

## Research Article

# A Method of Vehicle Route Prediction Based on Social Network Analysis

Ning Ye,<sup>1,2,3</sup> Zhong-qin Wang,<sup>1</sup> Reza Malekian,<sup>4</sup> Ying-ya Zhang,<sup>1</sup> and Ru-chuan Wang<sup>1,2,3</sup>

<sup>1</sup>Institute of Computer Science, Nanjing University of Post and Telecommunications, Nanjing 210000, China

<sup>2</sup>Jiangsu High Technology Research Key Laboratory for Wireless Sensor Network, Nanjing 210003, China

<sup>3</sup>Key Lab of Broadband Wireless Communication and Sensor Network Technology of Ministry of Education, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

<sup>4</sup>Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria 0028, South Africa

Correspondence should be addressed to Reza Malekian; [reza.malekian@up.ac.za](mailto:reza.malekian@up.ac.za)

Received 8 January 2015; Revised 7 April 2015; Accepted 9 April 2015

Academic Editor: Jian-Nong Cao

Copyright © 2015 Ning Ye et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A method of vehicle route prediction based on social network analysis is proposed in this paper. The difference from proposed work is that, according to our collected vehicles' past trips, we build a relationship model between different road segments rather than find the driving regularity of vehicles to predict upcoming routes. In this paper, firstly we depend on graph theory to build an initial road network model and modify related model parameters based on the collected data set. Then we transform the model into a matrix. Secondly, two concepts from social network analysis are introduced to describe the meaning of the matrix and we process it by current software of social network analysis. Thirdly, we design the algorithm of vehicle route prediction based on the above processing results. Finally, we use the leave-one-out approach to verify the efficiency of our algorithm.

## 1. Introduction

Intelligent Transport Systems (ITS) [1, 2] are an important tool in the delivery of sustainable transport policies. They are widely implemented within cities to manage traffic and to influence travel behavior. Vehicular ad hoc networks (VANETs) [3] are a significant component of ITS. In recent years, the intelligent transportation technology based on VANETs has gradually become a research focus in the Internet of things with the rapid development of wireless communication and sensor technology. VANETs [4] are comprised of OBU (On-Broad Unit), RSU (Road Side Unit), TCC (Traffic Control Center), and Internet. VANETs facilitate the vehicle-to-vehicle and vehicle-to-RSU communication, providing new technology support for detection of urban traffic condition.

VANETs enable enhancing safety level and the ability to driver assistance for vehicles. Through wireless communication between the vehicles, VANETs can help drivers acquire

driving information beyond the scope of their vision and perception and then timely handle these potential hazards to avoid traffic accidents. Also, VANETs can depend on road traffic state information to provide real-time traffic guidance services for drivers and then help them to choose a reasonable route to avoid traffic congestion. So, in VANETs, a vehicle can perceive shared information from other vehicles and communicate with surrounding vehicles so as to optimize the drivers' upcoming routes, to shorten driving time to destinations, and to improve driving experience.

Vehicle route prediction is of great application significance in VANETs. It can be used to effectively inform drivers which of the upcoming road segments will be frequently congested and inform them about related business information that drivers will be interested in. For example, a driver can depend on road congestion from VANETs to timely optimize the upcoming route. As we all know, most vehicles are equipped with navigation software to help drivers select a better driving route. However, the software is to find several

routes between given origins and destinations by combining some path algorithms based on historical traffic data, for example, Google Map and Baidu Map, and also lacks a real-time traffic situation. For the same beginning and end point, if one of recommendation routes with smoother road segments has been selected at same time, in this case, the original relatively smooth roads will become congested and the original congested roads will become smooth [5]. Route prediction could solve the problem easily. After the system of VANETs predicts which route a driver will take in the future, it could know the total number of vehicles on one's road segment in a period of time, thus inferring the corresponding road congestion. In summary, the system of VANETs based on route prediction enables a more holistic view of the traffic situation, a greater use of automatic responses during key events, a better understanding of how systems work together and how to resolve problems as they occur, and greater flexibility in terms of mixing and matching solutions.

This paper is organized as follows. The next section describes related work. Section 3 presents how to build a road network model and the corresponding matrix and introduces two concepts of social network analysis into the model. In Section 4, we propose an approach for route prediction based on social network analysis. Section 5 describes the origin of experimental data, evaluation metrics, and results. Finally, the paper is concluded in Section 6.

## 2. Related Work

Here we introduce previous work in the field of route prediction. Karbassi and Barth [6] proposed a car-sharing application in the case of given driving start and end points, while our method to predict driving routes in this paper is under the unconscious perception, that is, without inputting the end point, our system can calculate the possible vehicle driving routes automatically according to the current position of a vehicle. Also, some methods of vehicle route prediction mainly depend on historical driving data to acquire possible routes in the future. For example, Krumm [7] predicted a driver's near-term future path by a Markov model. Similarly, Simmons et al. [8] adopted a hidden Markov model to predict destinations and routes and Froehlich and Krumm [9] got route regularity by analyzing vehicle routes collected by 250 drivers and then the closest match algorithm returned an ordered list of route candidates based on route regularity. All of the previous work considers each entire driving route as a data analysis unit and views two interconnected road segments as an item. However, they mainly focus on two interconnected road segments so that it is easy for people to ignore possible relationship of disjunct road segments. For instance, Road 1 is connected to Road 2 while Road 2 is connected to Road 3. So there is a direct relationship between Road 1 and Road 2 and an indirect relationship between Road 1 and Road 3. Here, we assume the urban road network with social characteristics. In view of social network analysis [10], both relationship features are taken into account in this paper to predict a driver's route.

