Scale-Free Properties of Human Mobility and Applications to Intelligent Transportation Systems

Danielle L. Ferreira*, Bruno A. A. Nunes† and Katia Obraczka§
*Federal University of the State of Rio de Janeiro, Rio de Janeiro, RJ, Brazil
†University of California, San Francisco, CA, USA
§University of California, Santa Cruz, CA, USA
Email: danielle.ferreira@uniriotec.br, bruno.astutoarouchenunes@ucsf.edu and katia@soe.ucsc.edu

***Abstract*—Characterizing and modeling node mobility is of critical importance in building intelligent transportation systems and their applications. In this paper, we discuss the scale-free properties of some important human mobility characteristics, namely spatial node density and mobility degree, and show that they exhibit behavior that can be described by a power-law. Based on their power law characteristics, we derive analytical models for the spatial node density and mobility degree and showed that the data generated by the proposed analytical models closely approach empirical data extracted from the real mobility traces. Another contribution of our work is to use the proposed analytical models to build a synthetic mobility regime that is suitable for simulations of intelligent transportation systems. Finally, through network simulations, we show that ad-hoc network routing behavior under our mobility regime closely approximates routing behavior when the corresponding real trace is used.**

I. Introduction

As computing and sensing devices become more prevalent and embedded in everything around us and wireless communication more ubiquitous, they have enabled a variety of emerging applications such as Intelligent Transportation Systems, or ITS. According to the European Union's Directive 2010/40/EU [1], ITS embodies services that employ "information and communication technologies in the field of road transport, including infrastructure, vehicles and users, and in traffic management and mobility management, as well as for interfaces with other modes of transport". It also includes the use of information and communication technologies to improve public- and mass transit systems' efficiency and safety.

Understanding how people move in different environments and at different time scales is thus critical to enable ITS applications and services. The need for a deeper understanding of user mobility in wireless network environments has been well recognized (e.g., [2]) and has captured considerable attention from the networking community. CRAWDAD [3] is a notable example of an initiative funded by the US' National Science Foundation (NSF) whose goal is to make real traces of network user activity and mobility publicly available. However, even with such efforts, availability of real human and vehicular mobility traces is still quite limited and so is availability of real testbeds. As an alternative, a number of research efforts focus on extracting features from real mobility records (e.g., mobility traces) to build realistic mobility generators that will drive simulation platforms.

One of the main challenges in constructing mobility generators is developing models that can capture the complexity of human and vehicular mobility, and their key features, in real-world settings [4], [5], [6]. Two such key features are *clustering*, which can be defined as the tendency of people to agglomerate [7] and *geographical preference*, which refers to people's preferences for particular locales. The work in [8] proposes *spatial node density*, defined as the number of users located in a given unit area, as a way to measure the degree of clustering associated with a given user population. Spatial node density has considerable impact on fundamental network properties such as connectivity and capacity, which in turn have direct influence on core network functions like medium access and routing. Our prior work [8] showed that users tend to congregate and form clusters, rather than being homogeneously distributed over an area.

To date, only a few synthetic mobility regimes have attempted to model spatial node density. Some examples include [9], [10] and [11], which propose analytical models to study spatial node density under Random Waypoint (RWP) mobility. In [12], spatial node density has been modeled using first order ordinary differential equations (ODEs) whose parameters are extracted from real mobility traces. Using real traces to set values of model parameters is not ideal especially because of limited trace availability which may yield parameters that are specific to certain scenarios. In our previous work [13], we showed empirically, using real mobility traces collected in a variety of scenarios, that spatial node density and node mobility degree (i.e., the number of distinct cells visited by an individual node) observed in human mobility can be modeled by a power law. We then proposed a model to analytically describe the heavy-tail behavior exhibited by spatial node density and mobility degree resulting from user mobility, and confirmed that the proposed model closely approximates empirical spatial density distributions found in

This work was partially supported by grant CNS 1321151 from the US National Science Foundation and by the Brazilian Ministry of Education (under CAPES).

real mobility traces. As an example application of this analytical model, we used it to derive a mobility regime and showed how the proposed mobility regime closely resembles the real trace and the analytical model.

