# Mobility Dataset Generation for Vehicular Social Networks Based on Floating Car Data

Xiangjie Kong, *Senior Member, IEEE*, Feng Xia, *Senior Member, IEEE*, Zhaolong Ning, *Member, IEEE*, Azizur Rahim, Yinqiong Cai, Zhiqiang Gao, and Jianhua Ma

***Abstract*—Vehicular social networks (VSNs) have attracted the research community due to its diverse applications ranging from safety to entertainment. Social vehicles standing for private cars and floating cars standing for taxis are two important components of VSN. However, the lack of social vehicles data causes some factors are neglected including social aspects and macroscopic features, which blocks researching social attributes of vehicles in VSN. Generating a realistic mobility dataset for VSN validation has been a great challenge. In this paper, we present the detailed procedure to generate social vehicular mobility dataset from the view of floating car data, which has the advantage of wide universality. First, through the deep analysis and modeling of the dataset of floating cars and combining with the official data, we predict the origin–destination (OD) matrix of social vehicles with the gravity model, and then calibrate the OD matrix with the average growth factor method. Second, we construct network description after editing the road network. Third, we use simulation of urban mobility to reproduce the scenario in view of microsimulation by generating the mobility dataset of social vehicles based on floating car data and urban functional areas. At last, we prove the effectiveness of our method by comparing with real traffic situation in Beijing. The generated mobility model may not accurately represent the mobility of social vehicles in few spots, such as train station or airport, however, exploiting figures and facts of transportation in the city have been considered in the study to calibrate the model up to maximum possible realization.**

***Index Terms*—Human mobility, dataset generation, vehicular social networks, floating car data, urban functional areas.**

## I. Introduction

THE ever-increasing number of vehicles, roads, and advancement of roadside infrastructure with a diverse range of communication capabilities has led to an inter-dependent communication network of vehicles on the roads. Smart cars have provided users with a new mobile communication, and Vehicular Social Networks (VSN) have become a new typical Cyber-Physical Social System (CPSS) [1]. VSN can be formed when people drive or take a car, and viewed as a mobile communication network containing users and relationships between them [2]. VSN is a Vehicular Ad-hoc Network (VANET) and contains social relationships such as human mobility pattern, user selfishness, and user selectivity. The core contents of VSN are road network and mobile vehicles on the road. VSN can be regarded as a virtual community, which is formed by location and community interest groups as well as cooperation among drivers [3]–[5]. As shown in Fig. 1, VSN is composed of two layers, the upper layer includes social network service, map service, and video sharing service, while the lower layer contains dynamic communication topology, vehicles gathered hotspot and mobile user model. The two layers are connected by base

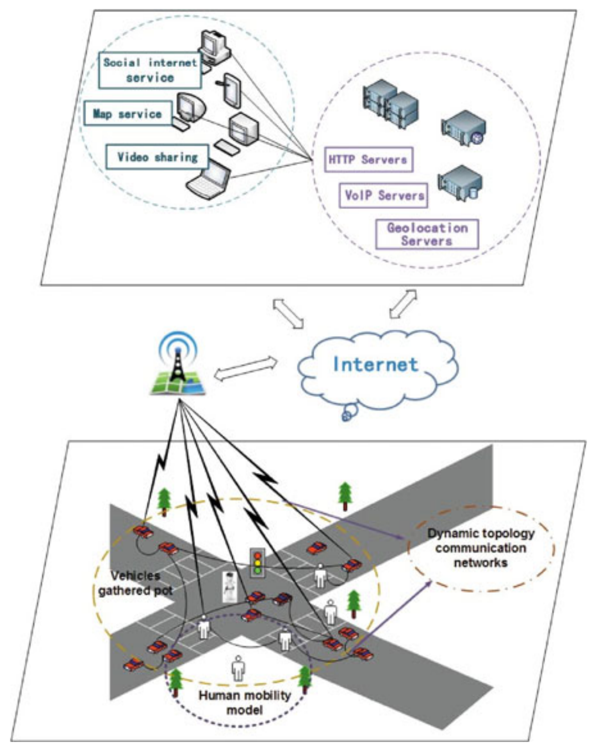



Fig. 1. The structure of VSN.

Manuscript received May 21, 2017; revised October 1, 2017 and December 11, 2017; accepted December 23, 2017. Date of publication January 1, 2018; date of current version May 14, 2018. This work was supported in part by the Natural Science Foundation of China under Grants 61572106 and 61502075, in part by the Natural Science Foundation of Liaoning Province, China, under Grant 201602154, in part by the Dalian Science and Technology Planning Project under Grants 2015A11GX015 and 2015R054, and in part by the Fundamental Research Funds for Central Universities under Grant DUT15YQ112. The review of this paper was coordinated by Prof. C. Assi. *(Corresponding author: Feng Xia.)*

X. Kong, F. Xia, Z. Ning, A. Rahim, Y. Cai, and Z. Gao are with the Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, Dalian 116620, China, and also with the School of Software, Dalian University of Technology, Dalian 116620, China (e-mail: xjkong@ieee.org; f.xia@ieee.org; zhaolongning@dlut.edu.cn; aziz.buner@outlook.com; 1198593462@qq.com; qiang.code@gmail.com).

J. Ma is with the Faculty of Computer and Information Sciences, Hosei University, Tokyo 184-8585, Japan (e-mail: jianhua@hosei.ac.jp).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2017.2788441

stations and the Internet. In practice, moving vehicles and users of the lower layer upload their position information and communication data by intelligent devices. After processing, the upper layer provides video, vehicle position, traffic condition to other users. Thus, users complete interaction with each other. VSN has the features of self-organizing, active community forming and rich social information. These unique characteristics bring new challenges and opportunities for the VSN applications and provide new research directions such as smart city, intelligent transportation, and social network [6]–[9].

As the cornerstone of VSN, obtaining vehicular mobility dataset in large scale accurately makes various kinds of research conceptions possible, including:

1) Making it possible to response the transport infrastructure condition rapidly and timely. For example, when a disaster occurs, traffic data can provide the most direct and effective information and decision support with the help of remote sensing imaging technology and regional traffic monitoring.
2) Making it possible to manage public travelling better. For example, we can observe traffic condition continuously by road detection technology, which can help us to plan the routes of public travelling.
3) Making real-time status tracking of goods flow possible. For example, the use of radio frequency technology can track goods in real time through the logistics of the whole process by dynamic monitoring of auxiliary to make a decision rapidly; this method can improve logistics efficiency and reduce logistics costs.

While people are traveling on the roads, they can join all kinds of social activities including music sharing, news sharing, voice chat in groups, which provides a chance to interact and know each other. In the society with highly developed information technology, not only mobile phones but also vehicles can be used for social activities. This newly emerging field of VSNs comes up with some challenges. For example, how to find a new VSN application, whether it can inspire friends to increase the energy-saving emission reduction and comply with traffic regulations, whether friends can share the road traffic information, and whether it can share social attributes between car friends [10]. To solve these problems, we need mobility dataset and social properties of vehicles so that data forwarding, data storing and vehicular friends mining can be analyzed. However, GPS data of social vehicles, which stand for private vehicles in this paper, cannot be obtained due to the restrictions of privacy and security, which causes the mobility dataset of social vehicles cannot be acquired currently. However, according to Beijing Transport Annual Report of 2015, social vehicles occupy about one-third in ground traffic, so the mobility dataset of social vehicles is indispensable while evaluating the performance of protocols or new traffic surveillance systems.

To solve the problems aforementioned, we present a method to generate a large-scale mobility dataset for VSNs based on floating car data. Floating car data is a method to sense the traffic status on the road network. In this work, we refer taxis with GPS equipments as floating cars. The mobility of floating cars differs in nature as compared to social cars, and mobility model purely based on floating cars may not represent the social cars. Thus, we combined the valuable information available from Beijing Traffic Development Annual report and exploit the mobility features and facts to calibrate the model. Our most important contributions can be summarized as follows:

1) We dig out the human mobility pattern based on urban functional areas after deep analysis of the floating car dataset and the characteristics of each trip.
2) We propose a mobility dataset generation method of social vehicles based on floating car data, which is perfectly universal and considers connections between different urban functional areas.
3) We fulfill scenario construction from microscopic view by generating about 14 million travelling trajectories of social vehicles in three days and complete the experiments based on floating car data and urban functional areas inside fifth circle highway of Beijing.

The rest of this paper is organized as follows. Related work is summarized in Section II. Section III presents our mobility dataset generation method in detail, and Section IV illustrates scene realization. In Section V, we conclude our work and point out the future direction.

## II. Related Work

### A. *The Application of Urban Big Data*

Nowadays, with the rapid development of the information society, a massive variety of urban data is available, and researchers have made many efforts to excavate these data.

