Modeling and Analysis of Large-Scale Urban Mobility for Green Transportation

Feng Xia, *Senior Member, IEEE*, Azizur Rahim, Xiangjie Kong, *Senior Member, IEEE*, Meng Wang, Yinqiong Cai, and Jinzhong Wang

***Abstract*—Context-aware applications of vehicular social networks (VSNs) play a significant role to achieve the goal of green transportation by sharing driving experience to reduce gasoline consumption. One of the main challenges is to evaluate the performance of these applications, which relies on the underlying VSN mobility model. In this paper, we investigate big urban traffic data to characterize essential features of urban mobility and construct large-scale green urban mobility models. We exploit the road and traffic information to enhance the trip generation algorithm and traffic assignment technique based on the weighted segments of roads. Besides, we perform extensive observations and corrections on the OpenStreetMap imported to simulation of urban mobility to make it analogous to real-world road topology. The experimental results and validation process show that the generated mobility models reveal realistic behavior required for analysis of context-aware applications of VSNs for green transportation systems.**

***Index Terms*—Dataset generation, green transportation systems, mobility modeling, vehicular social networks (VSNs).**

I. Introduction

According to the Shanghai Urban and Rural Construction and Transportation Development Research Institute annual report, road traffic has shown an average annual growth about 2%.[1] Road traffic is one of the major causes of environmental pollution and hazes in many cities around the globe. The research community has been attracted by energy-efficient protocols and systems to enhance transportation efficiency and achieve the goal of smart and green cities. The drivers and passengers may share their road experience and information to reduce fuel consumption and air pollution. Vehicular social networks (VSNs) are built on the top of vehicular *ad hoc* networks enabling commuters to interact and share the information with neighboring nodes (vehicles, passengers, and drivers) [1]. Rapid advancement in communication technologies enables these devices to interact and share information using processing and communication capabilities [2]. Liu *et al.* [3] presented carbon recorder, a mobile social application not only to encourage drivers to improve their behavior but also provide social awareness and a platform for data collection that can be exploited by the research community and traffic authorities for green transportation systems. Chen *et al.* [4] presented a crowdsourcing-based application for smart parking in smart cities to avoid idle cruising. The smart parking application integrates crowdsourcing with the traditional road navigation system to collect and share real-time information about parking availability, which can be used to reduce energy consumption. Kong *et al.* [5] presented a taxi recommendation system to overcome the problem of idle cruising and enhance fuel consumption. Similarly, NaviTweet [6] is a centralized socially aware interactive navigation system allowing users to provide information to the onboard vehicular navigation system. NaviTweet allows drivers to calculate a personalized route and form a VSN group with other commuters on the same path to enhance driving experience in terms of fuel consumption and time. Every user of the group can send voice tweets, and the NaviTweet server periodically aggregates these tweets into a tweet digest and send to the members of the specific VSN group. Similarly, Hu *et al.* [7] presented the concept of a social drive to encourage the user to improve their driving behaviors in terms of fuel consumption. A social drive provides the integration of onboard devices to social networks to share users' trip information and experience.

Inheritance of social network theories into the vehicular environment and human behaviors, which socialize the communication of vehicles on roads has emerged in the form of VSNs. Compared to other *ad hoc* networks and delay tolerant networks, the mobility of nodes is highly dynamic and restricted to roads in VSNs. The vehicular mobility is highly influenced by a number of factors including traffic regulation, city development and planning, working hours/days, work and lifestyle of inhabitants, social values, road blocks, and so on. Validation

Manuscript received July 1, 2017; revised October 10, 2017 and November 25, 2017; accepted December 10, 2017. Date of publication December 20, 2017; date of current version April 3, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61572106, in part by the Natural Science Foundation of Liaoning Province, China, under Grant 201602154, and in part by the Dalian Science and Technology Planning Project under Grant 2015A11GX015 and Grant 2015R054. Paper no. TII-17-1424. *(Corresponding author: Xiangjie Kong.)*

F. Xia, A. Rahim, X. Kong, M. Wang, and Y. Cai are with the Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, School of Software, Dalian University of Technology, Dalian 116620, China (e-mail: f.xia@ieee.org; aziz.buner@gmail.com; xjkong@ieee.org; meng_wang2016@outlook.com; yinqiongcai@outlook.com).

J. Wang is with the Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, School of Software, Dalian University of Technology, Dalian 116620, China, and also with Shenyang Sport University, Shenyang 110102, China (e-mail: wwjz432@126.com).

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/TII.2017.2785383

[1]http://www.shsz.org.cn/news/20173/i-2017331154044.html

and evaluation of different proposed protocols for VSNs rely on simulation. The value of the validation and the credibility of results are, therefore, highly dependent on the underlying mobility model used for simulation. Unfortunately, the simulation performance evaluation of context-aware application for VSNs are often biased by the underlying mobility models. An incorrect representation of vehicular mobility can lead to misleading results and conclusions, even if a flawless network-level simulator is used for performance evaluation. The results obtained will be considered more accurate and credible if and only if the underlying mobility model is analogous to the real-world car traffic. As an instance, Baumann *et al.* [8] analyzed the performance of routing protocols for different vehicular mobility models and concluded that nonrealistic mobility models may lead to misleading conclusions or overestimate the performance of routing protocols in the vehicular environment. In the existing work, it is assumed that a synthetic mobility model will be more realistic if it is based on behavioral models [9]. However, most widely used models are based on simplistic stochastic model, random models, small scale, or based on limited real-time vehicular mobility traces. Recently, Vegni and Loscri [10] presented a literature review on VSNs considering the social features and aspects for a diverse range of applications of next generation vehicles. Human behavior and social characteristics largely impact on the mobility of vehicles in such networks. The resulting mobility patterns of vehicles can dramatically affect the performance and behavior of different proposed applications and protocols [11]. These observations have led the research community to make substantial progress in the quality of mobility modeling for vehicular networks over the last few years.

In VSNs, vehicles on the roads opportunistically communicate with each other and may encounter other vehicles/commuters with a similar interest profile. These commuters may also interact other commuters facing similar traffic condition, environmental factors, the same mobility pattern, or belong to the same community. This inherent similarity of commuters could impact the mobility modeling in VSNs. Human mobility shows a high degree of spatial and temporal regularity that can be represented by a single probability distribution [12]. In this paper, we go a step further to infer the important macroscopic features of mobility from existing mobility traces and generate mobility models that can be used for performance evaluation of context-aware applications of VSNs. The contributions of our work presented in this paper can be summarized as follows.

1) Large-scale mobility traces are explored to infer important features and facts representing key aspects of vehicular mobility, such as origins and destinations.
2) We measured the attraction and repulsion of different hotspot regions by indegree/outdegree and use statistical analysis to characterize the origin–destination (O-D) models [9] for various scenarios by the unique probability distribution function.
3) We exploited the road and traffic information to adjust road network topology and calibrate the trip generation algorithm and the assignment technique using the unconstrained gravity model and the weighted Dijkstra algorithm [13].
4) Extensive observations and corrections are performed on the OpenStreetMap OSM[2] to make it analogous to real-world road topology, and extensive simulations are performed to validate the effectiveness of the generated mobility models.

The rest of this paper is organized as follows: In Section II, we present the existing work; in Section III, we present the data analysis and characterization of mobility aspects; mobility model generation process is presented in Section IV; experimental observations and analysis are presented in Section V; and finally, we conclude our work in Section VI.

II. Related Work

The surge in demand for socially aware applications for VSNs has pushed the research community to seek for different possible solutions and models in the vehicular environment, including green transportation systems, smart cities, intelligent traffic control systems, road safety, entertainment along the roads, and sharing road experience, are few to mention. The performance and evaluation of the proposed protocols and solutions depend upon the adopted mobility model [8].

From the literature review, we observe that significant efforts have been made to achieve accuracy and realism required for mobility modeling. Most of the studies found in the literature about mobility modeling and simulation of vehicular networks reinforce the demand for more realistic mobility models [9], [14], [15]. The O-D model is one of the macroscopic aspects that directly impacts the mobility model. However, the existing models found in the literature either assume a non-characterized random distribution or neglected the O-D model. Simulation-based vehicular mobility models, such as car traffic in Berlin [16] and real-world scenarios in the city of Turin [17] are generated using microscopic-level simulators. These models lack the statistical characterization and authenticity required for the realistic vehicular mobility model.