## 3. The Representation of Road Network Based on Social Network Analysis

At first, the road network can be represented as a graph  $G = (V, E)$ , where each road segment is defined as a vertex  $V$  of the graph and the intersection between road segments is defined as an edge  $E$  of the graph [11]. When a vehicle is driven from a starting point A to an end point B, an ordered set of roads in the route is defined as  $R_i = \{r_1, r_2, \dots, r_n\}$ , where  $r_i$  represents the  $i$ th road and  $n$  is total number of just-driven roads. So, in graph theory, a driven route can be represented as  $R_i = (v_1, e_{1,2}, v_2, \dots, v_n, e_{n,n+1}, v_{n+1})$ . In addition, the weight of every directed edge in the graph should be known. The weight is initialized to 1 in the initial road network model because different vehicles will pass through these road segments at least once in real life.

Then we illustrate the meaning of the road network with social characteristics. Social network is a set of social actors and the relationships between them; that is, a social network is composed of social actors and the relationship between actors [12, 13]. We suppose that the set of all of collected vehicle routes is  $R = \{R_1, R_2, \dots\}$ , where all of the elements are different from each other. As mentioned above, we regard the different elements  $v_j$  in the set  $R_i$  of a vehicle route as corresponding social actors in social network. So all of the social actors from different routes are represented as a set  $N = \{v_1, v_2, \dots, v_m\}$ , where elements  $v_1, v_2, \dots, v_m$  are different. Also, while a directed edge  $e_{ij}$  represents the direct relationship  $v_i \rightarrow v_j$  between adjacent points, all pairs of relationships will be formed as a set  $E = \{e_{1,2}, e_{2,3}, \dots, e_{k,k+1}\}$ , where there are  $k$  pairs of relationships in the graph. So we propose the strength of relationship here. If we find that most vehicles once passed through road  $v_i$  to road  $v_j$  from the collected data,  $v_i \rightarrow v_j$  is an undoubtedly strong relationship between them. And the relationship strength can be measured, which can be represented as the number  $n$  of driving through road  $v_i$  to road  $v_j$  in the data set. As a result, on this basis the weight in the graph is corrected to  $n + 1$  (the initial weight is 1).

Next, we can transform the graph of road network into the format of a matrix. Assume that a matrix to represent a network model is expressed in

$$A = \begin{Bmatrix} - & a_{12} & \cdots & a_{1n} \\ a_{21} & - & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & - \end{Bmatrix}, \quad (1)$$

where the elements of every row and column are from the set  $N$  of points (social actors) and the matrix element  $a_{ij}$  represents the relationship strength between roads. Suppose that there are  $n$  driving routes from  $v_i$  to  $v_j$ ; then  $a_{ij} = n + 1$ . Besides, we do not focus on the relationship between actors and themselves, so the diagonal values are denoted as “-” in the matrix.

Finally, we introduce two concepts from social network analysis to help us predict vehicle routes—point centrality of roads and cohesive subgroup of roads.

(1) *Point Centrality of Roads.* If many vehicles have traversed the same road segment, we think the road segment is very important for drivers. In other words, when vehicles are near the important road segment, it is much more likely to pass through the road for drivers. Indeed, the possibility of traversing other roads connected to it will also increase. So we consider that the road has a higher point centrality of roads than other roads. When we study point centrality of one point in the graph, it is not enough to only concern on the point but those points at a distance of 2 from it in a directed graph. So we should evaluate relationships between the point and either adjacent ones or nonadjacent ones.

Here we introduce how to measure point centrality of roads in a graph. Suppose that in the graph  $v_{i-1}$  is a node directly connected to another  $v_i$ ,  $v_{i-2}$  is a node directly connected to node  $v_{i-1}$ , and  $v_{i-2}$  is indirectly connected to  $v_i$ . We also assume that the number of nodes directly connected to  $v_i$  is  $n$  and the number of nodes indirectly connected to  $v_i$  is  $m$ . Then the expression of point centrality of node  $v_i$  is represented in

$$C(v_i) = \sum_{k=1}^n (d_k(v_{i-1}, v_i) + d_k(v_i, v_{i-1})) + \sum_{k=1}^m (d_k(v_{i-2}, v_{i-1}) + d_k(v_{i-1}, v_{i-2})), \quad (2)$$

where  $d_k(v_{i-1}, v_i)$  represents the distance from the  $k$ th node  $v_{i-1}$  to  $v_i$ . For example, in Figure 1 Road 1 is, respectively, directly connected to Road 2 and Road 5 and indirectly connected to Road 3 and Road 4.  $e_{i,j}$  represents the distance from node  $i$  to node  $j$ . So point centrality of Road 1 is calculated as follows:

$$\begin{aligned} C(v_1) &= \sum_{k=1}^2 (d_k(v_{i-1}, v_i) + d_k(v_i, v_{i-1})) \\ &\quad + \sum_{k=1}^2 (d_k(v_{i-2}, v_{i-1}) + d_k(v_{i-1}, v_{i-2})) \\ &= (1 + 3) + (1 + 2) + (2 + 1) + (5 + 6) = 21. \end{aligned} \quad (3)$$

(2) *Cohesive Subgroup of Roads.* There are some small groups composed of road segments in the road network graph. When a vehicle is on one of these roads, there is a higher probability to traverse the other roads in the small group. So such a small group is named as cohesive subgroup of roads.