In this paper, we extended our previous work as follows: (1) We show how our mobility regime can be used to generate synthetic mobility traces for scenarios motivated by ITS applications; (2) We evaluate the fidelity of our mobility regime by comparing mobility features (i.e., spatial node density and mobility degree) from synthetic traces generated by the model against mobility traces generated using a well-known mobility trace generator; (3) We also evaluate how accurately our mobility regime reproduces user mobility characteristics, i.e., spatial density and mobility degree by conducting a comparative study using four well-known mobility regimes, namely: Random Waypoint mobility (RWP), Natural [14], Clustered Mobility Model (CMM) [15], and Self-similar Least Action Walk (SLAW) [16]. Our results show that our mobility regime is the one that most closely approximates the real trace; (4) Additionally, we expand our study of node mobility degree behavior and show that, similar to a campus scenario, mobility degree in a vehicular scenario also follows a Power Law; (5) We conduct a comparative study of the proposed mobility regime when evaluating network routing and show that routing exhibits comparable performance under our mobility regime when compared to the real trace. We also show that our model's fidelity to the real trace is considerably higher when compared to existing mobility regimes.

The remainder of this paper is organized as follows: the next section describes the mobility datasets used in our study. In Section III, we present our empirical study on the power law properties of human mobility. Section IV introduces our analytical model for spatial node density and mobility degree and verifies that it matches well the same metrics extracted from real mobility datasets. In Section V, we present the Scale-Free Mobility Regime (SFMR), a waypoint-based mobility regime capable of generating mobility traces whose spatial node density and mobility degree resembles closely the ones measured in real human mobility scenarios. In Section VI, we describe the experimental methodology we use to evaluate SFMR's performance and in Section VII, we present SFMR's performance results. Section VIII shows how to use SFMR to generate mobility traces for ITS-inspired scenarios and Section IX concludes the paper with some directions of future work.

II. Mobility Traces

In our study, we use real traces which are summarized in Table I in terms of number of users/nodes, trace duration, and data sampling period . These traces were collected in scenarios that are quite diverse: (1) *Quinta* [17] was collected is a city park in Rio de Janeiro, Brazil; (2) *Dartmouth* [18] logs user access to Dartmouth College's campus WLAN in the form of AP association and disassociation events (denoted as "A/D events" in Table I's Data Sampling column); (3) *SF Taxis* [19] refers to the vehicular mobility trace collected in the city of San Francisco, California, USA, where a fleet of approximately 500 taxi cabs was equipped with GPS trackers and had their positions logged for a period of 24 days. Note that two of the traces, i.e., *Quinta* and *SF Taxis* were collected using GPS devices, while the third one, textitDartmouth, logs user activity in a WLAN environment.

Trace	# users	# Cells	Duration	Sampling
Quinta [17] (GPS)	97	16	900s	1s
SF Taxis [19] (GPS)	483	1600	24 days	1 to 3 mins
Dartmouth [18] (WLAN)	6524	1776	60 days	A/D events

TABLE I
Summary of user mobility traces considered in our study.

a) **Cells:** The area in which mobile users move is divided into equal sized squares, or *cells*. When considering infrastructure-based wireless LAN (WLAN) traces, such as the Dartmouth trace, every cell corresponds to an AP. We employ similar criteria (AP average transmission range) for the GPS traces and in our experiments we used $140m$-by-$140m$ cells.

b) **Spatial Node Density and Mobility Degree:** *Spatial node density* is defined as the number of nodes located in a given cell while *Mobility degree* is the number of cells visited by a node.

c) **Node Speed and Pause Time:** We compute node speed as $\frac{d}{\Delta t}$ where d is the distance traveled between two consecutive entries in the GPS trace at times t_1 and t_2 and $\Delta t = t_2 - t_1$. Pause time is calculated for the Quinta trace [1] as $P = \Delta t$, if $d <$ *threshold*, or zero otherwise. We use *threshold*$= 0.5m$ since, in the Quinta trace, data is sampled every 1sec and pedestrians do not typically move much in 1sec.

III. Power Law and Human Mobility

In this section, we show that both spatial node density and mobility degree resulting from human movement in different scenarios exhibits heavy tail behavior. Power laws are expressions of the form $P(x) \propto x^{-\alpha}$, where α is a constant parameter and x are the measurements of interest. Few physical phenomena follow a power law for all values of x [20]. Usually, only the tail of the distribution, i.e., starting from a given minimum value, x_{min}, follows a power law. Thus, given a set of values that correspond to the observed data and the hypothesis that the data was extracted from a distribution that follows a power law, we want to verify if this hypothesis is plausible.