Harri *et al.* presented a framework for vehicular mobility modeling and argued that trace-based models are more accurate as compared to synthetic models [9]. A number of traces exist that are directly recorded by logging the position of vehicles along their trajectories. Recently, real GPS traces of 632 taxis were recorded in Beijing [18]. The focus of this study is to optimize data dissemination in vehicular networks, not mobility modeling or characterization. Another mobility pattern of vehicles [19] was recorded from GPS traces of 4000 taxis in Shanghai. This dataset is much bigger in terms of number of vehicles as compared to the previous one. Due to high dynamic network topology in the vehicular environment, an end-to-end stable path between source and destination is rarely available. Thus, the traditional mobile *ad hoc* network routing protocols do not suit in such environment. The authors used the store-carry-forward mechanism of delay-tolerant networks to study the impact of the mobility model on data delivery and proposed

[2]https://www.openstreetmap.org

TABLE I
MOBILITY TRACES FOR VEHICULAR NETWORKS

Name	Nature	Macroscopic Data	Duration	Vehicles	Availability	Limitations
BERKELEY [21]	Synthetic	Road-side detectors	72 h	NA	Upon Req	Limited over a geographic location
CANTON [22]	Synthetic	Survey	12 h	26000	Online	Unrealistic macroscopic features
CHICAGO [23]	Real world	GPS data	17 days	1600	Online	Small scale
UMass DieselNet[3]	Real world	GPS data	60 days	35	Online	Limited number of vehicles
FRANFRUT[4]	Real world	GPS data	NA	400	NA	Small-scale mobility
LUXEMBOURG [24]	Synthetic	Road-side detectors	12 h	150000	Online	Limited to major roads
PORTO [25]	Synthetic	Perception	40 min	10000+	Upon Req	Based on simple assumption
PORTLAND [26]	Synthetic	Survey	15 min	16000+	NA	Covers only 15 min
SEATTLE [27]	Real world	GPS data	13 days	1200	Online	Limited to highway scenario
SHANGHAIGRID [19]	Real world	Real-world trace	24 h	1100+	Online	Lack of social demographic information
TORONTO [21]	Synthetic	Road-side detector	24 h	NA	Upon Req	Limited to highway scenario
TURIN [28]	Synthetic	Perception	1 h	1500	Upon Req	Based on authors' perception
ZURICH [29]	Synthetic	Survey	3 h	400+	Upon Req	Based on simple assumption

a routing protocol. However, the underling mobility model is based on a minor subset of the overall traffic in Shanghai. As mentioned earlier that an end-to-end path does not exists from a source to destination due to highly dynamic network topology in the vehicular environment. Xia *et al.* [20] analyzed the mobility pattern of 12 096 taxis in Beijing to study the intercontact interval of nodes and their impact on data delivery. In this study, the authors analyzed intercontact interval and clustering but did not study or model the trip generation or trip distribution. Intercontact interval and clustering are of great importance for data dissemination and information sharing in VSNs. However, both of the aforementioned work did not consider the impact of social demographic information on the route selection or trip distribution. The route selection of taxis and private vehicles differ in reality, and the resulting dataset might not reflect the mobility of social vehicles in the urban areas. As an instance, the mobility of taxis is restricted to specific lines or roads to avoid traffic jams or for security measurements. In other words, taxis are not allowed to travel along specific roads/lines in the city, while private vehicles are exempted from such restrictions.

A number of mobility traces have been recorded directly using position of vehicles. The routes of 500 taxis were recorded using GPS traces in San Francisco [30]. This dataset suffers from limited update frequency, time, space, and number of vehicles. Similar to taxis data, the mobility of 1200 buses were recorded using automatic vehicle position over an area of 5100 km^2 in Seattle [27]. The mobility of buses is restricted to specific routes and differs than mobility of privately owned vehicles. In [21], authors used empirical data collected from dual loop and metal detectors to measure per-lane intervehicle arrival time and spacing. The data are recorded for a time span of 24 h at two different locations, Berkeley and Toronto. This data can be fed to a microscopic simulator to derive the mobility traces of each vehicle over time, however, this dataset is limited to a highway scenario.

These models have considered multiple features and characteristics to achieve reality. However, these approaches have certain limitations if social and demographic information are considered. Some of these models are limited to highway scenarios or limited time interval. Mobility model presented in [31] and [32] has considered the characterization of mobility modeling in urban scenarios. The first one does not consider the characterization of the O-D model, while the proposed model in the second work does not suit to urban scenarios, which are different from the Cologne city. The other mentioned models do not take the traffic assignment into consideration. The traffic flow of buses and taxis are quite different than private or social vehicles, which significantly influence the mobility model and performance evaluation of socially aware applications of VSNs. An overview of existing mobility models is given in Table I. In this study, we go a step forward to exploit the traffic flow of private and public vehicles to improve traffic assignment with the characterization of the O-D model.

III. Data Analysis and Characterization

In this section, we present the data analysis and characterization process for origins, destinations, traveled distance, and travel time. These are the factors that influence the macroscopic O-D model for vehicular mobility.

A. Dataset

The work presented in this paper is based on a dataset provided by Shanghai QiangSheng Holding Co., Ltd. The dataset includes detailed traces of vehicular mobility of 13 750 taxis for one month. The data are published on the platform called Shanghai Open Data Innovation Application Contest.[5] The dataset contains the GPS records for one month of time (2015.04.01–2015.04.30). The available information from the dataset are given in Table II. Every vehicle has a unique identifier (TaxiID) representing the attributes reported in data collection. These attributes include a warning (W), status (S), lamp (L), longitude, latitude, speed, angle, and satellite. S represents the status of a taxi; a taxi is occupied if it is 1 or taxi is vacant if it is 0. We processed the raw data to extract the meaningful information that can be used for large-scale mobility modeling. The dataset from taxis may not accurately demonstrate the mobility pattern of social vehicles. However, this dataset gives insights to a number of parameters, which may be exploited for large-scale vehicular mobility including traffic density on different edges,

[3]DieselNet: https://dome.cs.umass.edu/umassdieselnet
[4]http://www.simtd.de/index.dhtml/deDE/index.html
[5]http://sodata.io/data

TABLE II
AVAILABLE INFORMATION IN DATASET

TaxiID	W	S	L	Bridge	Brake	Receive Time	GPS Time	Longitude	Latitude	Speed	Angle	Satellite
15520	0	1	3	0	0	2015-04-11 18:41:33	2015-04-11 18:41:39	121.481955	31.273432	0.0	74.0	9
27455	0	1	1	0	0	2015-04-11 10:06:14	2015-04-11 10:06:20	121.397398	31.131070	0.0	223.0	8
13682	0	0	0	0	1	2015-04-11 15:06:29	2015-04-11 15:06:35	121.280967	31.280967	0.0	170.0	7
15193	0	0	0	0	0	2015-04-11 10:58:59	2015-04-11 10:59:04	121.347783	31.216647	73.9	342.0	8
28511	0	0	0	0	0	2015-04-11 14:30:40	2015-04-11 14:30:46	121.388660	31.112390	59.8	151.0	8

hotspots, and a number of trips originating/ending in different zones. Besides, the online applications for taxi calling such as Didi Chuxing and Uber have significantly influenced the human mobility and have attracted the commuters to commute using taxis. Thus, the taxi data show resemblance to social mobility in terms of trips originating from and drawn to a particular destination. The dataset contains the mobility traces of vehicles either empty or occupied. Because we are only interested in the latter one, we process the dataset to infer the statistic of commuters using taxis.

B. Analysis and Characterization

In this study, we considered four characteristics of the existing mobility trace to infer the meaningful information using the statistical inference techniques described later. These features include the number of origins, the number of destinations, travel distance, and travel time. The raw data were preprocessed to transform it to measures representing the mobility in terms of macroscopic aspects. The proposed models and underlying methodology provide insight into synthetic modeling and characterization of the large-scale simulation scenario based on limited real traces. A random variable x represents each considered characteristic and the sample data $X_1, X_2, X_3, ...X_n$ is obtained from the trace files. We assume that the measurement of these variables is independent with each other and each sample data can be represented by a unique probability distribution function $f(x|\theta_1....\theta_k)$. Here, our objective is to infer the probability distribution function and missing variables, which represent the sample data so that it can be used to replicate it in other scenarios. The first step is to make a hypothesis about a suitable probability distribution function $f(x|\theta_1....\theta_k)$, which can represent the sample data. Once a hypothesis is made about one or more candidate probability functions, the second step is to estimate the values of their parameters. A mathematical function called *estimator* is used to estimate the unknown parameters, and there are many ways to decide about the form and quality of an estimator. We exploit the *maximum-likelihood estimators* (MLEs) [33] due to its strong and unique features distinguishing it from alternative methods. Besides, MLEs technique is one of the most popular techniques used for deriving estimators. For a sample data $X_1, X_2, X_3, ...X_n$ and hypothesized probability density function with one unknown parameter, the likelihood function is given by

$$L(\theta) = f_\theta(X_1) f_\theta(X_2) f_\theta(X_3) ... f_\theta(X_n). \qquad (1)$$