In all vehicles' driving routes, some are relatively regular. As shown in Figure 2(a), if Road A is only connected to Road B, the vehicle through Road A will inevitably run toward Road B. Also there are also some parts of vehicle driving routes that appear frequently. In Figure 2(b), three roads A, B, and C intersect at a point and Road B leads to urban business areas while Road C leads to suburbs. So the probability of running towards Road B after Road A will be greater than

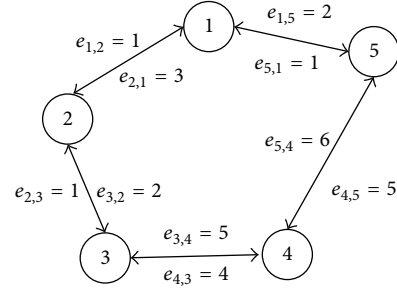


FIGURE 1: An example of a graph.

towards Road C. In other words, the relationship between Roads A and B is much closer than the relationship between Roads A and C. Besides, although two roads are not directly connected, they have also strong relationship. As shown in Figure 2(c), Road B leads to urban commercial areas while Road C leads to suburbs; the probability of running towards Road B after Road D is also greater than towards Road C, where Road A can be regarded as an intermediary between Road D and Road B. Similar to hunting job in real life, the social process that many people find jobs under the help of others often requires a third person (an intermediary) to participate in, but the relationship between two people except for the intermediary is not very close. Therefore, the roads with indirect connections could exist in a cohesive subgroup.

The purpose of studying cohesive subgroup is to discover some small groups with strong relationship in the road network. In the multivalued model, we firstly need to dichotomize the matrix corresponding to a road network, which means to convert all values  $a_{ij}$  to the value 0 or 1. So the matrix after dichotomization can be viewed as an adjacency one of a directed graph. Here, we will provide the description of the dichotomization regularity in detail. Assuming that first there are  $n$  directed edges in the road network and the weights of each directed edge, respectively, are  $x_1, x_2, \dots, x_n$ , then, after dichotomization, corresponding weights  $y_i$  of a directed edge are described in

$$y_i = \begin{cases} 1 & x_i \geq M \\ 0 & x_i < M, \end{cases} \quad (4)$$

where  $M$  represents a threshold and if  $x_i \geq M$ ,  $y_i = 1$  and if  $x_i < M$ ,  $y_i = 0$ .

After transforming a road network into a dichotomization matrix, we describe how to get cohesive subgroups. In the directed graph, the shortest distance of any two points in a cohesive subgroup is less than  $n$ . Assuming that  $d(i, j)$  represents the shortest distance from the nodes  $v_i$  and  $v_j$ , and, in the subgroup with a set  $N_s$ , if  $d(i, j) \leq n$ , for all  $v_i \in N_s$ , there is no point in the directed graph whose shortest distance to any point in a subgroup is more than  $n$ , where  $n$  is the maximum distance between members of a cohesive subgroup, so we describe the set  $N_s$  as a cohesive subgroup. If  $n = 2$ , the members of cohesive subgroup are connected directly (the shortest distance equals 1) or connected indirectly through a common adjacent point



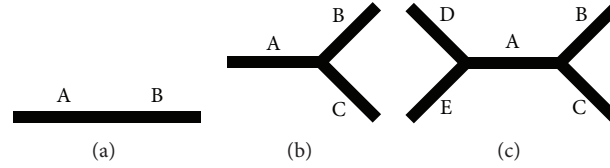


FIGURE 2: Different relationships between roads.