In [11], the authors focused on the next-generation social network of vehicles after reviewing the literature on VSN and summarized characteristics of social networks from the latest technology to mobile applications and challenges. Due to limited connection to Internet contents and services on the highway, Luan *et al.* [12] proposed the SOR application, which is designed to help vehicles interact with each other during traveling on the highway. This application encourages each user to maintain a personal blog similar to Facebook and twitter so that they can upload some videos and images to share information with others. In order to keep communication stable without accessing to the Internet on the roads, the authors in [13] developed Verse to promote social activities on the highway, which can provide friends recommendation for passengers to identify potential friends effectively. Normally, these friends have similar interests, and they can maintain a relatively reliable wireless communication link.

Xu *et al.* [14] put forward a new online architecture, which can analyze the real-time traffic data to predict conditions of next stage. Apart from the research of traffic flow prediction, traffic safety also received a lot of attention in the context of road expansion recently. Kong *et al.* [15] employ particle swarm optimization method to predict traffic congestion effectively and improve travel efficiency. The authors in [16] discovered the defects of existing technologies after reviewing much work on traffic safety. Although a lot of technologies have considered distance, speed, and other basic physical information, these factors do not reflect the physical diversity and rapidly changing traffic conditions. Therefore, in order to improve the level of road safety, developing a real-time traffic data service technology is advocated. In terms of detection of urban functional areas, Yuan *et al.* [17] proposed a data driven architecture to detect different functional areas in the city. Latent Activity the Trajectory (LAT) concept has been introduced and they divide Beijing into different functional areas based on LAT and traffic data. Kong *et al.* [18] take into account the relationship between picking up and dropping off to identify different urban functional regions, and recommend the best locations for taking a taxi.

### *B. Simulation of Urban Mobility Tool*

Simulation of Urban Mobility (SUMO) is an open source tool to simulate traffic condition, which is used to simulate the movement of vehicles in the city [19]. Viewed from the simulation content, SUMO is a space-continuous, discrete-time microscopic simulation package, including road network import and demand modeling components. With the help of SUMO, researchers are able to study urban traffic conditions more deeply. One of the most common scenarios is V2V data transmission, and SUMO can generate the vehicle movement trajectory. The microscopic mobility models implemented by SUMO are Krauss car-following model and Krajzewiczs lane changing model [20]. After the simulation, we can use TraCI to import these tracks into mobile network protocol simulation tools like NS2 and NS3. Compared with commercial software Vissim, although the traffic light algorithm in SUMO is relatively rough, its fast execution time and open API interface make SUMO become a strong candidate to evaluate a new traffic control algorithm. In addition, SUMO can produce a variety of output files to represent noise emissions, pollution emissions and energy consumption. HARMONOISE model is used to generate noise emissions and pollution emissions. Energy consumption is realized by persistence model stored in HBEFA database [21].

Compared to microscopic simulators, macroscopic traffic simulators focus on the traffic flows but do not take into consideration the behavior of a single vehicle in the traffic flow. For the VSN, the behavior of a single vehicle is vital. Thus microscopic simulators are more appropriate. As a microscopic traffic simulator, SUMO supports on-line interaction and closed-loop feedback through the TraCI interface, that is why we choose it to complete our simulation

### *C. Mobility Modeling & Dataset Generation*

Härri *et al.* [22] are the first to put forward the guidelines of vehicle trajectory generation model, and provide an overview and comparison of different mobility models proposed for VANETs. Uppoor *et al.* [23] optimize social vehicle trajectories of Köln by simulation. In order to obtain the dynamic balance, they use Gawron algorithm to iterate the experiment. After that, each vehicle is assigned to appropriate path, which can reduce road congestion and optimize the allocation of entire traffic flows. Finally, they generate the trajectories of social vehicles lasting 24 hours in Köln, German. They prove this mobility dataset would improve the performance of network protocol in simulation, and how incomplete mobility dataset of vehicle would lead to over-optimistic network connectivity. However, it is impossible to generate dataset by their method if there is no relatively accurate government research report. Pigné *et al.* [24] generate movement trajectories of vehicles in Luxembourg using microscopic simulation tool based on the traffic flows obtained from induction. However, traffic inductions are only installed on the main roads in the city or highways out of the city and do not cover the whole city. Besides, because the trajectories are macro, they cannot show the specific driving conditions of each car. Also, the movement trajectories do not cover the whole city scope and have a shorter duration.

In terms of road traffic image, Ferreira *et al.* [25] reappear the traffic scenarios by using macroscopic traffic information and capturing real-time trajectories of vehicles by three-dimensional imaging technology. In the experiment, they use private aircraft to photograph traffic conditions of Porto, Portugal every five seconds, which lasts one week. By studying the image, authors reconstruct the positions and trajectories of 10566 cars in the city. Thakurzx *et al.* [26] use roadside surveillance cameras to gain coarse information of traffic flows in ten urban areas, such as London, Sydney, Toronto. The data can be used to calibrate microscopic vehicular mobility. However, their work costs high consumption of resources and needs the support of strong image processing technology. Besides, aircraft takes photos in a short time, so it cannot be used to generate the large-scale dataset.

So far, the vehicle trajectory by simulation in Zurich, Switzerland is the largest scale that has been generated. Based on the parallel system proposed in [27], the traffic conditions lasting 24 hours provided by Swiss Regional Planning Bureau and the national travel report of Swiss in 1994, Cetin *et al.* [28] put forward a large-scale and multi-tasking microscopic simulation system and generate movement trajectories of vehicles covering 650002 km. Their dataset has a long duration and large coverage, but they had not taken into account of the variety of traffic situation every day. Codeca *et al.* [29] create a simulation scenario of Luxembourg, a mid-sized European city. The scenario covers an area about 156 $km^2$ which includes different types of roads up to 931 km. They retrieve all needed information from road topology and use the population data published on Luxembourg National Institute of Statistics and Economic studies (STATEC) to generate traffic demand by ACTIVITYGEN tool in SUMO. Finally they compare the results of simulation with the Typical Traffic option in Google Maps.

Gramaglia *et al.* [30] use fine-grained traffic counts on the highways near Madrid to generate vehicular mobility trajectories for different time-spans of several workdays. They mainly focus on the highways, which characterize high speed and continual overtaking. So they use microscopic car-following and lane-changing models to describe the behavior of drivers. Similarly, Bedogni *et al.* in [31] contribute to the ongoing efforts and develop dependable and publicly available mobility traces. The authors implement an original version of SUMO to allow importing OSM data in a neat automated fashion. Then they generate an original dataset of road traffic in Bologna, Italy. Finally they provide a novel validation methodology that builds on open data provided by navigation service, and use to assess the quality of the proposed Bologna dataset. However, the automated fashion conversion of OSM data in SUMO may not represent the actual road traffic scenario.

Compared to previous work, based on the data of floating cars, the method we proposed in this paper breaks through the dependence of traffic simulation on the investigation report and has good reproducibility, which is a new attempt of microscopic simulation scene. Also, the comprehensive social vehicular dataset which is obtained through our method will contribute to the simulation and evaluation of the communication protocol in VSN.

## III. Mobility Dataset Generation Method

Vehicular social networks highlight the social element, which means relationship is one of the key factors. If the possible social relationship can be dig out, ITS will develop rapidly, and quantities of needs in modern society will be satisfied, such as reducing resource consumption and pollution emissions, broadcasting traffic condition between cars on the road. In terms of the current situation of the development of VSN, although we have big traffic data like RFID data, taxi GPS data, bus GPS data, video data and POI data, it is still not enough [31]. To achieve the goals aforementioned, we need more trajectories of social vehicles representing travel behavior of urban residents. However, due to the privacy and security of social vehicles, it is unable to get the required data. Therefore, the development of VSN is blocked by the lack of social vehicle trajectory. In this paper, we propose a mobility dataset generation method of social vehicles based on floating car data. According to the report of Beijing Transportation Bureau, the taxi trips occupy over 12 percent of traffic flows on road surfaces, so the taxi trips take a significant portion of people's urban mobility [32]. Since both floating car and social vehicles can represent the people's urban mobility pattern, we can build mobility model based on the floating car data, and predict the traveling data of social vehicles according to the model. With the help of SUMO, we can generate a large number of social vehicle trajectories with high reality in the city. Besides, if floating car data or other transport vehicles in other cities can represent human's urban mobility(actually, there are over 30 cities around the world having over 10,000 taxicabs and it takes a large proportion in civil transportation), social vehicles traveling data can also be generated by the method proposed in this paper.

As shown in Fig. 2, our proposed dataset generation method mainly includes three parts: demand description, network description, and simulation. Demand description is leveraged to calculate the OD Matrix of social vehicles. We get the traffic volume of social vehicles with the amount ratio of social vehicles and taxis on main streets in different urban functional areas. Then we generate the OD Matrix between the various functional areas of social vehicles using the Gravity Model and employ the Average Growth Factor Method to calibrate it. In the part of network description, we mainly deal with the road network obtained from OSM. We modify it so that we can obtain a road topology for simulation, which matches to the real world as much as possible. Finally, with road network data and OD Matrix, we use SUMO to complete the simulation in the part of traffic simulation.