The third step is to examine the probability distribution function(s) determined in the first two steps and to evaluate how

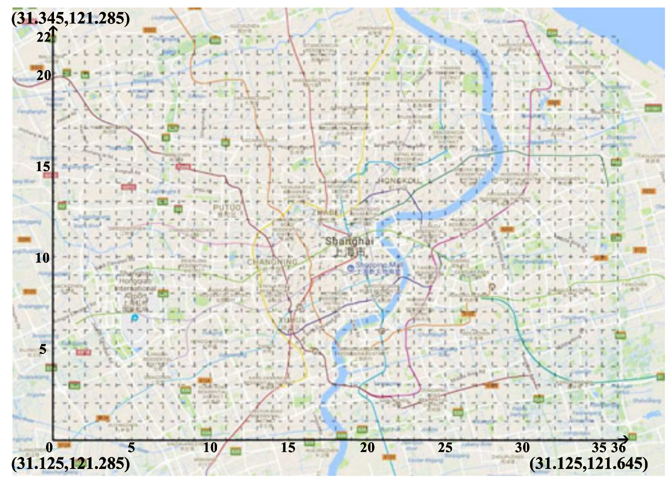

Fig. 1. Simulation area (31.125, 121.285–31.345, and 121.645).

closely these identified functions are representing our sample data. If more than one distributions are representative, we need to specify the distribution that provides the best fit. We perform a *goodness-of-fit* test for each hypothesis made in step one. We use the *Chi-Square* test [34] and empirical distribution function (EDF) test to evaluate the *goodness-of-fit* for fitted distribution. We perform the *goodness-of-fit* tests for all variables considered in our study. Besides, we use the quantile–quantile plot to evaluate the ability of a candidate probability function to represent the sample data.

C. Origins or Departures

The number of origins or departures from a particular geographic location influence the mobility model. We selected the area: 31.125, 121.285–31.345, and 121.645, and divided it into small areas of 0.01 degree (about 1 square kilometer) resulting in 792 small regions and assigned numbers (0–791), as shown in the Fig. 1. The top left corner is area 0, and the bottom right corner is area 791. The number of total taxis is 13 750 with 17.7 million trips. Our objective is to extract information from the raw dataset and infer which probability distribution can represent the number of origins or departures from each geographic location. We also aim to extract the information to identify the hotspots and compare with the information provided by government authorities.[6] We started the process by counting the number of origins/departures from each subarea using the

[6]http://www.stats-sh.gov.cn/

TABLE III
STATISTICAL SUMMARY AND QUANTILES OF ORIGINS/DEPARTURES

Data	N	Min	Max	Q1	Median	Q3	Mean	Std. Dev.	Variance	Skewness	Kurtosis
Weekdays morning	792	0.00	269.00	3.00	18.00	53.00	34.6805560	43.1169610	1859.0723000	1.8313848	3.7759544
Weekend morning	792	0.00	318.00	2.00	16.00	49.00	32.9936870	41.4596960	1718.9064000	1.8069208	4.3316709
Weekdays evening	792	0.00	510.00	2.00	17.00	56.75	41.6477270	62.6054880	3919.4472000	2.9717440	11.8030570
Weekend evening	792	0.00	502.00	3.00	18.50	65.00	48.2790400	73.8753460	5457.5668000	2.8613793	10.0953460

TABLE IV
GOODNESS-OF-FIT ANALYSIS FOR ORIGINS/DEPARTURES

Distribution	Weekdays Morning		Weekdays Evening		Weekend Morning		Weekend Evening		Ho
	χ^2	p-Value	χ^2	p-Value	χ^2	p-Value	χ^2	p-Value	
Gamma-Poisson	580.161172	1.000000	547.867876	1.000000	771.281187	0.685500	773.356664	0.665000	Accept
Poisson	42402.3640	<0.0001	41209.5494	<0.0001	74440.6218	<0.0001	89416.3450	<0.0001	Reject
Binomial	48677.7691	<0.0001	45980.1629	<0.0001	81060.1749	<0.0001	98930.8610	<0.0001	Reject

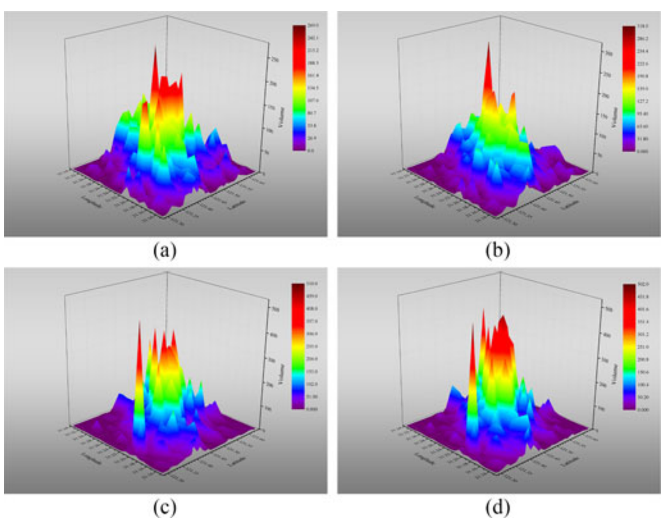

Fig. 2. Distribution of origins/departures. (a) Weekdays morning. (b) Weekdays evening. (c) Weekend morning. (d) Weekend evening.

following equation:

$$O_i = \sum_j T_{ij} \tag{2}$$

where T_{ij} is the total number of trips between origin i and destination j. O_i is the number of trips originating from origin i. In our analysis, we considered four different time slots to characterize the pattern: weekdays morning (7:00–9:00), weekdays evening (17:00–19:00), weekend morning (7:00–9:00), and weekend evening (17:00–19:00). From Fig. 2, it is clear that origins on weekdays are uniformly distributed, while only few spots showing high numbers of origins on weekends as compared to weekdays.

Based on the histogram and statistical summary given in Table III, we made a hypothesis about the possible distribution functions that may represent the sample data. *N* represents the sample size (number of small areas). Q1 represents the first Quartile of sample data and Q3 represents the third Quartile (median of the upper half of dataset). Exploiting the MLEs technique and *Chi-square goodness-of-fit* test, we observed that Gamma-Poisson distribution is more likely to represent the origins, as shown in Fig. 2. The *Chi-square* test computes, among other values, the χ^2 statistics and the p-value. The χ^2 and p-value are presented in Table IV. To summarize, the smaller the χ^2 and the higher the p-value, the more representative the hypothetical distribution. Subplots (a) and (b) in Fig. 3 show the histogram

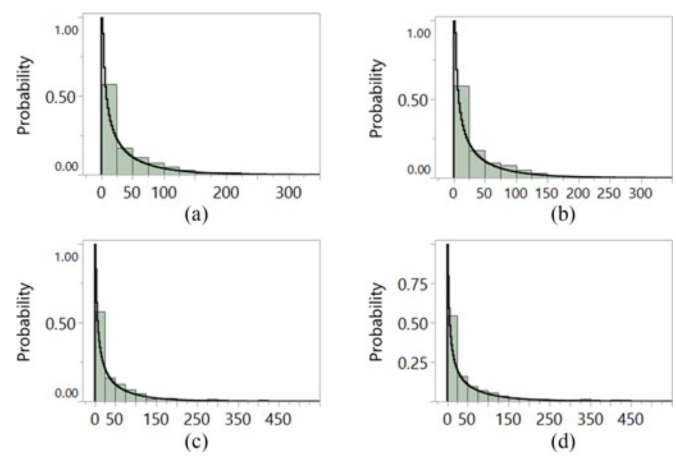

Fig. 3. Histogram and fitted distribution for origins. (a) Weekdays morning. (b) Weekdays evening. (c) Weekend morning. (d) Weekend evening.