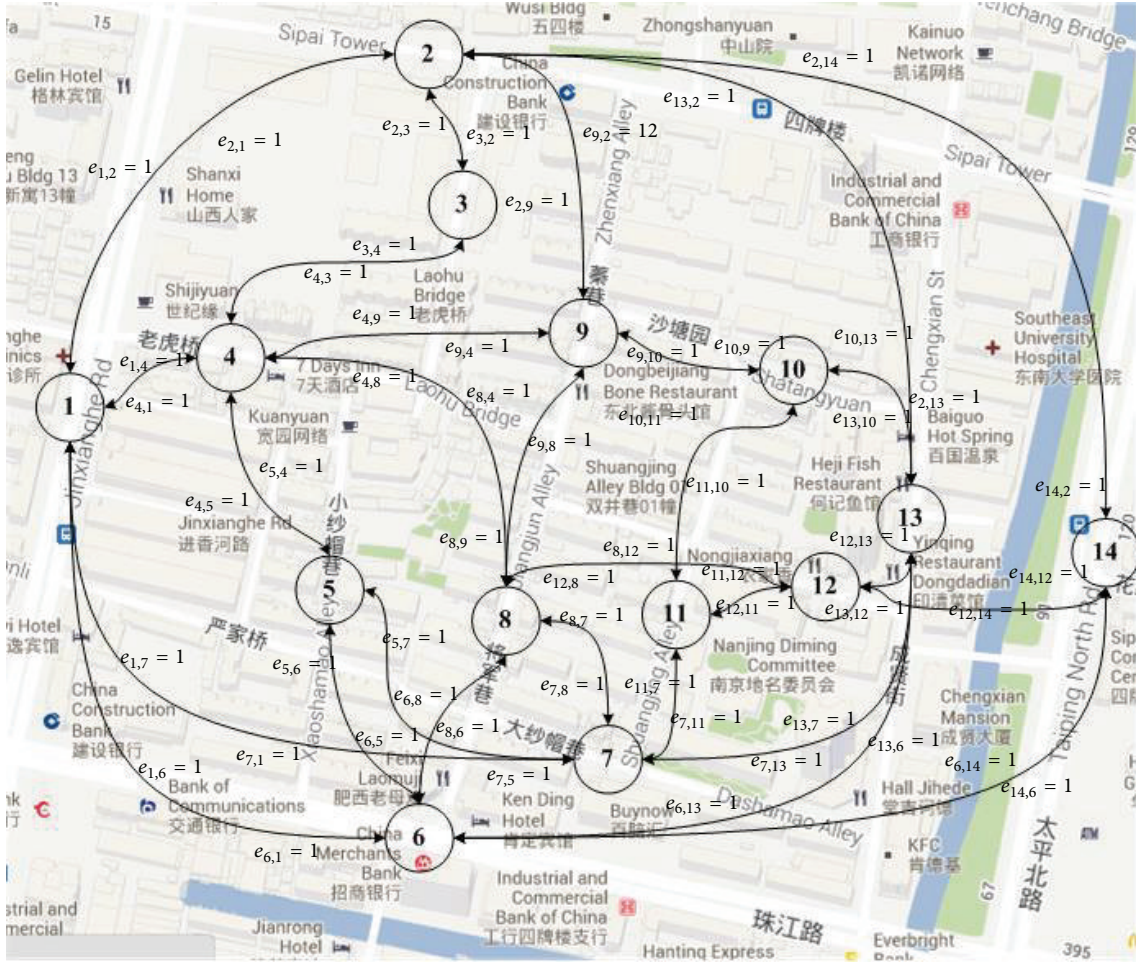


FIGURE 3: The representation of graphs based on road network map.

(the shortest distance equals 2). The larger the  $n$  is, the looser the limitation to every member will be.

#### 4. The Design of Route Prediction Algorithm Based on Social Network Analysis

##### 4.1. A Graph of Road Network

(1) *Build an Initial Graph of Road Network.* Firstly, we need to convert the road network map into an initial graph. According to the previous definition, we need to number each road. Each of the roads is seen as a node and the intersection of them is viewed as an edge, where the initial weight of each edge is 1. In Figure 3 we have numbered all

roads. For example, Jinxianghe Road is numbered as 1. Next, we draw different edges according to the relationship between different roads. For example, Jinxianghe Road is connected to Sipai Tower, and then there are two weighted edges in the initial conditions, which are  $e_{1,2} = 1$  and  $e_{2,1} = 1$ .

(2) *Correct the Weight of Each Edge.* As mentioned above, we focus on the study of relationship between roads, and an effective method to indicate relationship strength between roads is to calculate the number of routes that include the two interconnected roads from historical data. Assuming that there are  $n$  historical routes from road segment  $v_i$  to  $v_j$ , then the weight  $e_{i,j}$  corresponding to edges is represented as  $n + 1$ . The higher the number is, the closer the relationship

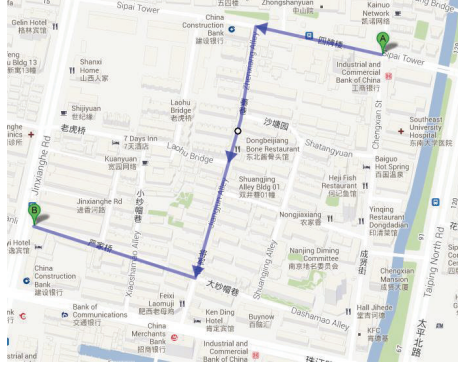


FIGURE 4: A driving route from starting point A to end point B.

between the two roads is. For example, in Figure 4 there is a driving route from the starting point A to the end B, that is, Sipai Tower  $\rightarrow$  Zhengxiang Alley  $\rightarrow$  Jiangjun Alley  $\rightarrow$  Dashamao Alley  $\rightarrow$  Jingxianghe Road. So the weights corresponding to edges are adjusted as  $e_{2,9} = 2$ ,  $e_{9,8} = 2$ ,  $e_{8,7} = 2$ , and  $e_{7,1} = 2$ .

**4.2. Analysis of Road Relationship.** After the road network graph correction, we can establish a matrix. So the matrix is the basis of our route prediction. We use relevant software (such as UCINET [14] or Pajex [15]) to dichotomize the matrix, which could help us analyze point centrality and the cohesive subgroups of nodes from that.

#### 4.3. Route Prediction Algorithm

**Input.** (1)  $V = \{v_1, v_2, \dots, v_n\}$  ( $n \geq 2$ ) is an ordered set of just-driven road segments, where  $v_i$  is the  $i$ th road, and the number of roads in the ordered set is at least two because we can obtain driving direction of vehicles, which is helpful to the vehicle route prediction in the following; (2) point centrality of roads is represented as  $C = \{c_1, c_2, \dots, c_n\}$ , where  $c_j$  is the value of point centrality of the  $j$ th road; (3) the set  $S = \{s_1, s_2, \dots, s_k, \dots, s_m\}$  indicates cohesive subgroups of roads, where  $s_k$  is the  $k$ th cohesive subgroup; that is,  $s_k = \{v_{k1}, v_{k2}, \dots\}$ , and  $m$  is the total number of cohesive subgroups; (4)  $N$  represents the prediction distance; that is, our algorithm will understand the number of roads to be driven on in the future.