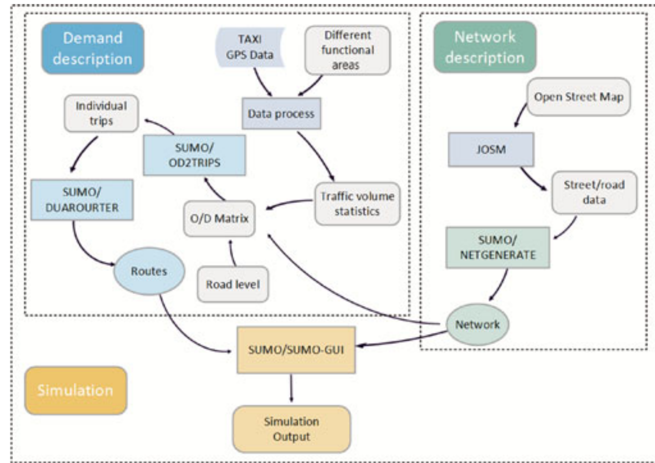



Fig. 2. Flow diagram of dataset generation methodology.

### A. Demand Description

Urban functional area refers to an area which surrounds some special functional buildings and has a similar function from the macro perspective. The urban functional area is different from the administrative region, but both of them can be refined into one or more parts to some extent. Besides, urban functional areas associate tightly with human mobility pattern, that when people reach or leave a region and where people come from and leave for [32]. There is a relationship between a particular type of land use, which is called functional area more accurately, and the traffic it generates [33]. In other words, the functional areas where a trip begins at and ends at basically determine the trip's characteristic. Floating cars are universal in cities, and they are equipped with the Global Positioning System (GPS) to record vehicle location, direction, and instantaneous velocity regularly. Through map matching algorithm and some other algorithms, vehicle information and road data can be connected in time and space. Moreover, we can aggregate the trips of floating cars which have the same characteristic to find the interior principles.

For the collected taxi GPS data, we clean them to remove the data exception due to the incomprehensive storage and inappropriate operating, such as the duplicate data, incomplete data and error data. To get a more accurate result and reduce the statistical error, we choose the region segmentation method. We divide the city into small grids according to the longitude and latitude with 0.05 range and mark each grid with a unique ID. Then we count the ID of grids that each region includes. As for the grid located at the intersection of two functional areas, we allocate it to the bigger one. The mobility traces of floating car (taxi data) may not accurately represent the mobility pattern of social vehicles. However, the mobility pattern of floating cars gives insights to a number of parameters which can be considered for social cars mobility generation, such as traffic density on different roads, number of pick-up points, and number of destinations; to mention a few. Also, the surge in development of online taxi-calling applications had significantly attracted the people to use taxis. Thus taxi mobility data shows similarity to social car mobility in terms of trips originated from and drawn to a particular point of interest. The taxi data contains the records for both cases, such as the empty taxi and occupied by a passenger. However, we are only interested in the later one. So, we process the floating car data to infer the statistics of social cars occupied by passengers.

In order to get the OD Matrix of social vehicles, we do as follows:

1) Traffic volume (the number of vehicles departing from the traffic region) and traffic absorptive volume (the number of vehicles arriving at the traffic region) of each region plays an important part in our prediction of social vehicles' traces. And we can get the data according to the method as follows. We calculate the traffic absorptive rate of each functional area from the OD matrix of floating car, and then we get the traffic volume of the social vehicles in each functional area using these formulas:

$$SA_i = \sum_{j=1}^{N} SG_j \tag{1}$$

$$SG_j = \sum_{k=1}^{n} \alpha_k \times SR_k \tag{2}$$

where $SA_i$ represents the total number of social vehicles in functional area $i$ which is divided into $N$ girds. $SG_j$ represents the total number of social vehicles in grid $j$ which contains $n$ roads. $SR_k$ represents the number of floating cars in road $k$. The number of floating cars in road $k$ multiply with $\alpha_k$, which is the corresponding ratio of social vehicles and floating car on road $k$, to get the number of social vehicles in road $k$. Based on information provided in Beijing Traffic Development Annual report, every road has its specific value of $\alpha$, however, for few roads, the value of $\alpha$ was not defined in the documents, so we used the average value. And through (1) and (2), we get the traffic volume of social vehicles in each functional area. Then we use the traffic volume and the absorptive rate we have already obtained to calculate the traffic absorptive volume in each functional area. Now that we obtain the traffic volume and traffic absorptive volume of social vehicles in every functional area.

2) We use the data of floating car to get the value of parameters of the Gravity Model. Data are classified according to the characteristic of each trip, in other words, according to the type of functional area where the floating car arrive at and depart from. And this kind of characteristic reflects on the Gravity Model is to influence its parameter $\gamma$. Because the OD matrix of floating cars we obtained from the raw dataset is a sparse matrix, we choose unconstrained Gravity Model to solve the problem [33].

$$T_{ij} = \frac{\alpha(O_i \times D_j)^{\beta}}{R_{ij}^{\gamma}} \tag{3}$$

$$\ln T_{ij} = \ln \alpha + \beta \ln O_i D_j - \gamma \ln R_{ij} \tag{4}$$

where $T_{ij}$ represents the traffic volume from traffic region $i$ to $j$. $O_i$ and $D_j$ represents the traffic volume and the traffic absorptive volume of functional area $i$ and $j$ separately. $R_{ij}$ means the impedance from area $i$ to $j$, here we choose the distance as the traffic impedance. We put groups of data into the formula and calculate the value of three parameters in each group by the Generalized Least Squares(GLS) method.

3) We use Gravity Model to predict the OD matrix of social vehicles. According to the completed work, we get the traffic volume and traffic absorptive volume of the social vehicles in different functional areas, and the corresponding value of the Gravity Model's parameters in different groups. As pointed out by authors in [34] that gravity model does not hold for urban mobility. If the area is divided into smaller regions, it is possible to predict the trip generation but not the attraction unless we don't have reliable urban digital traces. In our case, we exploit the Beijing Traffic Development Annual reports and taxi GPS data to extract geo-located information such as point of interests (POIs) and number of trips ending in the specific smaller regions to model the attraction. Then we get the preliminary OD matrix of social vehicles by the unconstraint Gravity Model. Because we use the unconstraint Gravity Model, the results do not satisfy the constraint condition. Then we employ the Average Growth Factor Method[1] for iteration to calibrate it. From the theory of the Growth Factor Method, which is a model to predict short-term traffic distribution, there is a linear growth coefficient relationship between the preliminary OD matrix and the final OD matrix. So we can employ the Average Growth Factor Method for iteration to calibrate the results. In the Average Growth Factor Method, the growth rate of $T_{ij}$ is the Average of the growth rates of $O_i$ and $D_j$.

First, we get some parameters after the former calculation, including $T_{ij}, O_i^m, D_j^m$, plus estimates $O_i^* = (O_1^*, O_2^*, \cdots, O_m^*)$ of all future origins by zone and plus estimates $D_j^* = (D_1^*, D_2^*, \cdots, D_m^*)$ of all future destinations by zone.

Second, we make $m = 0$, and calculate $F_{O_i}^m$ (the growth rate of the traffic volume of functional area $i$) and $F_{D_j}^m$ (the traffic absorptive volume of functional area $j$) through (5) and (6).

$$F_{O_i}^m = \frac{O_i^*}{O_i^m} \tag{5}$$

$$F_{D_j}^m = \frac{D_j^*}{D_j^m} \tag{6}$$

Third, we calculate approximate value of $T_{ij}^{m+1}$ through (7) and (8). As there are many kinds of Growth Factor Function, we use the Average Growth Factor Method because of its feasibility.

$$T_{ij}^{m+1} = T_{ij}^m \times f(F_{O_i}^m, F_{D_j}^m) \tag{7}$$

$$f_{average}(F_{O_i}^{m+1}, F_{D_j}^{m+1}) = \frac{1}{2}(F_{O_i}^m, F_{D_j}^m) \tag{8}$$

where $f_{average}$ means the growth rate of the traffic volume from traffic region $i$ to $j$.

Fourth,we compute resulting errors on growth rates and determine whether the result is convergent through (9) and (10).

$$1 - \varepsilon < F_{O_i}^{m+1} = O_i^*/O_i^{m+1} < 1 + \varepsilon \tag{9}$$

$$1 - \varepsilon < F_{D_j}^{m+1} = D_j^*/D_j^{m+1} < 1 + \varepsilon \tag{10}$$

[1]Average Growth Factor Method http://facweb.knowlton.ohio-state.edu/pviton/courses2/crp5700/Avg-growth-beamer.pdf

Fifth, we make $m = m + 1$ and goes back to second step. We employ the Average Growth Factor Method for iteration to calibrate results, until the parameters of $F_{O_i}^m$ and $F_{D_j}^m$ are less than 3%.