TABLE V
ESTIMATED PARAMETERS α AND β FOR GAMMA-POISSON (DEPARTURES)

Data	α	β
Weekdays morning	34.6806	73.0866
Weekend morning	32.9937	75.2180
Weekdays evening	41.6477	96.5155
Weekend evening	48.2790	115.6210

TABLE VI
STATISTICAL SUMMARY AND QUANTILES OF DESTINATION/ARRIVALS

Data	N	Min	Max	Q1	Median	Q3	Mean	Std. Dev.	Variance	Skewness	Kurtosis
Weekdays morning	792	0.00	824.00	13.00	55.00	177.75	132.4457100	183.2260600	33571.7900000	2.0933057	4.1388198
Weekend morning	792	0.00	372.00	5.00	18.00	62.75	47.6654040	69.0503060	4767.9448000	2.2071163	4.8335369
Weekdays evening	792	0.00	895.00	16.00	80.00	256.00	164.9709600	197.7112700	39089.7450000	1.5417836	1.8777701
Weekend evening	792	0.00	635.00	6.00	30.50	99.75	72.6224750	106.3383900	11307.8540000	2.6547442	8.2404730

TABLE VII
GOODNESS-OF-FIT ANALYSIS FOR DESTINATION/ARRIVALS

Distribution	Weekdays Morning		Weekdays Evening		Weekend Morning		Weekend Evening		Ho
	χ^2	p-Value	χ^2	p-Value	χ^2	p-Value	χ^2	p-Value	
Gamma-Poisson	816.671513	0.256100	815.235984	0.267600	624.116087	1.000000	797.582053	0.427900	Accept
Poisson	200499.4070	<0.0001	79123.3056	<0.0001	187426.8560	<0.0001	123164.5180	<0.0001	Reject
Binomial	238843.3830	<0.0001	90751.5574	<0.0001	229781.3200	<0.0001	139069.3360	<0.0001	Reject

and fitted distribution for origins in the morning and evening on weekdays, respectively. Similarly, subplots (c) and (d) show the histogram and fitted distribution for origins in the morning and evening rush hours on weekend, respectively. Gamma-Poisson distribution more likely represents the sample date. A Gamma-Poisson random variable x has probability mass function

$$f(x) = \frac{\Gamma(x+\beta)\alpha^x}{\Gamma(\beta)(1+x)^{\beta+x}x!} \quad x = 0, 1, 2, ... \tag{3}$$

where the values of parameters α and β for Gamma-Poisson distribution are given in Table V.

D. Destinations or Arrivals

From the raw data, we use the following equation to extract the number of arrivals/destinations in each zone/subarea:

$$D_j = \sum_i T_{ij} \tag{4}$$

where T_{ij} is the total number of trips between origin i to destination j. D_j is the number of arrivals in zone j. Table VI represents the statistical summary and quantiles of arrivals across different zones. Based on the statistical summary, we can get some conclusions about the destination areas. Smaller values of $Q1$ on the weekend show that few vehicles arrive in few areas as compared to arrivals on weekdays. The high value of variance on weekdays in the morning as well as in the evening show a great variability of the data, which means a high number of destinations on weekdays as compared to weekend. Based on the statistical observation, it is possible to formulate an initial hypotheses about the candidate probability function(s) $f(x|\theta_1 ....\theta_k)$ that could possibly represent this sample data. Similar to our analysis in previous subsection, we follow our methodology to infer the parameters of considered probability distributions functions, and then, use the *Chi-square* method to evaluate and decide about the most appropriate probability distribution function that represents our sample data. Table VII presents the χ^2 and the p-value for Gamma-Poisson, Poisson, and Binomial distributions. The results are compared and it is observed that the most suitable distribution is Gamma-Poisson distribution. The estimated values of parameters α and β for Gamma-Poisson for arrivals are given in Table VIII.

Subplots (a) and (b) in Fig. 4 represent the histogram and fitted distribution for arrivals during the morning and evening rush hours on weekdays, respectively, while subplots (c) and (d)

TABLE VIII
ESTIMATED PARAMETERS α AND β FOR GAMMA-POISSON (ARRIVALS)

Data	α	β
Weekdays morning	132.4460	245.5080
Weekend morning	47.6654	97.0557
Weekdays evening	164.9710	300.3080
Weekend evening	72.6225	154.4220

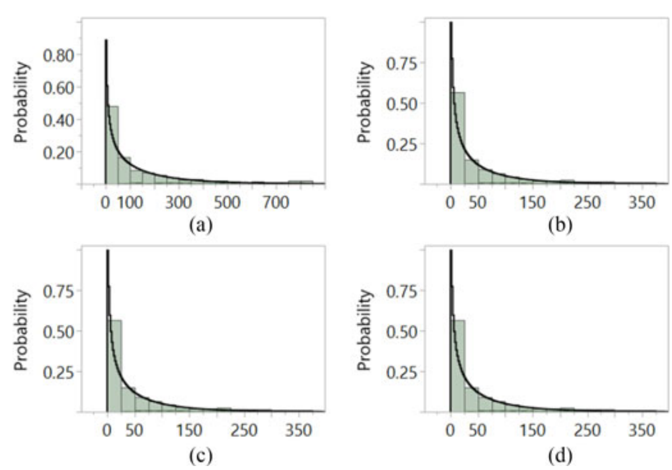

Fig. 4. Histogram and fitted distribution for destinations. (a) Weekdays morning. (b) Weekdays evening. (c) Weekend morning. (d) Weekend evening.

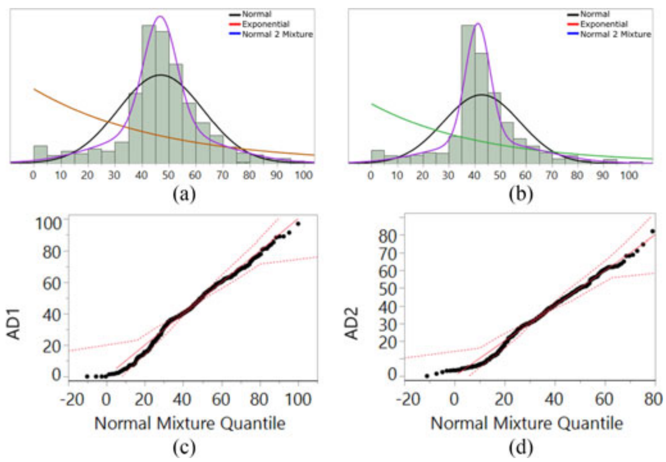

Fig. 5. Histogram and fitted distribution for travelled distance. (a) Distribution on weekdays. (b) Distribution on weekend. (c) Weekdays' quantile. (d) Weekend's quantile.

show the histogram and fitted distribution for arrivals during the morning and evening rush hours on weekend, respectively.

E. Traveled Distance and Travel Time

Besides the number of origins and distribution across the city average traveled distance and average travel time are also important macroscopic features that should be considered while defining a macroscopic mobility model. The traffic volume varies at different intervals that affects the total travel distance or travel time. During rush hours, the drivers either select a long route to avoid traffic congestions or spend a lot of time on road signals toward their destinations. Consequently, the distance and time depend upon the traffic volume at a different interval. So, it is of great importance to analyze the variation during weekdays/weekend and consider to define a macroscopic mobility model. The origins and distribution are analyzed in morning and evening, but average traveled distance and travel time do not need such consideration. In this section, we analyze average traveled distance and travel time on weekdays and weekend.

We analyze the average traveled distance and observe that more vehicles depart from origins toward their destinations on weekdays with longer distance as compared to the weekend. Commuters choose their destinations near to their residential areas on weekend, because the commercial areas are situated far from residential areas from where these commuters commute on weekdays. We perform the analysis to fit a candidate probability function to this sample data and infer the information that can be used while generating mobility models on weekdays and weekend. To evaluate the distribution that more likely presents the average traveled distance, we compare the quantile of fitted distribution and sample data. Fig. 5(a) and (b) shows the comparison of normal, exponential, and normal 2 mixture distribution for weekdays and weekend, respectively. It is observed that normal 2 mixture distribution is the most suitable to represent the average traveled distance on weekdays and weekend, respectively. Fig. 5(c) and (d) shows quantile comparison of fitted distribution against actual sample data. In both scenarios, weekdays and weekend, normal 2 mixture is the more likely representing the sample data. However, the distribution parameters are different for weekdays and weekend, which are given in Table IX.

TABLE IX
ESTIMATED VALUES OF PARAMETERS FOR TRAVELED DISTANCE

	$\mu 1$	$\mu 2$	$\sigma 1$	$\sigma 2$
Weekdays	44.953137	44.753263	19.929913	5.724430
Weekend	33.842494	34.013563	3.789011	16.224680

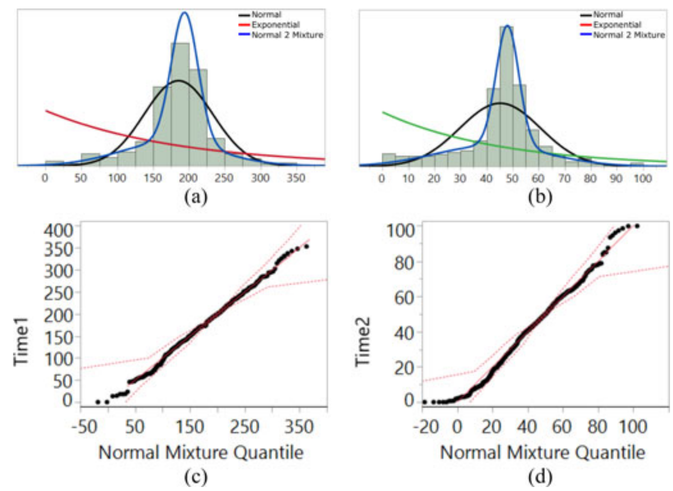

Fig. 6. Histogram and fitted distribution for travel time. (a) Distribution on weekdays. (b) Distribution on weekend. (c) Weekdays' quantile. (d) Weekend's quantile.