Here, we need to illustrate that GPS points themselves are often noisy and some contain invalid sensor data, so if we only transform the raw GPS data into trips without cleaning those, trips comprised of these GPS points will be errors in some extent. At present, there are many methods [16–18] to optimize and clean these raw GPS data in order to improve accuracy. In this paper, we mainly focus on route prediction rather than trip cleansing, so the data inputting into our algorithm should be available ones after cleaning to make sure of prediction accuracy.

**Output.**  $R = \{R_1, R_2, \dots, R_k\}$  is a set of possible routes that a driver will take in the future, where  $R_1$  represents the most possible upcoming route (but the route is not an entire

one because of the limitation of prediction distance), that is, where  $\{v_1, v_2\}$  is the known driving road segments and  $\{v_{i1}, \dots, v_{in}\}$  is the predicted driving road segments, while  $R_k$  represents the least possible upcoming route.

Steps for route prediction are as follows.

(1) Assume that the vehicle has been driven through at least two road segments and the set  $V = \{v_1, v_2, \dots, v_n\}$  ( $n \geq 2$ ) represents the just-driven roads. Then select two road segments  $v_n$  and  $v_{n-1}$  from the set  $V$ , where  $v_n$  is the current road that a vehicle is driven on and  $v_{n-1}$  is the former driven road.

(2) Traverse the cohesive subgroups  $s_j$  from the set of cohesive subgroups of roads  $S$  to judge whether the roads  $v_{n-1}$  and  $v_n$  belong to  $s_j$  and find all of cohesive subgroups with road segments  $v_{n-1}$  and  $v_n$ .

(3) Find road segments except  $v_{n-1}$ ,  $v_n$ , and roads indirectly connected to  $v_n$  from those cohesive subgroups, and then rank the road segments in descending order with point centrality value corresponding to these roads according to the set  $C$  of point centrality of roads. Assume that final results of road segments are inserted into the set  $P$  and the number of elements of  $P$  is  $m$ .

(4) The road segment with a larger point centrality indicates that vehicles may more likely pass through it, so the road that the vehicle is more likely to drive through is  $v_{n+1} = P_1$ , where  $P_1$  represents a road segment with maximum point centrality. If the prediction distance is  $n$ , where  $n \leq N$ , the first prediction routes are as follows:

$$\begin{aligned} R_1 &= \{v_{n-1}, v_n, P_1\} \\ R_2 &= \{v_{n-1}, v_n, P_2\} \\ &\vdots \\ R_m &= \{v_{n-1}, v_n, P_m\}, \end{aligned} \quad (5)$$

where  $R_1$  represents the most likely upcoming route and  $R_m$  is the least one.

(5) The above routes comprised of road segments are regarded as inputting parameter  $V$  of our prediction algorithm and follow recursive procedure until prediction distance  $n$  is equal to  $N$ .

To clearly illustrate the process of our algorithm, we create the dichotomization matrix and describe the directed graph shown in Figure 5 based on above descriptions. Assume that a vehicle has driven through two roads  $v_{13}$  and  $v_{12}$ , and the driving direction is from  $v_{13}$  to  $v_{12}$ . We also assume the prediction distance  $N = 3$  here.

(1) Obtain all cohesive subgroups with  $v_{13}$  and  $v_{12}$ , that is,  $\{v_{13}, v_{12}, v_{11}\}$  and  $\{v_{13}, v_{12}, v_{14}\}$ . Combine with those to form a set  $\{v_{13}, v_{12}, v_{11}, v_{14}\}$  in Figure 5(a).

(2) Regardless of elements  $v_{13}$  and  $v_{12}$ , calculate point centralities of  $v_{11}$  and  $v_{14}$ , respectively. Assume that the value of point centralities of  $v_{11}$  and  $v_{14}$  are, respectively, 7 and 3. Then rank  $v_{11}$  and  $v_{14}$  based on the value of point centralities in descending order to get an ordered set  $P = \{v_{11}, v_{14}\}$ .

(3) In Figure 5(b), when the prediction distance  $n$  is equal to 1, then  $R_1 = \{v_{13}, v_{12}, v_{11}\}$ ,  $v_{n-1} = v_{12}$ ,  $v_n = v_{11}$ , and  $P = \{v_7\}$ . Next repeat the above steps.