## B. Network Description

We can download free city map data from OpenStreetMap (OSM) or other open source websites. OSM data can be uploaded by any users, so there are not only advantages but also disadvantages. The benefit is everyone can maintain and modify the map data. However, map data has deviations compared with the actual data due to its open attribute. In order to construct a precise road network for simulation, we have to correct the road topology so that it can match to the real world.

1) The number of lanes of each edge in OSM depends on its category, and this is different from the actualities in the majority of cases. The influence of these differences on the simulation results is relatively large, so we survey the number of lanes of the roads from some open data published by Beijing Transportation Bureau for comparison with the road network from OSM.
2) There are some redundant traffic movement restrictions on some segments which cause vehicles have to travel long detours. We find out these wrong restrictions by checking the intersections which are easy to arise congestion during the simulation against the Google Street View Map. These incorrect points will not affect the whole of street layout, but they will impact on the road choosing of simulation. So we select these points manually and correct them using Java OpenStreetMap Editor (JOSM).
3) Some correct turn restrictions are acting on superfluous road segments. These errors will cause higher traffic volume because vehicles may travel a longer path to their destinations, so we separate these roads by JOSM.
4) The topological information of complex intersections in OSM is unfit to be directly converted for simulation by NETCONVERT. When encountering some exceedingly complex intersections, NETCONVERT interprets the road topology as if several junctions co-existed. As a consequence, the vehicles from different directions cannot move straight through the junction and causes heavy jam during the simulation, because these vehicles have to make several lane changes and pass several traffic signals. The result is the impossibility for vehicles in traffic flows to merge at the intersection correctly. To fix such a problem, we add a circuit to join road segment links referring to the same physical intersection in Google Map.
5) To solve the problems mentioned above with the road network, we fixed more than 200 roads where the number of lines was not matching the actual road network. Also, removed ten roads that do not exist now. Besides, more than 5520 redundant traffic moments, correct restrictions, and traffic signals at junctions were fixed to make the road network topology analogous to the real world.

After preprocessing the data, we get road network data that match to the real world.

```
<tripdef   id="<ID>"   depart="<TIME>"   from="ORIGIN_EDGE_ID">
      to="DESTINATION_EDGE_ID"  fromtaz="<ORIGIN_DISTRICT_ID>"
      totaz="<DESTINATION_DISTRICT_ID>"
      [type="<VEHICLE_TYPE>"]   [PERIOD="<INT>" REPNO="<INT>"]
      [color="<COLOR>"]/>
```

(a)

```
    <vehicle            id="<ID>"            depart="<TIME>"
        departLane="ORIGIN_LANE_ID"
        arrivalLane="DESTINATION_LANE_ID"
        fromtaz="<ORIGIN_DISTRICT_ID>">
        totaz="<DESTINATION_DISTRICT_ID>"
    <route edges="<EDGE_ID>" color="<color>"/>/>
```

(b)

Fig. 3. File information descirbtion of SUMO.

## C. Simulation

After obtaining road network data and OD Matrix of social vehicles between different functional areas, we complete the necessary conditions for the simulation to generate the dataset. In order to generate the trajectories of social vehicles, we need the tool, SUMO.

Firstly, in order to get the distinct file for simulation, we need to know the edges that every region includes and the weight of each edge. To achieve this goal, we match image including information of functional areas to road network at first so that we can find the boundary of every functional area and mark down their coordinates based on OpenCV. In SUMO, every edge is composed of a group of lanes. Every edge has an edge ID, and every lane has a lane ID. Then we can get the ID and start point coordinate of each lane from the net file. By these coordinates of functional area boundary and lanes start point, we use the ray method to determine lanes within or outside the polygon. Finally, we get the edge ID in each region. For social functional regions, some are surrounded by other regions from the perspective of space. In this case, we need to remove the lanes of the small region inside it. In order to better reflect the real situation of traffic assignment, in other words, to improve the reality of the simulation results, we investigate road traffic index by the Beijing Municipal Commission of Transportation to get the weight of each road.

Secondly, we use OD2TRIPS to convert OD Matrix to the single trip of each vehicle. The trip information appears as illustrated in Fig. 3(a), where each trip has a unique ID. ‘depart’ represents departure time of the vehicle, which uses this trip for the first time. ‘from’ means the road from where the trip starts and ‘to’ is the road at where the trip ends. ‘fromtaz’ is the starting functional area and ‘totaz’ is the final area. ‘period’ is the interval time of vehicles use the same trip. ‘repno’ is the number of vehicles that share this trip. ‘color’ represents the color of vehicles.

Thirdly, after getting the trip information, we use DUAROUTER to get all the routes of vehicles, as shown in Fig. 3(b). SUMO assigns the weight of roads depending on its level of service, and vehicles will choose routes using Dijkstra algorithm in SUMO. Each route information contains two parts: one part is the vehicle information, and the other one is the route

TABLE I
FORMAT OF ORIGINAL DATASET

| Number of attributes | Attribute name | Notes |
|---|---|---|
| 1 | Taxi ID | The unique ID of taxi |
| 2 | Trigger event | 0 = get off, 1 = pick up |
| 3 | Operation state | 0 = vacant, 1 = carring |
| 4 | GPS time | DATE |
| 5 | GPS longitude | NUMERIC(10,7) |
| 6 | GPS latitude | NUMERIC(9,7) |
| 7 | GPS speed | INT |
| 8 | GPS direction | INT |
| 9 | GPS state | 0 = invalid, 1 = valid |

information. Vehicle ID represents the ID of the vehicle. 'depart' is the time to start the trip. 'departLane' represents the road information vehicle starting from, and 'arrivalLane' represents the final road information. 'fromtaz' is the starting traffic area and 'totaz' is the final traffic area.

After obtaining the routes of vehicles, we can complete simulation in SUMO at last. Through the output files of simulation, we can analyze some features of the generated dataset, such as the number of running, waiting, ended and inserted vehicles.

## IV. EXPERIMENTS

### A. Original dataset

The taxi trace data we used comes from 12,000 taxis in Beijing.[2] These taxis upload their location information by GPS devises at a frequency of 11 seconds. We conduct experiments using the dataset over a period of 30 days (from November 1,2012 to November 30,2012) and within the fifth circle highway of Beijing over 839.28 km$^2$ ( [116.1970 °E,116.5425 °E] and [39.7775 °N, 40.0335 °N] ), which is 15 GB before compressed and includes hundreds of millions of records. Original taxi GPS data are stored in a text document named after the storage time. These documents contain taxi GPS information. The concrete format of the dataset is shown in Table I.

Since all the records in the raw dataset are not correct, we deal with the data in advance. Firstly, we delete the record whose GPS state is '0', which means GPS devices do not work. Secondly, we delete the duplicate items. Then we put all the items with the same ID in one file and list them in chronological order. Therefore, we get the track of each taxi in one day. By gathering the OD information of each trip, we can obtain the OD matrix of taxis traveling. Besides, we delete the trip whose duration or distance is too short to occur in reality. Finally, we analyze some features of the trips we used in this paper, as shown in Fig. 4. These features include the travel distance and duration. The results indicate that majority of trips of floating cars are with a shorter period and distance.

### B. Mobiity Dataset Generation

Because urban functional areas affect residents' mobility pattern largely, we leverage functional areas to divide Beijing and

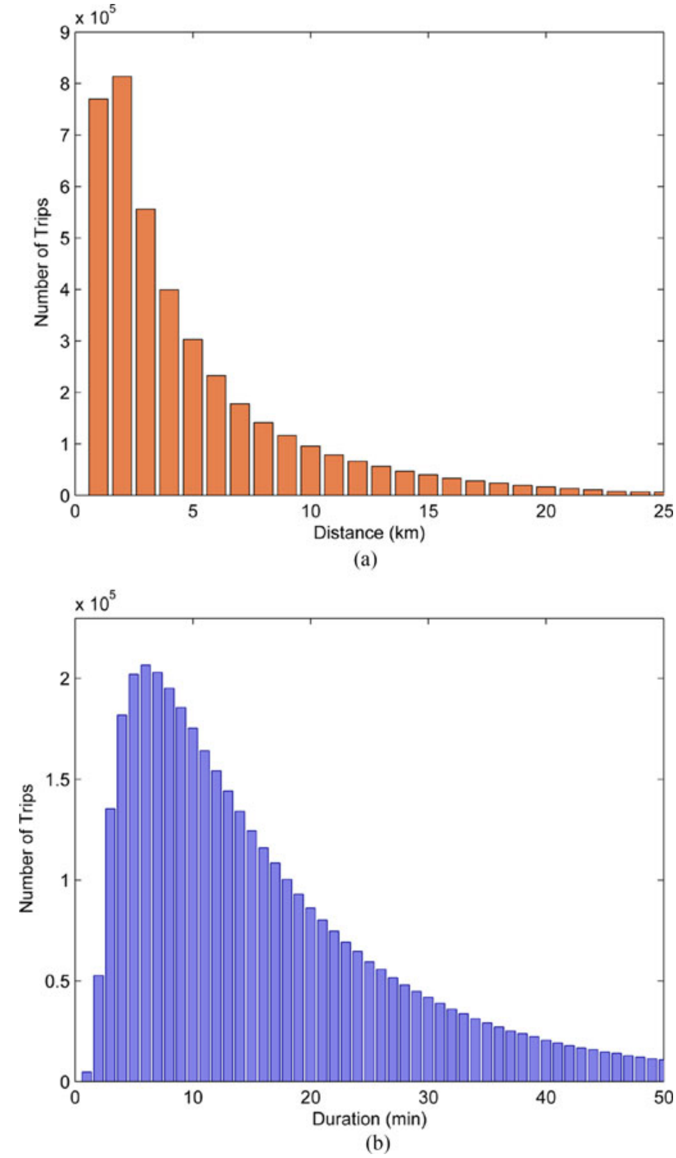



Fig. 4. Features of original dataset.