TABLE X
ESTIMATED VALUES OF PARAMETERS FOR TRAVELED TIME

	$\mu 1$	$\mu 2$	$\sigma 1$	$\sigma 2$
Weekdays	172.542300	194.000870	69.656107	18.678307
Weekend	41.840601	47.986713	21.725294	4.379850

Similar to average traveled distance, we also perform the analysis for average travel time on weekdays and weekend. Travel time from the source region to the destination region plays an active role in mobility modeling influencing number of vehicles along the routes. We take these two macroscopic aspects for characterization calibration of mobility models. Based on the results and comparison, as shown in Fig. 6(a) and (b), we infer that normal 2 mixture distribution is the most suitable candidate probability distribution to represent the average travel time on weekdays and weekend. Fig. 6(c) and (d) shows the quantile comparison for weekdays and weekend, respectively. The estimated values of parameters are presented in Table X.

IV. MOBILITY MODELING AND DATASET GENERATION

The dataset we analyzed here provides 25% of taxis deployed in the city. These data are realistic and provides meaningful insights into vehicular mobility in the city in terms of identifying the hotspot areas, road trajectories, social interaction of hotspot areas in terms of trips originating/ending from/in different areas, and so on. These features and aspects can be exploited for large-scale synthetic mobility modeling. In this section, we present mobility model based on data analysis in the previous section. We consider the possible macroscopic aspects with simulation tool to present the mobility models with a possible level of realism. We introduce realistic mobility models that may be used for simulation of VSNs analysis. These models are based on the real-time traffic data under a reliable and credible microscopic mobility simulator.

A. Microscopic Mobility

The macroscopic and microscopic characteristics are equally important and are required to be considered to generate realistic mobility models. Inappropriate representation of mobility characteristic of the mobility model, either macroscopic or microscopic, may lead to unreliable results and conclusion. Microscopic aspects of mobility models define and present the behavior of vehicles more precisely. We select an open-source discrete-time traffic simulator, called simulation of urban mobility (SUMO),[7] to simulate the microscopic mobility of vehicles. SUMO has the capability to model the individual behaviors of the vehicle more accurately. Also, SUMO provides support to import road maps from multiple sources, including OSM. Besides, SUMO supports scalable, time scheduled, multimodal traffic mobility modeling with portable libraries. SUMO provides flexible solutions for various research directions, such as navigation, traffic management, vehicle-to-vehicle communication, and routing, are few to mention. SUMO can be used to generate vehicular mobility models with macroscopic and microscopic behaviors, and the resulting trace files can be imported to network simulators, such as NS2 and NS3,[8] for performance analysis of socially aware data dissemination and routing protocols for VSNs.

B. Trip Generation and Distribution

With increasing interest of research community in social aware networking solutions for a diverse range of application, the demand for realistic mobility models grew in the last few years. The research community has focused on microscopic as well as macroscopic mobility aspects of vehicular mobility. Certain issues, including traffic demand, have encouraged the research community to investigate further and produce realistic mobility models for vehicular scenarios, i.e., VSNs. One of the basic building blocks for the vehicular mobility model is the O-D model. The simplistic or unrealistic approaches found in the literature encourage the research community to come up with mobility models that are more realistic and analogous to real-world traffic condition. In this section, we present the generation process for the characterized O-D model and uniform traffic assignment based on mobility traces.

The O-D matrix mimics the mobility pattern of vehicles in terms of trips from a specific origin to a specific destination.

[7]http://www.dlr.de/

[8]NS2 & NS3 are discrete-event network simulators. https://www.nsnam.org/

Area	1	2	3	.	.	j	.	.	n	O_i
1	T_{11}	T_{12}	T_{13}	.	.	T_{1j}	.	.	T_{1n}	O_1
2	T_{21}	T_{22}	T_{23}	.	.	T_{2j}	.	.	T_{2n}	O_2
3	T_{31}	T_{32}	T_{33}	.	.	T_{3j}	.	.	T_{3n}	O_3
. .	. .	. .	. .	. .	. .	. .	. .	. .	. .	. .
i	T_{i1}	T_{i2}	T_{i3}	.	.	T_{ij}	.	.	T_{in}	O_i
. .	. .	. .	. .	. .	. .	. .	. .	. .	. .	. .
n	T_{n1}	T_{n2}	T_{n3}	.	.	T_{nj}	.	.	T_{nn}	O_n
D_j	D_1	D_2	D_3	.	.	D_j	.	.	D_n	T

Fig. 7. Origin and destination trip matrix.

Trip generation and trip distribution are both equally important to generate an accurate O-D matrix. Trip generation is the first step toward a realistic mobility model. Restriction by traffic authority, demographic information, residential density, accessibility, social behaviors of commuters, and different time intervals of the day and week are some of the factors that affect the trip generations process. A simplistic approach may not be applied to model the trip generation. The process of trip generation requires maximum possible realization. So, we use the statistical inferences and characterization presented in the previous section to model the total number of trips generated from and attracted to each subarea during a particular time interval. We use different probability distribution functions with estimated value of parameters for trip generation. The second step is to define trip distribution, which represents the number of trips originating from particular area i to a destination area j. In other words, we need to define the number of edges that connect different geographic locations (vertices of the graph). Fig. 7 shows the notation of a trip matrix. As mentioned earlier, the whole area is divided into 792 subareas (vertices) and trips generated from and attracted by each region represent the edges between these vertices. The attraction and repulsion of these vertices are defined by indegree and outdegree, while the number of edges between vertices represents the number of trips generated/attracted from/to a specific location.

There are a number of methods for estimating the O-D matrix, such as roadside interviews, flagging methods, aerial photography, and car following. In contrast, we are interested in estimating the O-D matrix based on the description of the network and traffic counts in terms of origins and destinations. Also, some of the methods for estimating O-D matrices require additional information such as population and employment. The gravity model is one of the methods applied in transport engineering to define the O-D matrix based on traffic count. This model is similar to Newton's gravitational law. It is assumed that the trips generated from an origin to a destination are directly proportional to the total number of trips originated and attracted from/to a specific location. The following equation is used to estimate the number of trips originating from area i to area j:

$$T_{ij} = \alpha \frac{P_i E_j}{{d_{ij}}^2} \tag{5}$$

where P_i is the population in area i, E_j is the employment in area j, and d_{ij} is the distance or travel time between i and j. α is the parameter of calibration. The O-D matrix is a sparse matrix and the attraction and repulsion of hotspot regions is defined by number of origins and destinations. So, we use the following unconstrained gravity model to define the O-D matrix:

$$T_{ij} = \alpha \frac{(O_i D_j)^\beta}{{c_{ij}}^\gamma} \tag{6}$$

where c_{ij} is the cost from origin i to destination j. We select distance as the cost in our analysis. Utilizing the dataset analyzed in previous section and using the least-square method, we estimate the value of parameters α, β, and γ. Because the O-D matrix got by the unconstrained gravity model does not conform to the constraint conditions of trip distribution, we use the average growth coefficient method for iteration to calibrate the O-D matrix until the parameters of ${F_{O_i}}^m$ and ${F_{D_i}}^m$ are less than 3%.

$$f_{\text{average}}({F_{O_i}}^{m+1}, {F_{D_j}}^{m+1}) = \frac{1}{2}({F_{O_i}}^m, {F_{D_j}}^m) \tag{7}$$

$${T_{ij}}^{m+1} = {T_{ij}}^m({F_{O_i}}^m, {F_{D_j}}^m)/2 \tag{8}$$

where ${F_{O_i}}^m = O_i/{U_i}^m$, ${F_{D_i}}^m = D_i/{V_i}^m$, ${U_i}^m = \sum_j {T_{ij}}^m$, and ${V_j}^m = \sum_i {T_{ij}}^m$.