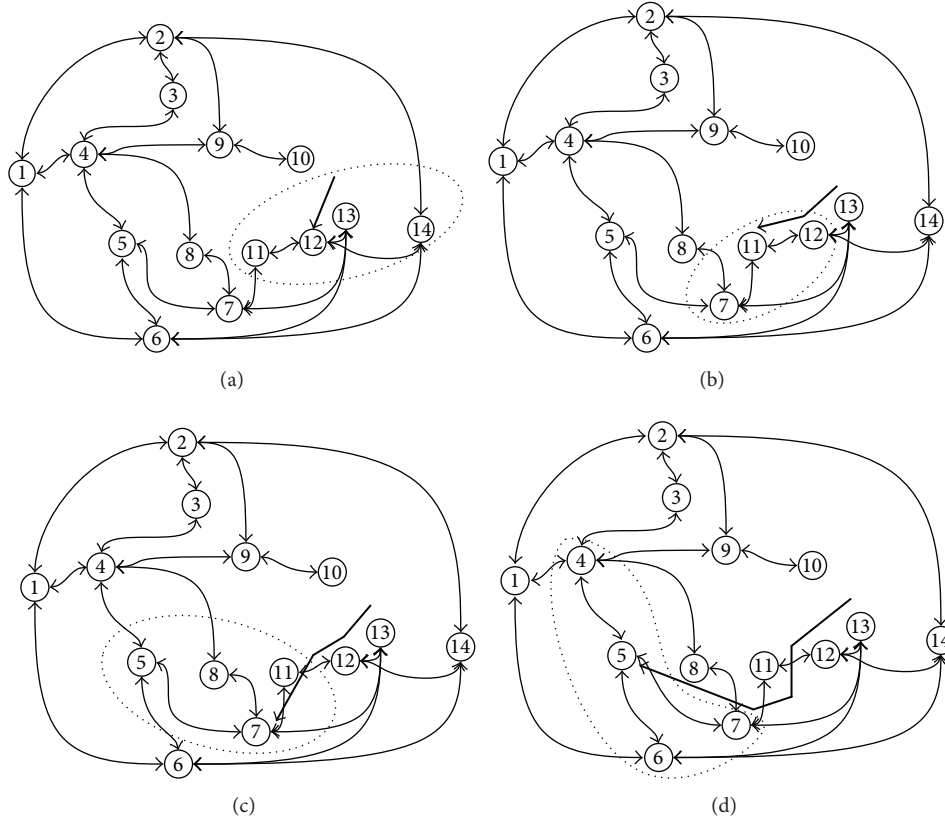


FIGURE 5: The process of route prediction algorithm based on social network analysis.

(4) In Figure 5(c), after  $n$  is added to 2,  $R_1 = \{v_{11}, v_7, v_8, v_5\}$ ,  $v_{n-1} = v_{11}$ ,  $v_n = v_7$ , and  $P = \{v_5, v_8\}$ .

(5) In Figure 5(d), after  $n$  is added to 3, the route consisting of roads with maximum point centralities is  $R_1 = \{v_{13}, v_{12}, v_{11}, v_7, v_5\}$ .

(6) The other possible routes with  $N = 3$  are  $R_2 = \{v_{13}, v_{12}, v_{11}, v_7, v_8\}$ ,  $R_3 = \{v_{13}, v_{12}, v_{14}, \dots\}$ , and so on.

## 5. Experiment

The data used to test our prediction algorithm come from Microsoft Multiperson Location survey (MSMLS) [19], where the data set is mainly comprised of the GPS data. Here we apply a leave-one-out approach [8], meaning that most of the data are used to correct weights of each edge described in Section 4 while the others are used to verify the efficiency of the prediction algorithm. We need to point out that there are some problems of the GPS data in the city, including dropouts produced in the area with intensive buildings and other cases like noise and offset of GPS [20]. Therefore, we have to correct part of data properly to ensure that the GPS data correspond to road segments in the city. Then we show vehicle routes consisting of the GPS data in Google Map software. Then we have established the road network model after correcting different edge weights.

### 5.1. Evaluation Criteria

(1) *Prediction Accuracy.* Our method can predict all the possible driving routes for drivers but a vehicle may only drive through one of those routes, so we need to exclude the redundant routes as large as possible to provide drivers with the best route. Therefore, we only consider the most possible route. As described before, prediction distance  $N$  limits the number of road segments in a prediction route. The evaluation criterion of algorithm's prediction accuracy is shown in

$$\text{Precision} = \frac{\text{Numof}_{\text{sim}}(R_1, R)}{\text{Numof}(R_1)} \times 100\%, \quad (6)$$

where  $R$  represents the actual driving route with  $(N + 2)$  road segments ( $N$  is the prediction distance),  $\text{Numof}_{\text{sim}}(R_1, R)$  represents the number of the same road segments between the most likely route  $R_1$  and  $R$ , and  $\text{Numof}(R_1)$  indicates the number of road segments in the route  $R_1$ .

(2) *Prediction Integrity.* Prediction integrity shown in (7) is used to evaluate integrity of the route from our prediction results compared with the actual driving route:

$$\text{Integrity} = \frac{\text{Numof}(R_1)}{\text{Numof}(AR)} \times 100\%, \quad (7)$$



where Numof(AR) represents the number of roads in the actual driving route.

**5.2. Experimental Results.** We both assume that the vehicle has driven through two roads, and in the process of dichotomization, we regard the average weight of each directed edges in road network model as the threshold  $M$  described in

$$M = \frac{1}{n} \sum_{j=1}^n x_j. \quad (8)$$

Figure 6 illustrates the relationship between prediction accuracy and prediction distance  $N$  in our algorithm and Jon Froehlich's. We understand that the prediction accuracy will drop with the increase of prediction distance  $N$  in both algorithms. And the larger the  $N$  is, the smaller the prediction accuracy is. In our algorithm, when  $N$  is from 1 to 3, the prediction accuracy drops slightly. But, from the beginning of  $N = 4$ , the accuracy decreases extremely. In addition, the accuracy of prediction with repeated data is higher than all of route data (i.e., part of test data does not appear in training data). In Jon Froehlich's algorithm, when  $N$  is from 1 to 3, the accuracy is lower than our algorithm, but, after  $N = 4$ , the accuracy is higher than our algorithm in both repeated and all trips.