view every functional area as traffic region. These areas include Diplomatic/Embassy area, Science/Education/Technology area, Emerging Commercial/Entertainment area, Developed Residential area, Emerging Residential area, Old Neighborhood area, Developed Commercial/Entertainment area, Historical Interests/Parks area and Nature area. Then we use the unconstrained Gravity Model to predict OD matrix of social vehicles according to OD matrix of floating cars and official report in Beijing.[3] Table II shows the parameter $\gamma$ values of a part of groups.

Road network data from OSM are vector data and contain every road with a unique ID. In order to get all the road IDs of different functional regions, we match image including function area to the road network and find the data of these roads based on mapping relation through ArcGIS. We import WGS-84 coordinate into the image and rectify it based on the real map

[2]Source of taxi trace data, http://www.datatang.com/data/44502

[3]Beijing Traffic Development Annual Report, http://www.bjtrc.org.cn

TABLE II
PARAMETER VALUES OF GRAVITY MODEL

| Original functional area | Terminal functional area | The value of $\gamma$ |
|---|---|---|
| Science/Education/Technology area | Nature area | 1.73 |
| Diplomatic/Embassy area | Science/Education/Technology area | 0.33 |
| Old Neighborhood area | Commercial/Entertainment area | 0.99 |
| Commercial/Entertainment area | Residential area | 0.73 |

TABLE III
FACTOR OF MAIN STREETS

| Street name | Social vehicle ratio | Taxi ratio | Growth factor |
|---|---|---|---|
| East fifth ring road | 59.52% | 4.78% | 12.45 |
| South fifth ring road | 37.24% | 0.58% | 64.21 |
| West fifth ring road | 68.50% | 3.26% | 21.01 |
| North fifth ring road | 59.97% | 4.52% | 13.27 |
| East forth ring road | 65.74% | 12.44% | 5.28 |
| South forth ring road | 71.41% | 7.69% | 9.29 |
| West forth ring road | 72.44% | 10.84% | 6.68 |
| North forth ring road | 60.50% | 18.45% | 3.28 |
| East third ring road | 55.38% | 20.78% | 2.67 |
| South third ring road | 57.88% | 15.72% | 3.68 |
| West third ring road | 62.19% | 16.62% | 3.74 |
| North third ring road | 59.03% | 20.10% | 2.94 |
| East second ring road | 65.92% | 19.60% | 3.36 |
| South second ring road | 52.53% | 10.88% | 4.83 |
| West second ring road | 63.68% | 20.02% | 3.18 |
| North second ring road | 69.63% | 18.99% | 3.67 |
| Changan Avenue | 65.53% | 17.15% | 3.82 |
| Liangguang Avenue | 45.50% | 14.86% | 3.06 |
| Qianmen West street | 50.36% | 20.75% | 2.43 |

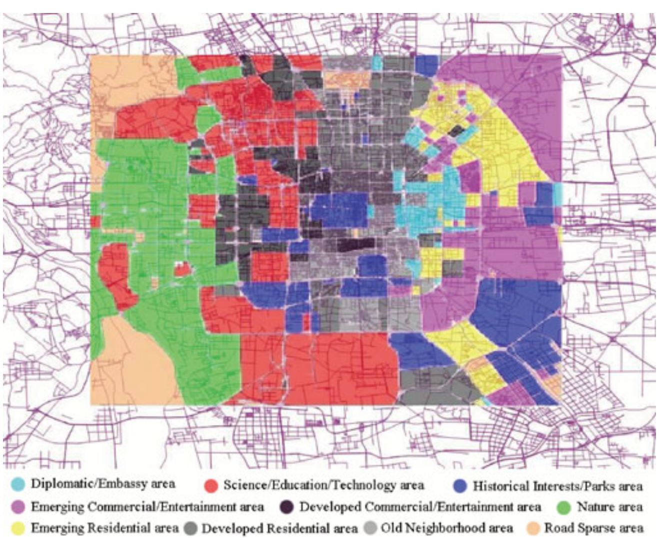



Fig. 5. Matching of road data and urban functional areas.

until overlapping between them completely. Hence, we get the corresponding map in Fig. 5. Purple lines represent the concrete road in Beijing, and different colors illustrate different kinds of functional areas. From Fig. 5, Beijing is divided into small traffic regions. Some regions are larger like Nature areas on the left, and some are smaller like Diplomatic/Embassy areas on the right.

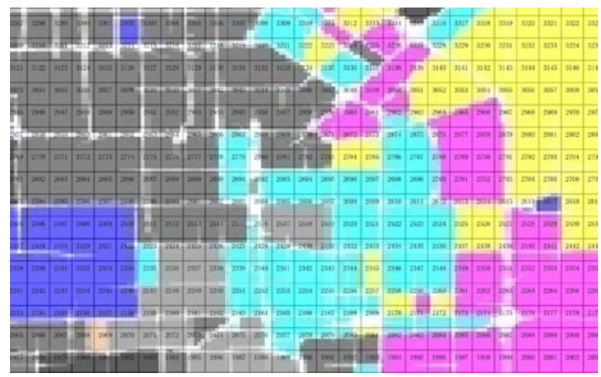


Fig. 6. Division by longitude and latitude.

Functional areas with different sizes are not conducive to our next work, so we further divide them into squares according to the latitude and longitude with 0.05 range, and mark each square with a fixed ID as shown in Fig. 6. The squares which are in the intersections of functional areas are assigned to the largest occupied one.

Based on official statistics, we can get OD matrix of passengers in taxis between different traffic regions. According to Beijing Traffic Development Annual Report, we get the following data: In 2012, there are 4190000 social vehicles and 66000 taxis. Social vehicles almost travel 9090000 times every day, and taxis travel 1990000 times every day on average according to the traffic report in Beijing 2013. Normally, the ratio of social vehicles traveling times and taxis traveling times is 4.57. Besides, we get more other data shown in Table III and Table IV. The statistics data are most on the main roads since it is impossible to acquire the ratio data in all roads, therefore as for the streets which cannot acquire its ratio data, we use the normal or average value.

Road data we used are from OSM[4], however, there were some errors in the data. As shown in Fig. 7, yellow vehicles can turn left and go straight through the intersection, however, cars cannot turn left here, and we need to modify it using JOSM, which is an open source tool to modify road data from OSM. Moreover, we want to simulate social vehicles travelling in Beijing rather than all the transportation tools. So we delete railways, sidewalks and so on. As shown in Fig. 8(a), we edit the map data using JOSM. After deleting the irrelevant data including railway data at the bottom right corner, comparing the map data with the real world and modifying it, the final road network topology is shown in Fig. 8(b).

[4]OpenStreetMap, http://www.openstreetmap.org

TABLE IV
FACTORS OF FUNCTIONAL REGIONS

| Number of functional regions | Roads inside | Factors |
|---|---|---|
| 0 | East third ring road, East second ring road, North second ring road | 5.25 |
| 1 | East third ring road | 2.67 |
| 2 | East second ring road, East second ring road, South second ring road | 4.10 |
| 3 | East fifth ring road | 12.45 |
| 4 | North forth ring road | 3.28 |
| 5 | North third ring road | 2.94 |
| 6 | West fifth ring road, West forth ring road | 13.85 |
| 7 | No main roads | 4.57 |
| 8 | East fifth ring road, North fifth ring road | 12.86 |
| 9 | East forth ring road, North forth ring road, East third ring road | 3.74 |