C. Traffic Assignment

Traffic assignment techniques compute the possible fastest and the possible route from source to the destination point. The traffic intensity varies over time. Besides, some special restrictions on some of the routes from traffic control authorities significantly impact the process of route selection. For example, public vehicles without some special permissions or tokens are not allowed to travel on roads passing through some educational or public security institutions due to various reasons. Similarly, some routes are specified for specific vehicles to travel along those routes, such as bus rapid transit. Consequently, selection of the fastest route from source to destination without prior knowledge of road topology and restriction from traffic control authorities may result in an appropriate and unrealistic traffic assignment.

Traffic volume on a specific edge from source area i to destination area j affects the process of the traffic assignment. To calculate an optimum path from source to destination is a challenging task and many factors influence this process, including traffic volume, the capacity of route, and travel time. We exploit the simplest assignment technique, accept all-or-nothing, to measure the traffic flow on each edge for all O-D pairs.

$$Va = \sum_{ij} T_{ij} P_{ij} a \tag{9}$$

$$P_{ij} a = \begin{cases} 1, & \text{if } T_{ij} \text{ travel on edge a} \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

Initially, we use the dataset analyzed in this paper to update the edges and use weighted Dijkstra's algorithm for the traffic assignment.

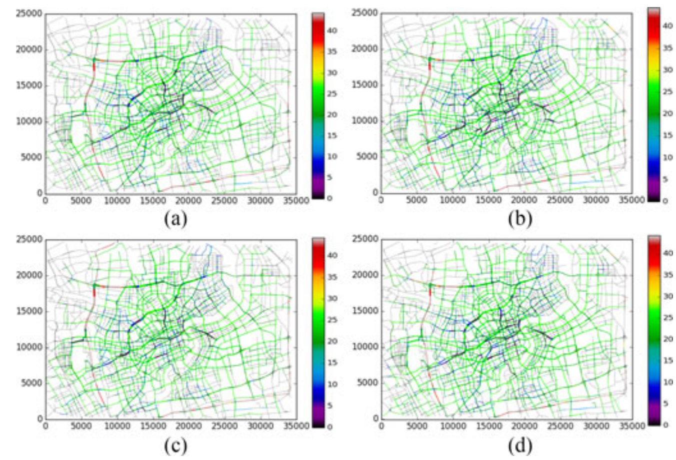

Fig. 8. Initial traffic flow. (a) Weekdays morning. (b) Weekdays evening. (c) Weekend morning. (d) Weekend evening.

D. Network Topology

In our simulation, we use the street layout of Shanghai city from OSM database. The layout provides details about road signals, topology, and information about different points of interest, including commercial, residential, and recreational spots. We extract the road topology for the particular area mentioned in Section III. However, comparing with actual road topology, this imported layout shows inconsistent information influencing the vehicular mobility. In order to make the road network topology analogous to real-world road network topology, we perform several operations on this imported map, which we describe in detail in the following section.

V. Experimental Results and Analysis

Our experiment and analysis are based on the strong statistical inference technique to model the number of trips originated from and attracted by particular subarea using a candidate probability function with known values of parameters. The resulting O-D data, generated O-D matrix for four different scenarios, and the OSM road layout are combined to produce vehicular mobility dataset for four different scenarios that may be used for research and analysis of different communication and application protocols for VSNs.

Initial analysis after running SUMO with input mentioned previously shows unrealistic behaviors, such as heavy traffic on some road segments and nonuniform traffic assignment. By observing the color of edges, black color is obvious. The dark color shows a low traffic flow or a heavy traffic jam on these routes, as shown in Fig. 8.

This undesirable situation is caused by several factors including inappropriate conversion of the OSM road layout, unnecessary traffic movement restrictions, inconsistent road information, the number of lanes, and unrealistic traffic assignment. In order to deal with this problem, we investigate the OSM road layout and identify the inconsistent road information, redundant traffic signals, road layout, and the number of lanes and roads that shows deviation from real road network topology observed from Google Map and Baidu Map.

The OSM project provides high-quality free maps that can be imported into SUMO. These maps are updated and contributed by a vast user community, but not specifically designed for vehicular mobility simulations [35]. The automated conversion of OSM map results in an inaccurate representation of road topology that significantly impairs the mobility generation process. The first source of errors which causes unrealistic vehicular mobility pattern is the number of lanes of each edge during the automatic conversion. In order to make the number of lanes consistent with real-world road topology, we use Java OSM[9] editor to repair the OSM data based on information provided by Shanghai Municipal Government.[10]

The vehicular mobility is highly influenced by the redundant road traffic signals and inconsistent restrictions on some edges or segments. The road layout imported from OSM database includes information about road junctions and signals. During OSM to SUMO conversion, SUMO assigns periodic ON/OFF timers to these signals based on the priority of edges. Besides, SUMO also uses an automated algorithm to place additional traffic lights. Similarly, wrong moment restriction on road segments causes traffic congestion. To deal with this situation, first, we analyzed the impact of these redundant traffic signal and removed the signals having a negative impact on traffic flow. Then, we used the JOSM editor to edit the OSM map to remove wrong traffic movement restriction or missing restriction required to regulate certain edges. For example, as shown in Fig. 9(a), according to Google Street View, there exists a no left turn from Sinan road to Taikang road, however, missing in the converted road topology as in Fig. 9(b), we revised the map as shown in Fig. 9(c). We investigated the area we used in our analysis and made the required correction on most of T-junctions.

The OSM data contain wrong restrictions on some of the road segments. These limitations do not affect the whole of the street layout, but affecting the microscopic vehicular mobility. We identify such wrong restriction by comparing with real-world road topology and correct the OSM data. In the initial OSM data, there is a no right turn from Jinzhou Road to Yunshun Road. Because this restriction is only valid for a portion of the edge, and we need to separate the portion, as shown in Fig. 10(a), we separate the segment of Jinqiao Road as shown in Fig. 10(b). Such kind of errors cause vehicles to travel a long distance to reach the destination or wait for an unspecified time interval to turn right/left.

As mentioned earlier, automatically converted OSM data into SUMO is simply unfit for microscopic vehicular mobility simulations. SUMO interprets the exceedingly intricate intersections, such as two closely segments in OSM topology as if two road junctions coexist, one next to the other. As a consequence, the number of traffic lights that regulate the car flows into the crossroad is doubled, resulting long waiting time for vehicles to cross the intersections. In order to fix such a problem, we join road segment links that refer to the same physical intersection and replace the traffic signals by circuits, as shown in Fig. 11.

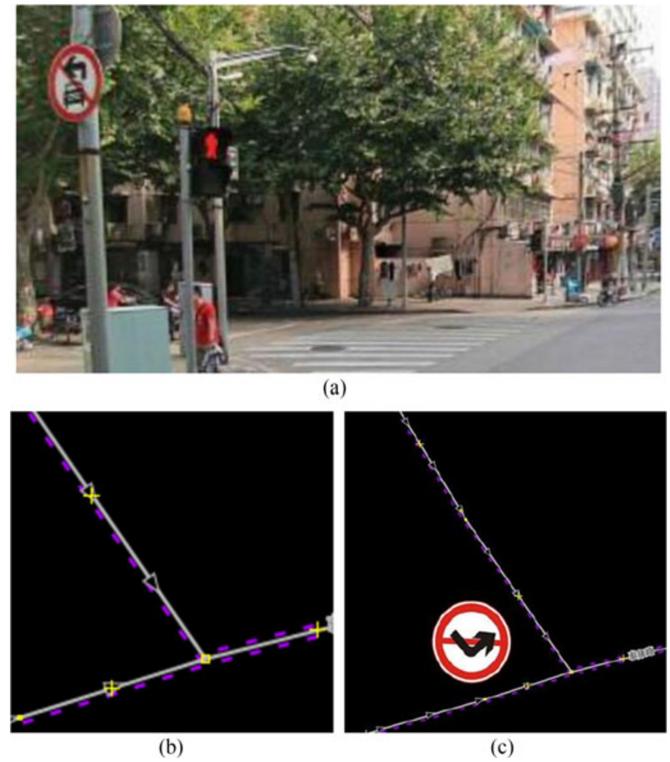

Fig. 9. Example or missing road restriction. (a) Observed from google map. (b) No restriction. (c) Restriction.