Jon Froehlich's method firstly analyzes route regularity from the collected GPS data; that is, a large portion of a typical driver's trips are repeated. So they exploit this fact for prediction by matching the first part of a driver's current trip with one of the set of previously observed trips. However, there are some problems in their method. First, if previously driving routes from a driver have been collected in the historical data but the driver's current trip never occurs before when predicting upcoming route, prediction accuracy will drop due to the new route even though road segments of the new routes coincide with previous routes in the collected data. Additionally, if previously driving routes from a driver never existed in the historical data set, then Jon Froehlich's algorithm will not find the driver's route regularity, which will greatly reduce the accuracy of the prediction algorithm. In our algorithm, we find relationship between road segments from all historical data rather than consider each entire route. And we also use social network analysis theory to explore the potential relationship between road segments. Even though the problem of above new routes will also exist, we know the internal relationship between each road to improve the accuracy of prediction. In addition, from the beginning of  $N = 4$ , the accuracy decreases extremely. The accuracy of Jon Froehlich's algorithm is lower than our algorithm, but, after  $N = 4$ , the accuracy is higher than our algorithm in both repeated and all trips. From the perspective of social network analysis, we find that the possibility of transmitting information from one side to another is lower when the relationship between two social actors is alienated. The greater the prediction distance  $N$  is, the less the relationship between the current road and the farthest one is. So the results illustrate that our algorithm has higher accuracy in short-term route prediction and Jon Froehlich's algorithm performs

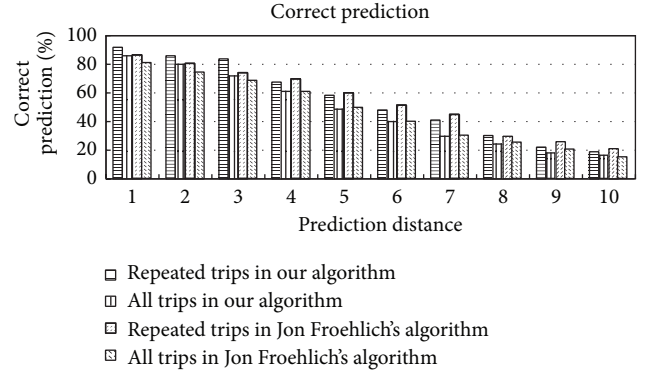


FIGURE 6: Correct prediction by prediction distance.

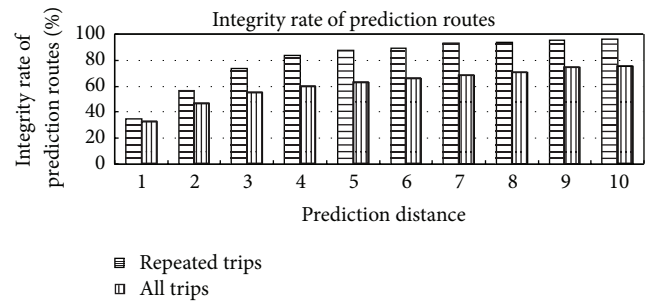


FIGURE 7: Integrity rate of prediction routes by prediction distance.

better in long-term prediction. In my opinion, the former is more important because the high prediction accuracy could afford much highly reliable application for drivers. With the increase of prediction distance, prediction accuracy drops sharply and it is little helpful for practical applications.

Figure 7 reveals the relationship between prediction integrity and prediction distance  $N$ . The larger the prediction distance is, the greater the integrity of predicted route compared to the actual route is. When  $N = 3$ , the route integrity in the set of repeated data can reach 73.57%; that is, the number of roads the vehicle has actually passed through is 5, including two driven roads.

The prediction accuracy and prediction integrity are contradictory. The larger the prediction distance  $N$  is, the lower the prediction accuracy is and the greater the route integrity is. The goal of vehicle route prediction is to restore the integrity of driving routes as large as possible. If the prediction distance  $N$  is too small, the significance of vehicles prediction will not be fully reflected. Therefore, when the prediction distance is equal to 1, there is little advantage to help the driver avoid congested roads. From Figures 6 and 7, in the real applications, we consider that the prediction distance  $N = 3$ .

Figures 8 and 9 are used to verify the effect of dichotomization on prediction accuracy and prediction integrity. Assume that the current vehicle has been driven through two roads and, in order to reduce the impact of prediction distance  $N$  on dichotomization as much as possible, suppose that  $N = 3$  in Figure 8 and  $N = 10$  in

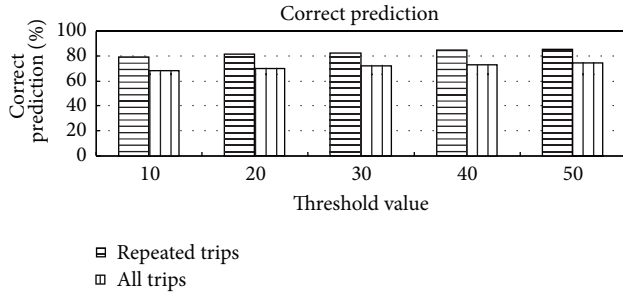


FIGURE 8: Correct prediction by threshold value.

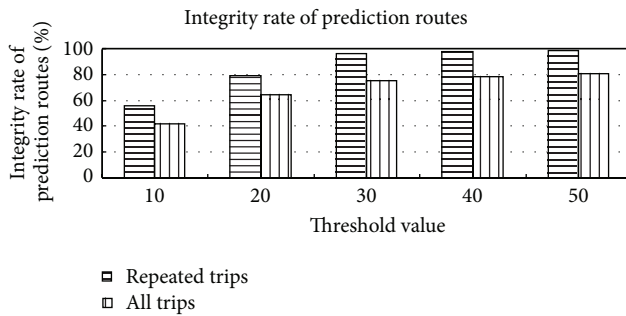


FIGURE 9: Integrity rate of prediction routes by threshold value.