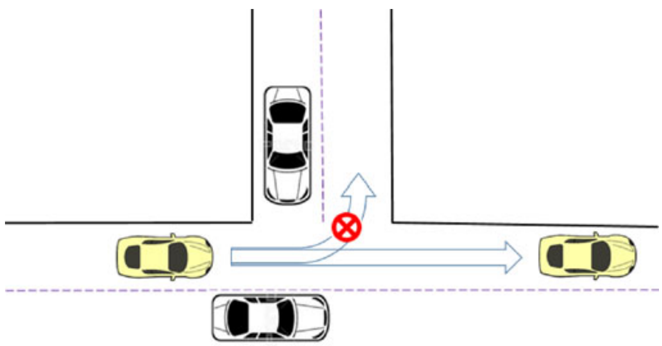
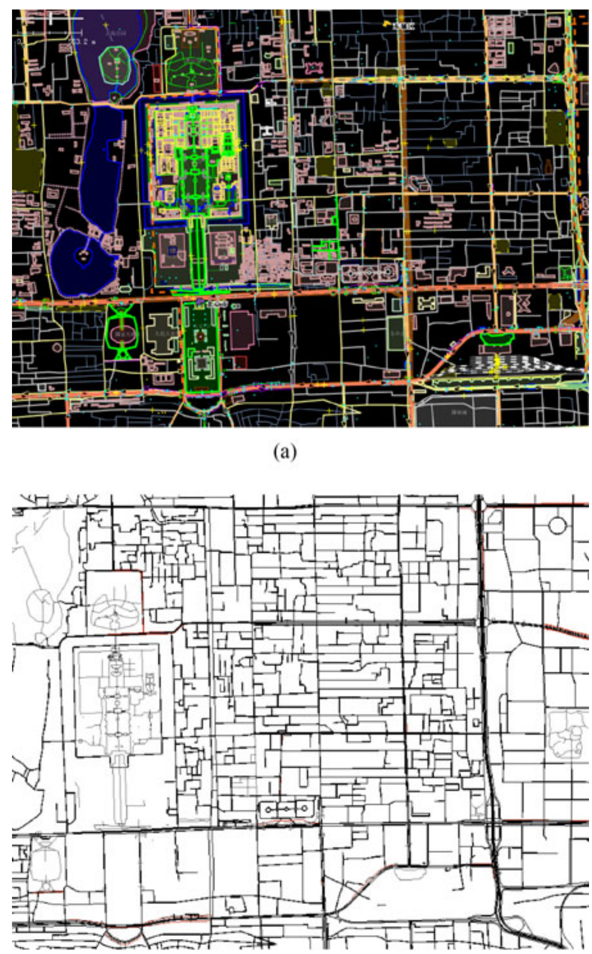


Fig. 7. Errors in road data.



(a)

(b)

Fig. 8. Modification of the road network data. (a) Open by JOSM. (b) Edited by SUMO.

### C. Experiment Results and Analysis

We generate mobility dataset of social vehicles traveling, and find some interesting problems. Fig. 9 shows road speed of vehicles, and different roads have their max vehicle speeds. From the view of road network composition, roads within the fifth circle highway are distributed densely and form complex road network system. Lower speed limited on these roads leads to traffic jams. Besides, the speed of provincial road, which is restricted greatly.

Then we simulate traffic condition within the fifth circle highway in Beijing with OD matrix of social vehicles traveling and road network data of different functional areas. We generate social vehicles trajectories at 8:00–9:00, 12:00–13:00, 17:00–18:00 and 22:00–23:00. These four time periods represent morning peak traffic, rush hour traffic, evening peak traffic, and night traffic respectively. Then we get routes, and each item is composed of vehicle ID, departure region, destination region, departure lane, destination lane, departure time and vehicle speed.

Fig. 10 which is obtained employing the validation mechanism proposed in [31], shows the comparison between real traffic condition and simulation result from 12:00 to 13:00 in Beijing. Fig. 10(a) is real-time road conditions information from Beijing Traffic Management Bureau, and Fig. 10(b) is the visualization of our simulation. Fig. 10(c) shows the route length as a function of the travel time, as obtained in simulation and from Google Maps. Some of the cars are restricted to traffic jams in the simulation, causing the travel time and route length abnormal, which do not occur in the Google Maps prediction. That explains the limited dispersion in the scatterplot. Observing the results, there are 26747 dots of SUMO data in the figure and about 2176 dots are outlying. In another word, the percentage of outlying behaviors (obvious abnormal data) is 8.13%. From results it is clear that scenarios are the same for most of the cases. The traffic conditions in Fig. 10(a) and 10(b) are almost unanimous in general and efforts are made to make the traffic mobility analogous to real world in terms of trip generated from

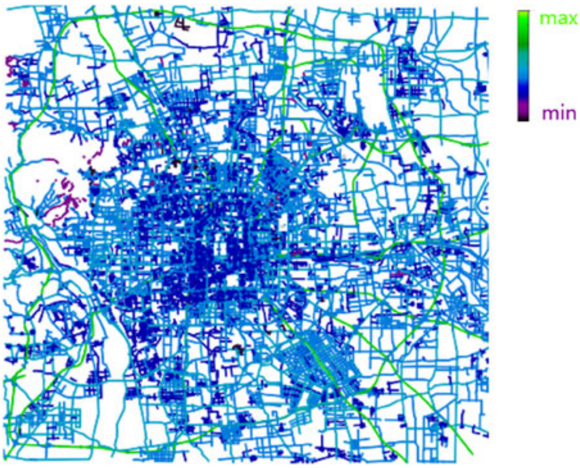



Fig. 9. Max speed of motorway in Beijing.

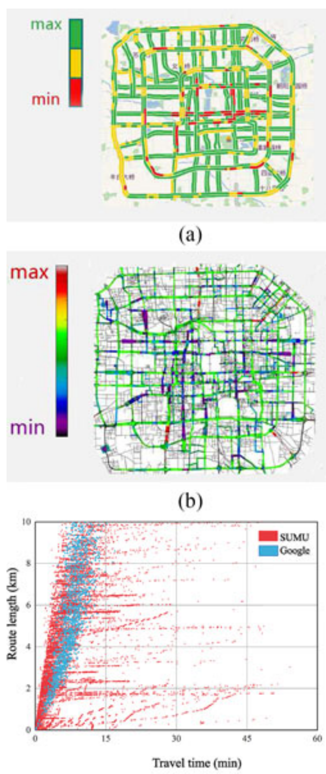



Fig. 10. Traffic condition comparison. (a) Real traffic condition in Beijing on 13:00. (b) Simulation results in Beijing on 13:00. (c) Comparison of travel time and distance.

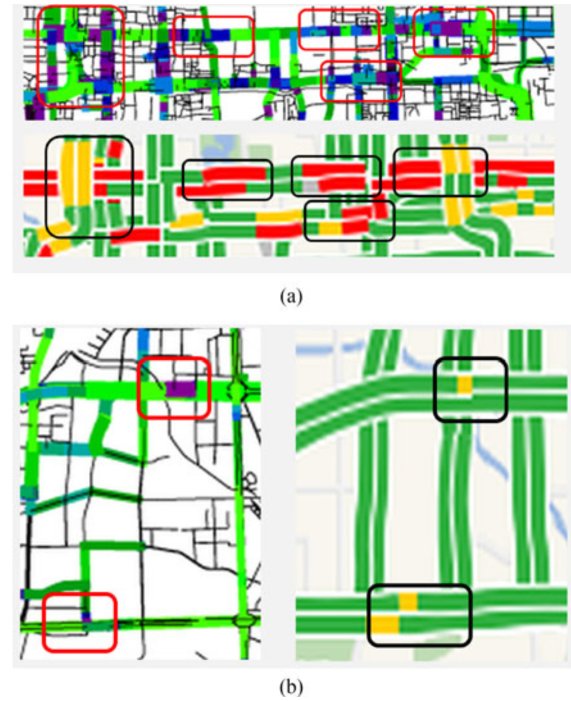



Fig. 11. Details between real traffic and simulation results. (a) Qianmen street and Changan Avenue traffic condition. (b) South third ring road and South forth ring road.

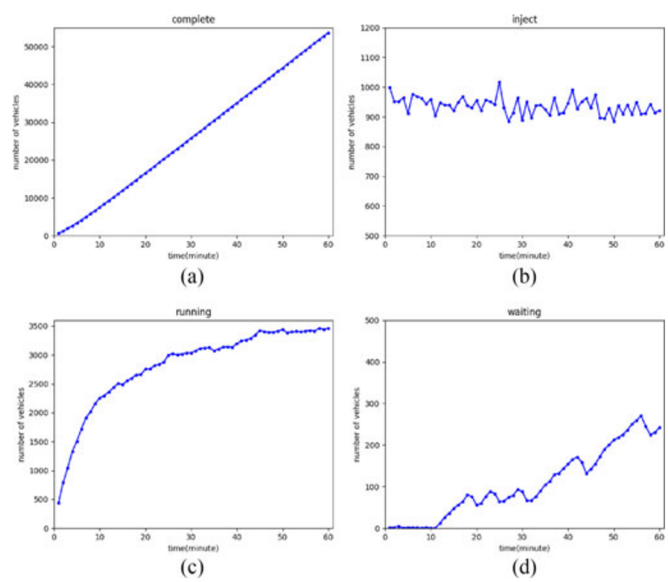



Fig. 12. The description of simulation traces data from 12:00 to 13:00.

and drawn to a specific location. Also, changes were made to make the road network topology analogous to real world scenario to avoid road congestions. Fig. 11(a) shows the congestion areas on Qianmen street and Changan Avenue. Fig. 11(b) demonstrates the open way in South third circle highway and South fourth circle highway.