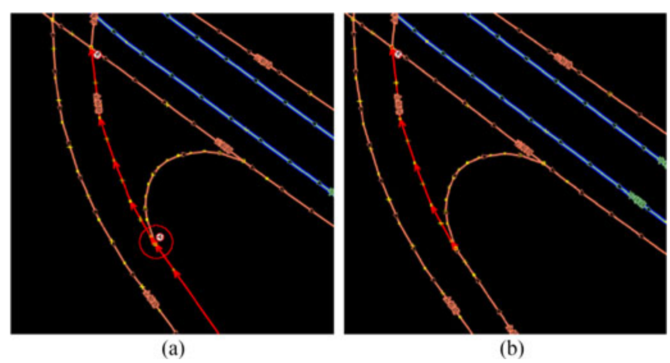

Fig. 10. Example of incorrect restriction. (a) Wrong restriction. (b) After correction.

We use the trace files to update the weights of road edges. However, route selections of taxis and social vehicles differ in reality. Similarly, drivers select the feasible routes from source to destination based on their experience. A simplest Dijkstra's algorithm on road topology with limited information results in traffic congestion on some routes; based on the initial weights of edges. In order to make the traffic assignment analogous to traffic flow in the city, we exploit Traffic Performance Index (TPI) of Shanghai city to update edges and perform other operations such as removing some edges and adding extra lanes that were missing during the automated conversion process. The traffic intensity varies over time, and traffic assignment depends upon

[9]https://josm.openstreetmap.de
[10]http://www.datashanghai.gov.cn/

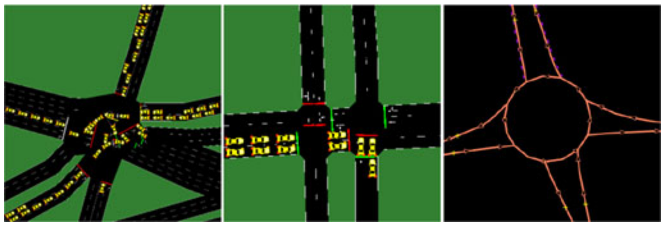

Fig. 11. Example of wrong restriction.

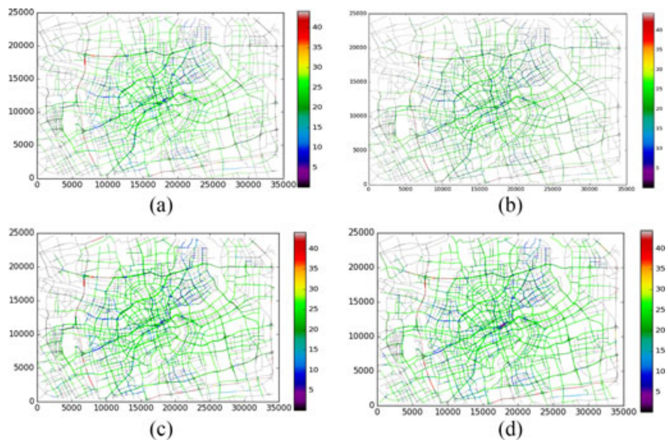

Fig. 12. Final traffic flow. (a) Weekdays morning. (b) Weekdays evening. (c) Weekend morning. (d) Weekend evening.

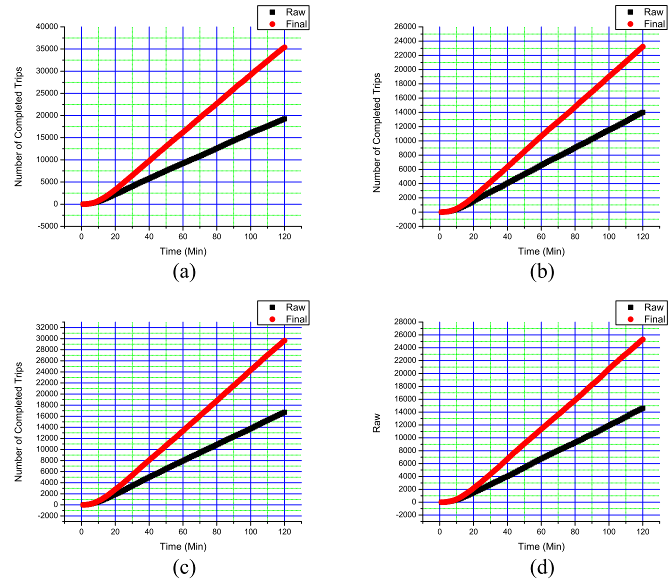

Fig. 13. Completed trips. (a) Weekdays morning. (b) Weekdays evening. (c) Weekend morning. (d) Weekend evening.

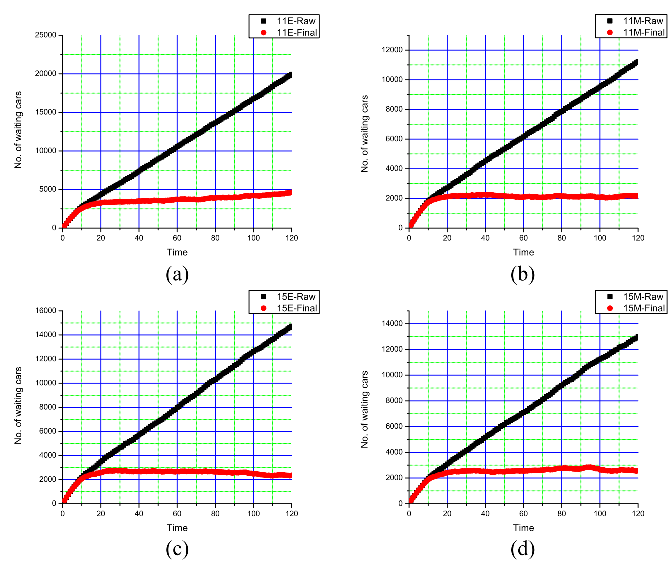

Fig. 14. Number of waiting cars. (a) Weekdays morning. (b) Weekdays evening. (c) Weekend morning. (d) Weekend evening.

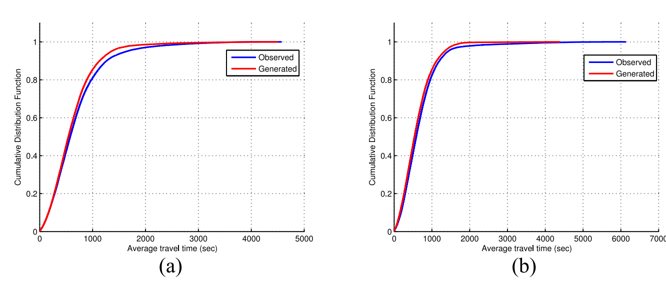

Fig. 15. Comparison of travel time. (a) Weekdays morning. (b) Weekdays evening.

the cost of the edge that reflects the traffic intensity on that edge. We use the TPI data to update all the edges in different time slots, such as morning peak and evening peak. With these changes, the traffic flow has significantly improved in terms of road utilization and uniform traffic distribution.

After making the aforementioned changes, we observe that traffic flow has been improved significantly, as shown in Fig. 12. The resulting datasets present the vehicular mobility during morning peak and evening peak both on weekdays and weekend. The updated road topology and the performed changes to OSM data, given in the previous section, provide realistic mobility environment for the traffic demand, which shows normal traffic flow in the area. The dark color is reduced that means the problem with road congestions is reduced significantly.

The behavior of resulting mobility models is compared with the simulation results before performing the aforementioned changes. The number of completed trips and number of waiting cars over time in the four scenarios before and after the changes are shown in Figs. 13 and 14, which show significant improvement and the number of completed trips increases over time, and waiting cars reduced significantly.

Moreover, the resulting mobility models show the same behaviors in terms of average traveled distance and travel time as characterized in Section III-E. Fig. 15 presents the validation of average traveled time for weekend and weekdays. It is possible to see that generated cumulative distribution function for both weekend and weekdays show behavior similar to observed dataset. Our observation shows that the generated mobility models show similarity with vehicular mobility observed in the real world, which can be used for performance evaluation of different network and application layer protocols for VSNs, such as data dissemination and routing. The performance evaluation of such application and protocols depend upon the underlying mobility model and network connectivity. In order to validate our generated mobility models in terms of network connectivity, we analyze these models in terms of node degree. The direct communication of mobile nodes is influenced by a number of factors in urban areas, such as multipath fading, signal propagation, limited transmission range of dedicated short range communications technologies, medium access schemes, and limited transmit power and energy budget of mobile devices [36], [37]. However, issues related to signal propagation and communication are out of the scope of this paper, and we assume a very simplistic model to measure a node degree with disk model of communication range of 200 m. Node degree represents the number of neighboring nodes within the communication range of the source node. Observing statistical summary of node degree for all the scenarios, we observe that 75% nodes have more than 33 neighboring nodes, while only 25% of nodes have 8 or less number of neighboring nodes within communication range. Thus, the generated mobility models show consistent behaviors in terms of connectivity required for routing and dissemination protocols analysis.