We fit the data from our mobility traces into a power law and compute its parameters by following the statistical framework described in [20]. We then apply a *goodness-of-fit* test also from [20], which generates a p value, used to test whether a distribution follows or not a power law distribution. In other words, the test checks if a distribution following a power law is a plausible fit for the empirical data. This test computes the distance between the empirical data distribution and the hypothesis of the

[1] We compute pause time for the Quinta trace as we will use for the experiments reported in Section VII-B. The Quinta trace was post-processed to account for possible GPS errors, as indicated in [17].

model. This distance is computed through the statistical test of Kolmogorov-Smirnov (KS), and is compared with the distance of measurements taken from a set of synthetic data drawn from the same model. The value of p is defined as a fraction of the distance of the synthetic data that is greater than the empirical distance. We use $p < 0.1$ [20] to reject the hypothesis that the empirical data follows a power law.

A. Spatial Node Density

This section presents the hypothesis test that the spatial node density is well represented by a power law distribution. Figures 1(a), 1(b) and 1(c) shows the cumulative distribution functions (CDFs) of spatial node density for the Quinta, Dartmouth, and the San Francisco Taxis traces respectively, along with the fitting of the data according to a power law. For the sake of comparison, these figures also plots the fitting of the same data using the exponential and log-normal distributions, as suggested in [20]. This is done in order to ensure that not only a power law distribution is a good fit for the data, but also provides better fit when compared to other distributions.

Additionally, the graphs in Figure 1 show the values for the parameters of the fitted curves. It also shows the values of p for the power-law fit for all three traces studied. We observe that p is well above the reference threshold of 0.1 used in [20] for all three traces, validating the hypothesis that the spatial node density distribution follows a power law with parameters α and x_{min} approximately equal to 2.5 and $10 - 20\%$ of the upper density, respectively.

As pointed out in the previous section, the value x_{min} which determines where the heavy tail behavior begins is sometimes imprecise. In our experiments we found that this value ranges from 10% to 20% of the upper density (i.e., the maximum value of density measured). These findings are consistent with the well known "80/20" rule [21].

Here, the exponent α represents the slope of the curve, and can be extracted from the observed data by using the following formula [20]: $\alpha = 1 + n[\sum_{i=1}^{n} \ln \frac{x_i}{x_{min}}]^{-1}$, where x_i are the measured values of x, and n is the number of samples above x_{min}.

The parameters of the exponential and log-normal distributions were extracted from the data set by fitting the best curve that minimizes the distance to the real data, using Matlab's fitting toolbox. Table II compares the fitting errors between the different distributions (i.e., power law, exponential, and log normal) and the traces. The power law distribution fitting yields errors at least 2 orders of magnitude smaller than the fittings using the other distributions.

B. Mobility Degree

Node mobility degree, or the number of different locations or cells that a node visits, is another important factor in mobile networks. For example, in disruption-tolerant networks (DTNs) or social networks, a node's degree of mobility will directly affect the node's node ability to relay messages since a node that visits a greater number of locations would potentially have more opportunities of contacts with other nodes. Thus, mobility degree can be used to decide whether a node is a good candidate to act as a message relay and/or how many copies of a message the node should carry.

Distribution	**SF Taxis** (density)	**Quinta** (density)	**Dartmouth** (density)
Power Law	**2.6306e-06**	**5.7428e-04**	**7.8624e-06**
Exponential	0.0173	0.0390	0.00380
Log-normal	0.0154	0.0192	8.4840e-04

TABLE II
MEAN SQUARE ERROR RESULTING FROM POWER-LAW, EXPONENTIAL, AND LOG-NORMAL FITTING OF THE TRACES' SPATIAL NODE DENSITY.

By applying the same method used in Section III-A, we show that the cumulative distribution of the number of distinct locations visited by a node also presents a heavy tail behavior, i.e., the hypothesis that node mobility degree follows a power law distribution is also plausible.