Figure 9. From the two figures we understand that, with the increase of the threshold  $M$ , the effect on prediction accuracy is slight, but that on prediction integrity is serious. We choose the route with highest probability from candidates; that is, the point centrality of roads in the route is the largest. If  $M$  is smaller, the requirement of members in the cohesive subgroups is stricter, but the corresponding roads are still in cohesive subgroups because the point centrality of each road is higher. Instead, the strict requirement of members in cohesive subgroups inevitably leads to decreasing the number of members so that a certain node in the directed graph becomes the leaf node after dichotomization; that is, there is no connection between a certain road and others.

## 6. Conclusion

This paper mainly defines the relationship between different roads based on the method of social network analysis so as to predict possible routes in the future. First of all, we introduce the method of road network modeling. Then we illustrate the concepts of point centrality of roads and cohesive subgroup of roads and, based on these concepts, we correct the existing road network model. Finally, we design a valid route prediction algorithm and verify the effectiveness by experiments.

In the following studies, we will pay attention to the impacts of road congestion and advertisement on the route choice. But we need to comprehensively consider the effect of different information for the driving routes. Therefore, we need to design a method of route prediction to think about the interconnection between predicted routes and the actual driving routes.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This research was performed in cooperation with the institution. The research is supported by National Natural Science Foundation of China (nos. 61170065 and 61003039), Peak of Six Major Talent in Jiangsu Province (no. 2010DZXX026), China Postdoctoral Science Foundation (no. 2014M560440), Jiangsu Planned Projects for Postdoctoral Research Funds (no. 1302055C), Jiangsu provincial research scheme of natural science for higher education institutions (no. 12KJB520009), and Science & Technology Innovation Fund for higher education institutions of Jiangsu Province (no. CXZZ11-0405).

## References

- [1] G. Dimitrakopoulos and P. Demestichas, "Intelligent transportation systems," *IEEE Vehicular Technology Magazine*, vol. 5, no. 1, pp. 77–84, 2010.
- [2] S. Grant-Muller and M. Usher, "intelligent transport systems: the propensity for environmental and economic benefits," *Technological Forecasting and Social Change*, vol. 82, no. 1, pp. 149–166, 2014.
- [3] S. Zeadally, R. Hunt, Y.-S. Chen, A. Irwin, and A. Hassan, "Vehicular ad hoc networks (VANETS): status, results, and challenges," *Telecommunication Systems*, vol. 50, no. 4, pp. 217–241, 2012.
- [4] H. Omar, "Introduction," *Journal of Pediatric and Adolescent Gynecology*, vol. 26, no. 3, supplement, p. S1, 2013.
- [5] N. Ye, Z.-Q. Wang, R. Malekian, Q. Lin, and R.-C. Wang, "A method for driving route predictions based on hidden Markov model," *Mathematical Problems in Engineering*. In press.
- [6] A. Karbassi and M. Barth, "Vehicle route prediction and time of arrival estimation techniques for improved transportation system management," in *Proceedings of the Intelligent Vehicles Symposium (IV '03)*, pp. 511–516, IEEE, June 2003.
- [7] J. Krumm, "A markov model for driver turn prediction," SAE Technical Paper, 2008.
- [8] R. Simmons, B. Browning, Y. Zhang, and V. Sadekar, "Learning to predict driver route and destination intent," in *Proceedings of the IEEE Intelligent Transportation Systems Conference (ITSC '06)*, pp. 127–132, IEEE, September 2006.
- [9] J. Froehlich and J. Krumm, "Route prediction from trip observations," SAE SP 2193:53, 2008.
- [10] T. Palonen and K. Hakkarainen, "Patterns of interaction in computer-supported learning: a social network analysis," in *Proceedings of the 4th International Conference of the Learning Sciences*, pp. 334–339, 2013.
- [11] C. Ünsalan and B. Sirmacek, "Road network detection using probabilistic and graph theoretical methods," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 11, pp. 4441–4453, 2012.
- [12] T. Palonen and K. Hakkarainen, "Patterns of interaction in computersupported learning: a social network analysis," in *Proceedings of the 4th International Conference of the Learning Sciences*, pp. 334–339, 2013.



- [13] A. Marin and B. Wellman, "Social network analysis: an introduction," in *The SAGE Handbook of Social Network Analysis*, pp. 11–25, SAGE, 2011.
- [14] S. P. Borgatti, M. G. Everett, and L. C. Freeman, *Ucinet for Windows: Software for Social Network Analysis*, 2002.
- [15] V. Batagelj and A. Mrvar, "Pajek-program for large network analysis," *Connections*, vol. 21, no. 2, pp. 47–57, 1998.
- [16] X. Cao, G. Cong, and C. S. Jensen, "Mining significant semantic locations from GPS data," in *Proceedings of the VLDB Endowment*, vol. 3, pp. 1009–1020, 2010.
- [17] M. J. Duncan, W. K. Mummery, and B. J. Dascombe, "Utility of global positioning system to measure active transport in urban areas," *Medicine and Science in Sports and Exercise*, vol. 39, no. 10, pp. 1851–1857, 2007.
- [18] J. Yuan, Y. Zheng, X. Xie, and G. Sun, "T-drive: enhancing driving directions with taxi drivers' intelligence," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 1, pp. 220–232, 2013.
- [19] J. Krumm and E. Horvitz, "The Microsoft multi-person location survey," Microsoft Research Technical Report, 2005.
- [20] Y. Ning, W. Zhong-qin, R. Malekian, W. Ru-chuan, and A. H. Abdullah, "Design of accurate vehicle location system using RFID," *Elektronika ir Elektrotechnika*, vol. 19, no. 8, pp. 105–110, 2013.