VI. Conclusion

This paper aims to characterize different aspects of vehicular mobility and develop large-scale datasets for evaluation and analysis of context-aware applications of VSNs for green transportation systems. We use statistical inference techniques to characterize mobility characteristics and perform simulations with real trace to infer different aspects that possibly impact the vehicular mobility. We generated mobility model for four different scenarios exploiting road and traffic information to make the generated mobility models analogous to real world. We use a statistical inference technique to characterize the origins and destinations across the city and use unconstrained gravity model for traffic generation. Besides, we exploit the real-time traffic information of private vehicles and road topology to make the traffic assignment analogous to real world using weighted Dijkstra algorithm. The resulting mobility models mimic the realistic vehicular moments in morning and evening peaks both on weekdays and weekend.

When it comes to the performance analysis in a vehicular environment, such as VSNs, several challenges exist that need to be tackled to make the performance analysis and results more credible and realistic. The mobility models should be more realistic and detailed in terms of macroscopic behavior, and traffic assignment as mobility pattern of private cars significantly differs than taxis and buses. Signal propagation and the underlying network architectures are the other factors that influence the performance and communication in a vehicular environment. In our future study, we intend to analyze different mobility models using socially aware protocols for VSNs to evaluate the impact of the road layout and traffic assignment in various scenarios.

References

[1] 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. 49–55, May 2017.

[2] R. Yu, J. Ding, X. Huang, M. T. Zhou, S. Gjessing, and Y. Zhang, "Optimal resource sharing in 5G-enabled vehicular networks: A matrix game approach," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 7844–7856, Oct. 2016.

[3] B. Liu, D. Ghosal, Y. Dong, C. N. Chuah, and M. Zhang, "CarbonRecorder: A mobile-social vehicular carbon emission tracking application suite," in *Proc. 2011 IEEE Veh. Technol. Conf.*, Sep. 2011, pp. 1–2.

[4] X. Chen, E. Santos-Neto, and M. Ripeanu, "Crowd-based smart parking: A case study for mobile crowdsourcing," in *International Conference on Mobile Wireless Middleware, Operating Systems, and Applications*. Berlin, Germany: Springer, 2012, pp. 16–30.

[5] 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.

[6] W. Sha, D. Kwak, B. Nath, and L. Iftode, "Social vehicle navigation: Integrating shared driving experience into vehicle navigation," in *Proc. 14th ACM Workshop Mobile Comput. Syst. Appl.*, 2013, pp. 16-1–16-6.

[7] X. Hu *et al.*, "Social Drive: A crowdsourcing-based vehicular social networking system for green transportation," in *Proc. 3rd ACM Int. Symp. Design Analysis Intell. Veh. Netw. Appl.*, 2013, pp. 85–92.

[8] R. Baumann, S. Heimlicher, and M. May, "Towards realistic mobility models for vehicular ad-hoc networks," in *Proc. 2007 Mobile Netw. Veh. Environ.*, May 2007, pp. 73–78.

[9] J. Harri, 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, Fourth Quarter 2009.

[10] A. M. Vegni and V. Loscr, "A Survey on Vehicular Social Networks," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2397–2419, Fourth Quarter 2015.

[11] W. Viriyasitavat, F. Bai, and O. K. Tonguz, "Dynamics of network connectivity in urban vehicular networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 3, pp. 515–533, Mar. 2011.

[12] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.

[13] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, no. 1, pp. 269–271, Dec. 1959.

[14] F. Mezghani, R. Dhaou, M. Nogueira, and A. L. Beylot, "Content dissemination in vehicular social networks: Taxonomy and user satisfaction," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 34–40, Dec. 2014.

[15] X. Kong, X. Song, F. Xia, H. Guo, J. Wang, and A. Tolba, "LoTAD: Long-term traffic anomaly detection based on crowdsourced bus trajectory data," *World Wide Web*, Aug. 2017.

[16] H. Conceição, L. Damas, M. Ferreira, and J. a. Barros, "Large-scale simulation of V2V environments," in *Proc. ACM Symp. Appl. Comput.*, 2008, pp. 28–33.

[17] C. Barberis and G. Malnati, "Epidemic information diffusion in realistic vehicular network mobility scenarios," in *Proc. 2009 Int. Conf. Ultra Modern Telecommun. Workshops*, Oct. 2009, pp. 1–8.

[18] J. Ahn, M. Sathiamoorthy, B. Krishnamachari, F. Bai, and L. Zhang, "Optimizing content dissemination in vehicular networks with radio heterogeneity," *IEEE Trans. Mobile Comput.*, vol. 13, no. 6, pp. 1312–1325, Jun. 2014.

[19] H. Y. Huang *et al.*, "Performance evaluation of SUVnet with real-time traffic data," *IEEE Trans. Veh. Technol.*, vol. 56, no. 6, pp. 3381–3396, Nov. 2007.

[20] C. Xia, D. Liang, H. Wang, M. Luo, and W. Lv, "Characterization and modeling in large-scale urban DTNs," in *Proc. 37th Annu. IEEE Conf. Local Comput. Netw.*, Oct. 2012, pp. 352–359.

[21] F. Bai and B. Krishnamachari, "Spatio-temporal variations of vehicle traffic in VANETs: Facts and implications," in *Proc. 6th ACM Int. Workshop Veh. Internetw.*, 2009, pp. 43–52.

[22] B. Raney, N. Cetin, A. Völlmy, M. Vrtic, K. Axhausen, and K. Nagel, "An agent-based microsimulation model of swiss travel: First results," *Netw. Spatial Econ.*, vol. 3, no. 1, pp. 23–41, 2003.

[23] M. Doering, T. Pögel, W.-B. Pöttner, and L. Wolf, "A new mobility trace for realistic large-scale simulation of bus-based DTNs," in *Proc. 5th ACM Workshop Challenged Netw.*, 2010, pp. 71–74.

[24] Y. Pigné, G. Danoy, and P. Bouvry, "A vehicular mobility model based on real traffic counting data," in *Proc. Int. Workshop Commun. Technol. Veh.*, 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] P. Wagner and K. Nagel, "Microscopic modeling of travel-demand: Approaching the home-to-work problem," Los Alamos National Lab., Los Alamos, NM, USA, Tech. Rep. no. LA-UR-98-3368; CONF-990112-ON: DE99002281; TRN: AHC29914%%74, 1998. [Online]. Available: https://www.osti.gov/scitech/biblio/334270

[27] S. Ahmed and S. K. Salil, "Characterization of a large-scale delay tolerant network," in *Proc. IEEE Local Comput. Netw. Conf.*, Oct. 2010, pp. 56–63.

[28] C. Barberis and G. Malnati, "Epidemic information diffusion in realistic vehicular network mobility scenarios," in *Proc. 2009 Int. Conf. Ultra Modern Telecommun. Workshops*, 2009, pp. 1–8.

[29] F. Cunha, A. Carneiro Vianna, R. Mini, and A. Loureiro, "Is it possible to find social properties in vehicular networks?" in *Proc. IEEE Symp. Comput. Commun.*, Jun. 2014, pp. 1–6.

[30] P. O. V. de Melo *et al.*, "RECAST: Telling apart social and random relationships in dynamic networks," *Performance Eval.*, vol. 87, pp. 19–36, 2015.

[31] 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.

[32] F. A. Silva, A. Boukerche, T. R. Silva, L. B. Ruiz, and A. A. Loureiro, "A novel macroscopic mobility model for vehicular networks," *Comput. Netw.*, vol. 79, pp. 188–202, 2015.

[33] G. Casella and R. L. Berger, *Statistical Inference*, vol. 2. Pacific Grove, CA, USA: Duxbury, 2002.

[34] A. M. Law, W. D. Kelton, and W. D. Kelton, *Simulation Modeling and Analysis*, vol. 2. New York, NY, USA: McGraw-Hill, 1991.

[35] L. Bedogni, M. Gramaglia, A. Vesco, M. Fiore, J. Hrri, 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.

[36] K. Zhang, Y. Mao, S. Leng, Y. He, and Y. ZHANG, "Predictive offloading in cloud-driven vehicles: Using mobile-edge computing for a promising network paradigm," *IEEE Veh. Technol. Mag.*, vol. 12, no. 2, pp. 36–44, Jun. 2017.

[37] Y. Wu, J. Chen, L. P. Qian, J. Huang, and X. S. Shen, "Energy-aware cooperative traffic offloading via device-to-device cooperations: An analytical approach," *IEEE Trans. Mobile Comput.*, vol. 16, no. 1, pp. 97–114, Jan. 2017.

Authors' photographs and biographies not available at the time of publication.