigate over a map to find city-wide critical areas. This tool allows for understanding both urban structures and human behavior in shaping urban transportation demand using a graph-based approach. The tools may need to be further adjusted in collaboration with real users."

He pointed out that "The regions in this system currently are purely from graph computing which may need to consider the functional and traditional definition of city areas. " He also suggested "The design characteristics of streets (such as number of lanes, length, and speed limit) can also be incorporated, as well as other city data." We will improve our approach according to these comments.

6.3 User Study of Node-Link Graph View

Since map views and temporal information views have been well studied in traffic visual analytics, we focused on evaluating the usefulness of the node-link graph view in the visual analytics process. Our hypothesis was that the high level visual abstraction and interactions provided by the node-link graph view can help users examine centralities more effectively. To test this hypothesis, we conducted a preliminary user study by using (V1) the map view coordinated with the node-link graph view; (V2) the map view only without the node-link graph view.

Participants, Tasks and Procedure: Fourteen participants (4 females and 10 males) aged from 22 to 35 were graduate students majoring in computer science. Four tasks were used in the study: (T1) find five regions with the highest Pagerank with V1; (T2) find five regions with the highest Pagerank with V2; (T3) find five regions with the highest betweenness with V1; (T4) find five regions with the highest betweenness with V2. Two different graphs (PK1 and PK2) were used for T1 and T2, and another two (BW1 and BW2) were used for T3 and T4. The study began with an introductory briefing of the concepts of Pagerank and betweenness to a participant. Then both the map view and the graph view were introduced to the participant and she/he explored the views freely for a few minutes. The participant then conducted the tasks one by one in a random order and was given 90 seconds for each task. PK1 and PK2 were randomly assigned for T1 and T2, and BW1 and BW2 were randomly assigned for T3 and T4.

Result: The answers of each participant were recorded and compared with the ground truth to compute the accuracy. For all participants, more correct regions were found with V1 (average accuracy 90% for T1 and 88% for T3) than with V2 (average accuracy 75% for T2 and 74% for T4). We then performed a statistical test on the result. First, we did an F-Test to determine if the variances of accuracy when using the two visual interfaces (V1 and V2) were equal. We got 1.916(F) < 2.576 (F critical one-tail), p-value=0.126, so we did not reject the null hypothesis. The variances of accuracy using the two interfaces were equal. Then, we did a two-tail test to test if the means of using the two visual interfaces (V1 and V2) are equal. We got 4.84(t stat) > 2.06(t critical two-tail), p-value=6.3e-5, and therefore, we rejected the null hypothesis. We concluded that the means of accuracy when using V1 and V2 differed significantly. This study showed that the graph view can help users examine centralities more effectively than only using the map view.

Table 2. *Size of the street-level graph and region-level graphs.*

	Origin Graph	3000 partitions	1000 partitions	100 partitions
Vertex No.	37,634	3,000	1,000	100
Edge No.	1,512,691	32,516	10,214	628

Table 3. *Computing time for the region-level graphs*

Graphs	Pagerank	Betweenness	Closeness
original graph	23.7sec	8.2hr	2.3hr
100 partitions	0.03sec	0.02sec	0.08sec
1000 partitions	0.3sec	2.0sec	0.5sec
3000 partitions	1.1sec	15.4sec	4.3sec

7 Implementation and Computational Feasibility

In our implementation, Pagerank is computed following the widely used approach [20]. The computation of betweenness is executed by a fast algorithm [4]. We report the computation performance on a desktop workstation (Intel Xeon E5520 with 4 cores at 2.27GHz and 16GB memory). Created from the raw data of 15,206 taxi trajectories (around 60 million sample points) with T as one day, the original street-level graph is created in 35.3 seconds. The partitions are computed very fast in less than 1 second. Table 2 shows the size of the original graph and a few region-level graphs. Table 3 shows the computation time of the centralities. The computation is slow on the original street-level graph, especially for betweenness and closeness centralities, since they need to compute shortest paths between each pair of vertices, a well-known time-consuming problem. After graph partitioning, the performance is greatly accelerated due to the small size of region-level graphs. The 100-partition graph was computed in milliseconds leading to interactive performance, which is critical for our visual analytics system.

8 Case Study

In this case study, we illustrate how an urban transportation expert, named Zhang, used TrajGraph to find important regions which are critical for Shenzhen's traffic problems. Zhang opened our web-based prototype system in his web browser. In Fig. 5(1), he looked at the city-wide graph in the node-link graph view and selected to show the betweenness centrality of the whole day. He found that node 95 was red, which indicated a backbone region on fast reaching paths in Shenzhen. Zhang zoomed in the graph view (Fig. 5(2)) and clicked node 95 to display the region on the map (Fig. 5(4)).