Fig. 12 illustrates a more specific data description about the simulation result from 12:00 to 13:00. According to Fig. 12(c), at the beginning of the simulation, some vehicles haven't injected into the simulation scenario, so the curve begins with 0 and gradually increase. Once it comes to saturation, the number of vehicles maintain a relatively stable state. At the same time, according to Fig. 12(d), some vehicles come into the traffic jam, which conforms to the actual situation: the noon peak of Beijing's transportation from 12:00 to 13:00. But the downward trend that the curve presents sometimes suggests that traffic jams can be eased. Thus it can be seen that our simulation result are coordinate with the Beijing's real transportation, and our method can be used to generate social vehicle's mobility dataset with high accuracy and reliability.

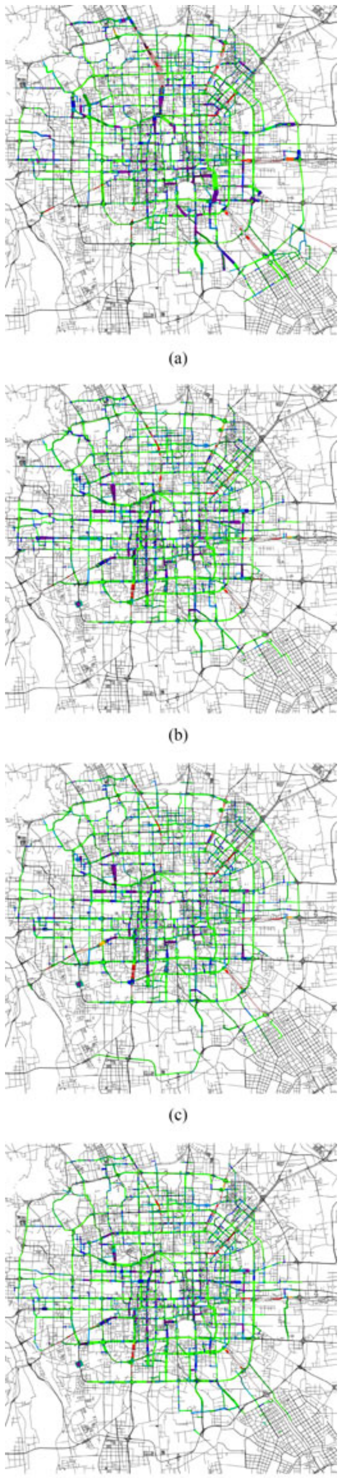



Fig. 13. Traffic condition comparison in different time periods. (a) 8:00–9:00. (b) 12:00–13:00. (c) 17:00–18:00. (d) 22:00–23:00.

Fig. 13 is the visualization of traffic simulation results in the four intervals. These figures show the traffic congestion conditions within the fifth circle highway. We find some interesting results: 1) In the four periods, different levels of congestion happened at the intersection. 2) Due to the speed and vehicle amount, external traffic zone is better than internal traffic in the city. 3) Compared with Fig. 9, vehicles on highways have faster speeds than others. 4) Although vehicle amount at night is less than that at day, there exist congestions in some roads during 22:00 to 23:00 as is shown in Fig. 13, which may be caused by huge vehicle amount. 5) Traffic in west and south is in a better condition than that in East and North through comparisons between four figures.

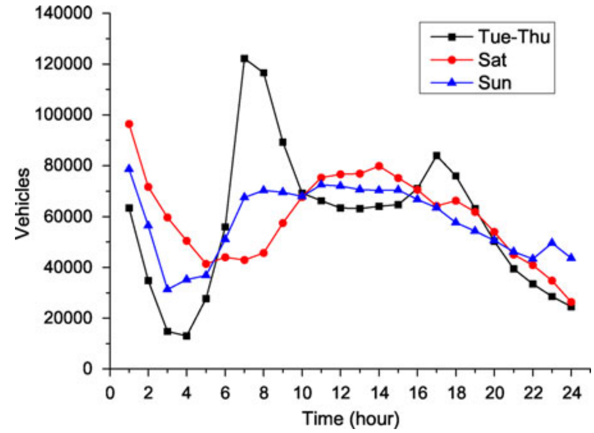



Fig. 14. Summary of trip information of 72 hours.

Compared with Fig. 5, we contrast traffic conditions in different time periods and found some interesting facts as following. For residential regions, traffic is always in a bad condition during the four periods because a large number of vehicles are travelling on the road in morning peak traffic, rush hour traffic, and evening peak traffic. Traffic condition in the commercial region is better in morning and noon than that at night. This phenomenon may be caused by recreational activities provided by shopping centers and recreation centers, because people prefer to relax after work. For sparse regions, small population and small amount of vehicles make the traffic fluent. In addition, vehicles converge upon school districts, which leads to traffic jam in morning peak traffic and evening peak traffic to education and technology function based regions.

Through analyzing traffic condition in one week in detail, we can divide it into five parts: traffic on Monday, traffic from Tuesday to Thursday, traffic on Friday, traffic on Saturday and traffic on Sunday. Their traffic conditions are different separately from each other. Here we choose traffic on Saturday, traffic on Sunday and traffic from Tuesday to Thursday for simulation. We generate social vehicle tracks within the fifth circle highway in Beijing of 72 hours. The statistics information is as shown in Fig. 14. It can be observed that: 1) Amount of vehicles rises in the morning on Saturday, which is more than Sunday and weekday, 2) Compared with Sunday, vehicle number grows more quickly than weekends because residents begin to work, 3) Vehicle number at weekends is less because people prefer to rest at home.

## V. Conclusion

The lack of social vehicle tracks impedes the research of social attributes between vehicles in VSN. To solve this problem, we propose the mobility dataset generation method of social vehicles traveling based on floating car data and social functional areas. Our method mainly includes network demand, demand description, and simulation of social vehicle trajectories with SUMO, and it can be used to construct traffic scenario for simulation of VSNs. In the experiments, we use SUMO to generate social vehicle tracks in four time periods and three days inside fifth ring road in Beijing. Compared with real-time traffic condition published by Beijing Traffic Management Bureau, our results are almost consistent, which shows our method has good accuracy and efficiency.

In our method, vehicles choose a route based on simulation tool, which may lead to a traffic jam in some roads. In the future, we will analyze all the roads in the whole city, and assign a specific value to every road. Then vehicles can choose their roads according to these values to make our dataset generation method better. Besides, the trip attraction of each region needs more realistic measurements to model efficiently model the O-D matrix. Also, mobility pattern of social vehicles may differ in few areas, such as train stations or other crowded public spots. The mobility model can be further enhanced by considering the data from parking areas in these locations, which is not available at the moment. However, we intend to get and analyze such data in our future work.