From the map view he immediately realized that this region was on one side of a mountain separating the city. He had general knowledge of Shenzhen that this mountain was critical for the transportation. It was of interest for him to study more pertinent details. Therefore he started to study the neighboring region (node 36) on the other side of the mountain by clicking the node in Fig. 5(2). What were the different transportation features of these two regions connected by a mountain tunnel? From Fig. 5(3) Zhang found that the two regions had quite similar time variation in their betweenness values during this day. The line charts revealed that both of them were very important regions connecting north and south parts of Shenzhen. Zhang further examined their Pagerank values in the temporal information view. As shown in Fig. 5(5), the two time lines now became different. It was obvious that at 6:00am region 36 had a higher score than region 95. Zhang would like to find the reason. He zoomed into the two regions on the map to study street level details. He associated the colors of the streets to the average travel time on them at 6:00 am. In Fig. 5(6), he identified that in the northern region (node 95), the major highway junction had a long travel time (shown in red on roads), which implied a traffic jam. Therefore, this region was not well used as a hub by taxi vehicles as taxi drivers may avoid this junction. So its Pagerank was low. On the other hand, the southern region (node 36) included two major junctions where the travel time was not red. Therefore, they were used for fast transportation to other areas which answered why the region's Pagerank was higher. Zhang reached a conclusion that region 95 was an important urban bottleneck in the early morning at that day. If this pattern was repeated for multiple days, it might be worth to find a solution in urban design or traffic control.

This case study illustrates how TrajGraph effectively supports interactive visual reasoning of real traffic situations in a large city. Besides intuitively conveying the knowledge derived from automatic computational analysis to users, it provides interactive visual exploration capabilities to allow users to have a comprehensive understanding of the complex situations.

9 Conclusion

We have proposed a visual analytics system, TrajGraph, to study urban transportation patterns from taxi trajectories. TrajGraph uses a new graph model to reflect urban network structures and incorporate real traffic information carried by taxi trajectory data. The graph is aggregated by graph partitioning to represent region-level information. Graph centralities of Pagerank and betweenness are computed to reveal the importance of different city regions in transportation. A set of visualizations and interactions allow users to investigate the importance in different levels of detail. Users can effectively conduct tasks in analyzing the roles of city roads for improving urban planning and amending city operations.

We will work further on the approach by (1) employing more graph measures and algorithms, such as minimum spanning tree and community analysis, and studying their functions in urban studies; (2) promoting its use by domain experts and analyzers. In the long run, we will see more and more such data with the widespread use of recording devices and systems. We hope our method can be further extended to public transits, fleet, and human trajectories.

Acknowledgments

We would like to thank Shuo Zhao for his help in the user study with Shenzhen drivers. We would like to thank Maogong Zheng and George Chen from Shenzhen Institute of Advanced Technologies for providing the data. We also thank the reviewers for their highly constructive feedback and suggestions. This work is partially supported by U.S. National Science Foundation under grant numbers 1535031, 1535081, and 1416509.

References

[1] N. Andrienko and G. Andrienko. Spatial generalization and aggregation of massive movement data. *Visualization and Computer Graphics, IEEE Transactions on*, 17(2):205–219, 2011.

[2] N. Andrienko and G. Andrienko. Visual analytics of movement: An overview of methods, tools and procedures. *Information Visualization*, 12(1):3–24, 2013.

[3] I. Boyandin, E. Bertini, P. Bak, and D. Lalanne. Flowstrates: An approach for visual exploration of temporal origin-destination data. *Computer Graphics Forum*, 30(3):971–980, 2011.

[4] U. Brandes. A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 25(163), 2001.

[5] J. Buhl, J. Gautrais, N. Reeves, R. Solé, S. Valverde, P. Kuntz, and G. Theraulaz. Topological patterns in street networks of self-organized urban settlements. *European Physical Journal B*, 49:513–522, 2006.

[6] Caliper. TransCAD Transportation Planning Software. Website, 2014. http://www.caliper.com/tcovu.htm.

[7] D. Chu, D. A. Sheets, Y. Zhao, Y. Wu, J. Yang, M. Zheng, and G. Chen. Visualizing hidden themes of taxi movement with semantic transformation. In *Pacific Visualization Symposium (PacificVis), 2014 IEEE*, pages 137–144. IEEE, 2014.

[8] Citilab. Cube. Website, 2014. http://www.citilabs.com/.

[9] T. Crnovrsanin, C. Correa, C. W. Muelder, and K.-L. Ma. Proximity-based visualization of movement trace data. In *IEEE VAST*, pages 11–18, October 2009.

[10] P. Crucitti, V. Latora, and S. Porta. Centrality measures in spatial networks of urban streets. *PHYSICAL REVIEW E*, 73(3):036125, 2006.

[11] M. Ester, H. peter Kriegel, J. S, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of KDD)*, pages 226–231, 1996.

[12] N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva. Visual exploration of big spatio-temporal urban data: A study of new york city

taxi trips. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2149–2158, Dec. 2013.

[13] Y. Gao, P. Xu, L. Lu, H. Liu, S. Liu, and H. Qu. Visualization of taxi drivers' income and mobility intelligence. In *ISVC (2)*, pages 275–284, 2012.

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

[15] A. Guttman. R-trees: A dynamic index structure for spatial searching. *SIGMOD Rec.*, 14(2):47–57, June 1984.

[16] M. A. Harrower and C. A. Brewer. ColorBrewer.org: An Online Tool for Selecting Color Schemes for Maps. *The Cartographic Journal*, 40(1):27–37, 2003.

[17] C. Hurter, B. Tissoires, and S. Conversy. Fromdady: Spreading aircraft trajectories across views to support iterative queries. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1017–1024, Nov. 2009.

[18] X. Jiang, C. Zheng, Y. Tian, and R. Liang. Large-scale taxi o/d visual analytics for understanding metropolitan human movement patterns. *Journal of Visualization*, 18(2):185–200, 2015.

[19] G. Karypis and V. Kumar. Multilevel graph partitioning schemes. In *Proc. 24th Intern. Conf. Par. Proc., III*, pages 113–122. http://glaros.dtc.umn.edu/gkhome/metis/metis/overview, 1995.

[20] P. Lawrence, B. Sergey, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford University, 1998.

[21] H. Liu, Y. Gao, L. Lu, S. Liu, L. Ni, and H. Qu. Visual analysis of route diversity. *IEEE Conference on VAST*, pages 171–180, 2011.

[22] D. Papadias, J. Zhang, N. Mamoulis, and Y. Tao. Query processing in spatial network databases. In *Proceedings of the 29th International Conference on Very Large Data Bases - Volume 29*, VLDB '03, pages 802–813. VLDB Endowment, 2003.

[23] P. Pavithra. Network structure and metropolitan mobility. *Journal of Transport and Land Use*, 7(2), 2014.

[24] A. Perer and B. Shneiderman. Balancing systematic and flexible exploration of social networks. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):693–700, Sept. 2006.

[25] M. Picone, S. Busanelli, M. Amoretti, F. Zanichelli, and G. Ferrari. *Advanced Technologies for Intelligent Transportation Systems*. Springer Publishing Company, Incorporated, 2014.

[26] S. Porta, P. Crucitti, and V. Latora. The network analysis of urban streets: A dual approach. *Physica A: Statistical Mechanics and its Applications*, 369(2):853–866, Sept. 2006.

[27] M. Rosvall, A. Trusina, P. Minnhagen, and K. Sneppen. Networks and cities: An information perspective. *Phys. Rev. Lett.*, 94:028701–1:028701–4, 2005.

[28] C. Rozenblat and G. Melancon. *Methods for Multilevel Analysis and Visualisation of Geographical Networks*. Springer, 2013.

[29] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages*, pages 336–, Washington, DC, USA, 1996. IEEE Computer Society.

[30] C. Tominski, H. Schumann, G. Andrienko, and N. Andrienko. Stacking-based visualization of trajectory attribute data. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2565–2574, 2012.

[31] H. Wang, H. Zou, Y. Yue, and Q. Li. Visualizing hot spot analysis result based on mashup. In *Proceedings of the 2009 International Workshop on Location Based Social Networks*, LBSN '09, pages 45–48, New York, NY, USA, 2009. ACM.

[32] Z. Wang, M. Lu, X. Yuan, J. Zhang, and H. van de Wetering. Visual traffic jam analysis based on trajectory data. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2159–2168, 2013.

[33] Z. Wang, T. Ye, M. Lu, X. Yuan, H. Qu, J. Yuan, and Q. Wu. Visual exploration of sparse traffic trajectory data. *IEEE Trans. on Vis. Comp. Graph.*, 20(12):1813 – 1822, 2014.

[34] E. Weiner. *Urban Transportation Planning in the United States: History, Policy, and Practice*. Springer; 4th ed., 2012.

[35] N. Willems, H. van de Wetering, and J. J. van Wijk. Visualization of vessel movements. In *EuroVis*, pages 959–966, Aire-la-Ville, Switzerland, Switzerland, 2009. Eurographics Association.

[36] J. Wood, J. Dykes, and A. Slingsby. Visualisation of origins, destinations and flows with od maps. *The Cartographic Journal*, 47(2):117–129, 2010.

[37] J. Yuan, Y. Zheng, C. Zhang, W. Xie, X. Xie, G. Sun, and Y. Huang. T-drive: Driving directions based on taxi trajectories. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '10, pages 99–108, New York, NY, USA, 2010. ACM.

[38] Y. Zheng, L. Capra, O. Wolfson, and H. Yang. Urban computing: Concepts, methodologies, and applications. *ACM Trans. Intell. Syst. Technol.*, 5(3):38:1–38:55, Sept. 2014.

[39] Y. Zheng, Y. Liu, J. Yuan, and X. Xie. Urban computing with taxicabs. In *Proceedings of the 13th International Conference on Ubiquitous Computing*, UbiComp '11, pages 89–98, New York, NY, USA, 2011. ACM.