## References

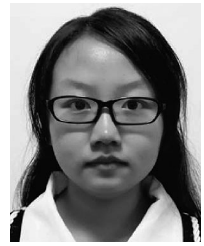
[1] G. Xiong *et al.*, "Cyber-physical-social system in intelligent transportation," *IEEE/CAA J. Automat. Sin.*, vol. 2, no. 3, pp. 320–333, Jul. 2015.
[2] Z. Wang, J. Liao, Q. Cao, H. Qi, and Z. Wang, "Friendbook: A semantic-based friend recommendation system for social networks," *IEEE Trans. Mobile Comput.*, vol. 14, no. 3, pp. 538–551, Mar. 2015.
[3] J. Lloret, A. Canovas, A. Catalá, and M. Garcia, "Group-based protocol and mobility model for vanets to offer internet access," *J. Netw. Comput. Appl.*, vol. 36, no. 3, pp. 1027–1038, 2013.
[4] J. Wan, D. Zhang, S. Zhao, L. Yang, and J. Lloret, "Context-aware vehicular cyber-physical systems with cloud support: architecture, challenges, and solutions," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 106–113, Aug. 2014.
[5] X. Hu, T. H. Chu, V. C. Leung, E. C.-H. Ngai, P. Kruchten, and H. C. Chan, "A survey on mobile social networks: Applications, platforms, system architectures, and future research directions," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1557–1581, Oct.–Dec. 2015.
[6] R. Yu, J. Kang, X. Huang, S. Xie, Y. Zhang, and S. Gjessing, "Mixgroup: Accumulative pseudonym exchanging for location privacy enhancement in vehicular social networks," *IEEE Trans. Dependable Secure Comput.*, vol. 13, no. 1, pp. 93–105, Jan./Feb. 2016.
[7] F. Xia, H. B. Liaqat, J. Deng, J. Wan, and S. Das, "Overhead control with reliable transmission of popular packets in ad-hoc social networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 9, pp. 7647–7661, Sep. 2016.
[8] X. Zheng *et al.*, "Big data for social transportation," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 3, pp. 620–630, Mar. 2016.
[9] Z. Ning, F. Xia, N. Ullah, X. Kong, and X. Hu, "Vehicular social networks: Enabling smart mobility," *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 46–55, May 2017.
[10] Q. Yang and H. Wang, "Toward trustworthy vehicular social networks," *IEEE Commun. Mag.*, vol. 53, no. 8, pp. 42–47, Aug. 2015.
[11] A. M. Vegni and V. Loscri, "A survey on vehicular social networks," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2397–2419, Oct.–Dec. 2015.
[12] T. Luan, R. Lu, X. Shen, and F. Bai, "Social on the road: Enabling secure and efficient social networking on highways," *IEEE Wireless Commun.*, vol. 22, no. 1, pp. 44–51, Feb. 2015.
[13] T. H. Luan, X. Shen, F. Bai, and L. Sun, "Feel bored? Join verse! Engineering vehicular proximity social networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 3, pp. 1120–1131, Mar. 2015.
[14] J. Xu, D. Deng, U. Demiryurek, C. Shahabi, and M. van der Schaar, "Mining the situation: Spatiotemporal traffic prediction with big data," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 4, pp. 702–715, Jun. 2015.
[15] X. Kong, Z. Xu, G. Shen, J. Wang, Q. Yang, and B. Zhang, "Urban traffic congestion estimation and prediction based on floating car trajectory data," *Future Gener. Comput. Syst.*, vol. 61, pp. 97–107, 2016.
[16] K. Chong and H. Sung, "Prediction of road safety using road/traffic big data," in *Proc. Int. Conf. Semantic Web Bus. Innov.*, 2015, pp. 23–27.
[17] N. J. Yuan, Y. Zheng, X. Xie, Y. Wang, K. Zheng, and H. Xiong, "Discovering urban functional zones using latent activity trajectories," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 3, pp. 712–725, Mar. 2015.
[18] X. Kong, F. Xia, J. Wang, A. Rahim, and S. K. Das, "Time-location-relationship combined service recommendation based on taxi trajectory data," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3, pp. 1202–1212, Jun. 2017, doi:10.1109/TII.2017.2684163.
[19] M. Behrisch, L. Bieker, J. Erdmann, and D. Krajzewicz, "Sumo—Simulation of urban mobility: An overview," in *Proc. 3rd Int. Conf. Adv. Syst. Simul.*, Barcelona, Spain, 2011, pp. 63–68.
[20] S. Krauss, P. Wagner, and C. Gawron, "Metastable states in a microscopic model of traffic flow," *Phys. Rev. E*, vol. 55, no. 5, pp. 55–97, 1997.
[21] D. Krajzewicz, G. Hertkorn, C. Rössel, and P. Wagner, "Sumo (simulation of urban mobility)—An open-source traffic simulation," in *Proc. 4th Middle East Symp. Simul. Model.*, 2002, pp. 183–187.
[22] J. Härri, F. Filali, and C. Bonnet, "Mobility models for vehicular ad hoc networks: A survey and taxonomy," *IEEE Commun. Surveys Tuts.*, vol. 11, no. 4, pp. 19–41, Oct.–Dec. 2009.
[23] S. Uppoor, O. Trullols-Cruces, M. Fiore, and J. M. Barcelo-Ordinas, "Generation and analysis of a large-scale urban vehicular mobility dataset," *IEEE Trans. Mobile Comput.*, vol. 13, no. 5, pp. 1061–1075, May 2014.
[24] Y. Pigné, G. Danoy, and P. Bouvry, "A vehicular mobility model based on real traffic counting data," in *Communication Technologies for Vehicles*. New York, NY, USA: Springer, 2011, pp. 131–142.
[25] M. Ferreira, H. Conceição, R. Fernandes, and O. K. Tonguz, "Stereoscopic aerial photography: An alternative to model-based urban mobility approaches," in *Proc. 6th ACM Int. Workshop Veh. Internetw.*, 2009, pp. 53–62.
[26] G. Thakurzx, P. Huiz, and A. Helmyx, "Modeling and characterization of urban vehicular mobility using web cameras," in *Proc. IEEE INFOCOM Workshops*, 2012, pp. 262–267.
[27] B. Raney *et al.*, "An agent-based microsimulation model of Swiss travel: First results," *Netw. Spatial Econ.*, vol. 3, no. 1, pp. 23–41, 2003.
[28] N. Cetin, A. Burri, and K. Nagel, "A large-scale multi-agent traffic microsimulation based on queue model," in *Proc. 3rd Swiss Transp. Res. Conf.*, 2003.
[29] L. Codeca, R. Frank, and T. Engel, "Luxembourg sumo traffic (lust) scenario: 24 hours of mobility for vehicular networking research," in *Proc. 2015 IEEE Veh. Netw. Conf.*, 2015, pp. 1–8.
[30] M. Gramaglia, O. Trullols-Cruces, D. Naboulsi, M. Fiore, and M. Calderon, "Mobility and connectivity in highway vehicular networks: A case study in madrid," *Comput. Commun.*, vol. 78, pp. 28–44, 2016.
[31] L. Bedogni, M. Gramaglia, A. Vesco, M. Fiore, J. Härri, and F. Ferrero, "The Bologna Ringway dataset: Improving road network conversion in sumo and validating urban mobility via navigation services," *IEEE Trans. Veh. Technol.*, vol. 64, no. 12, pp. 5464–5476, Dec. 2015.
[32] J. Yuan, Y. Zheng, and X. Xie, "Discovering regions of different functions in a city using human mobility and POIs," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2012, pp. 186–194.
[33] A. M. Voorhees, "A general theory of traffic movement," *Transportation*, vol. 40, no. 6, pp. 1105–1116, 2013.
[34] Y. Yang, C. Herrera, N. Eagle, and M. Gonzalez, "A multi-scale multi-cultural study of commuting patterns incorporating digital traces," in *Proc. NetMob*, 2013, pp. 1–3.

**Xiangjie Kong** (M'13–SM'17) received the B.Sc. and Ph.D. degrees from Zhejiang University, Hangzhou, China. He is currently an Associate Professor with the School of Software, Dalian University of Technology, Dalian, China. He has published more than 50 scientific papers in international journals and conferences (with 30+ indexed by ISI SCIE). His research interests include intelligent transportation systems, mobile computing, and cyber-physical systems. He is a Senior Member of CCF and a Member of ACM. He has served as a (Guest) Editor of several international journals, the Workshop Chair, or a PC Member of a number of conferences.

**Feng Xia** (M'07–SM'12) received the B.Sc. and Ph.D. degrees from Zhejiang University, Hangzhou, China. He was a Research Fellow with Queensland University of Technology, Brisbane, QLD, Australia. He is currently a Full Professor with the School of Software, Dalian University of Technology, Dalian, China. He has authored or coauthored 2 books and more than 200 scientific papers in international journals and conferences. His research interests include computational social science, network science, data science, and mobile social networks. He is a Senior Member of ACM and a Member of AAAS. He is a (Guest) Editor of several international journals. He serves as the General Chair, the PC Chair, the Workshop Chair, or the Publicity Chair of a number of conferences.

**Zhaolong Ning** (M'14) received the M.S. and Ph.D. degrees from Northeastern University, Shenyang, China. He was a Research Fellow with Kyushu University, Fukuoka, Japan. He is currently an Associate Professor with the School of Software, Dalian University of Technology, Dalian, China. He has authored or coauthored more than 70 scientific papers in international journals and conferences. His research interests include social computing, network optimization, and big scholar data.

**Azizur Rahim** received the B.Sc. degree from the University of Engineering and Technology, Peshawar, Pakistan, and the M.S. degree in electrical engineering from COMSATS, Islamabad, Pakistan. He is currently working toward the Ph.D. degree, under the supervision of Prof. F. Xia, at The Alpha Lab, School of Software, Dalian University of Technology, Dalian, China. His research interests include mobile and social computing, ad-hoc networks, VANETs, mobile social networks, and vehicular social networks.



**Yinqiong Cai** is currently working toward the B.S. degree at Dalian University of Technology, Dalian, China. She is currently a Senior Student, also a Research Assistant, in The Alpha Lab, School of Software, Dalian University of Technology. Her research interests include dataset generation, vehicular social network, and data mining.

**Zhiqiang Gao** received the Bachelor's and the Master's degree, both in software engineering, from Dalian University of Technology, Dalian, China, in 2014 and 2016, respectively. He is currently with UnionPay International, Shanghai, China. His research interests include mobile social networks, big traffic data, and socially aware networking.

**Jianhua Ma** received the B.S. and M.S. degrees from the National University of Defense Technology, Changsha, China, in 1982 and 1985, respectively, and the Ph.D. degree from Xidian University, Xi'an, China, in 1990. He is currently a Professor with the Faculty of Computer and Information Sciences, Hosei University, Tokyo, Japan. He has authored or coauthored more than 200 papers and edited more than 20 books/proceedings and more than 20 journal special issues. His research interests include multimedia, networks, ubiquitous computing, social computing, and cyberintelligence.