Figure 1 shows the CDFs of the distributions of the number of cells visited by users for the Dartmouth (Figure 1(d)) and SF Taxi traces (Figure 1(e)), along with the fitting of the data according to a power law, exponential, and log-normal distributions [2]. Here we can also observe that the curve that approaches the real data the most is the power law fit, which attests to the fact that most users tend to have low mobility or be stationary, while a small portion of users are highly mobile and visit a large number of locations. Table II shows the mean square error of each fit for the spatial node density metric, regarding Dartmouth and SF Taxi traces. Similar to the spatial node density results, the power law distribution also shows fitting errors for mobility degree at least 2 orders of magnitude smaller than the other distributions for both Dartmouth ad SF Taxis traces, as can be observed in Table III.

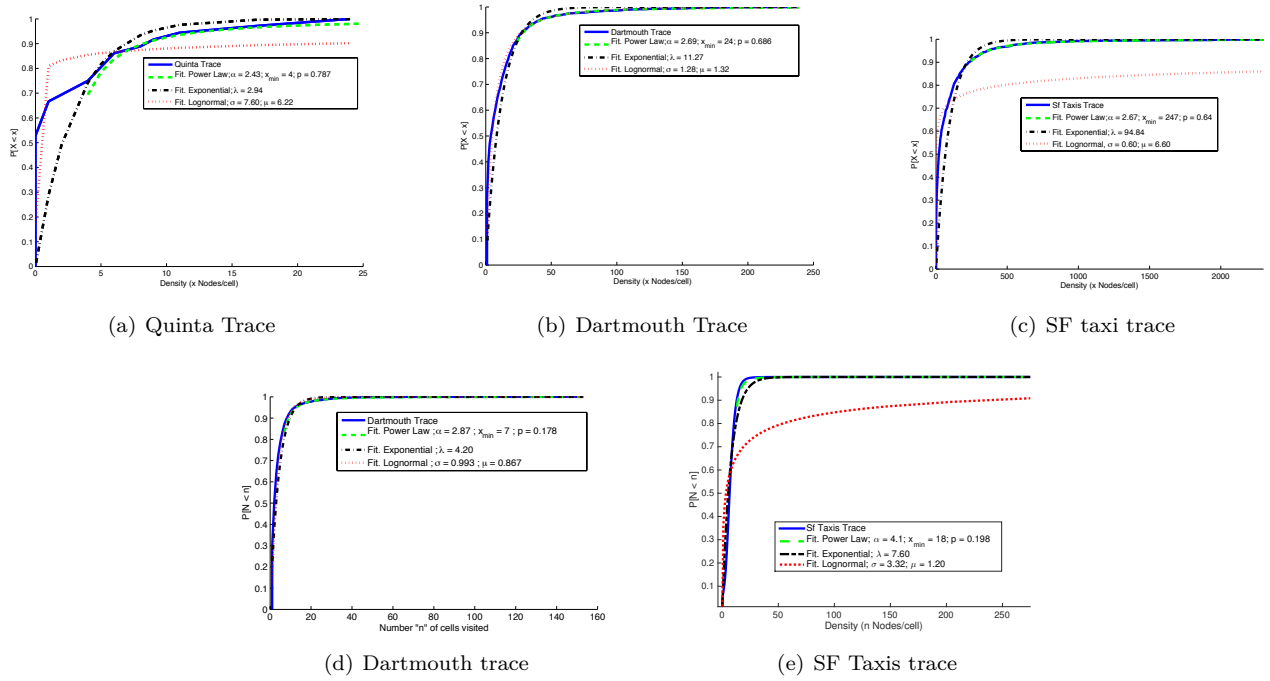
Distribution	**SF Taxis** (mob. degree)	**Dartmouth** (mob. degree)
Power Law	**6.0948e-05**	**5.8607e-05**
Exponential	0.0025	0.0022
Log-normal	0.0461	0.0014

TABLE III
MEAN SQUARE ERROR RESULTING FROM POWER-LAW, EXPONENTIAL, AND LOG-NORMAL FITTING OF THE TRACES' NODE MOBILITY DEGREE.

IV. Scale-Free Stochastic Model

We propose in this section an analytical model, named Scale-Free Stochastic Mobility (SFSM), which is based on the spatial node density and node mobility degree power-law behavior shown in Section III. SFSM's contributions include the ability to: (1) express analytically these key features of human mobility which explains the formation and maintenance of clusters, and (2) generate mobility regimes that follow the observed power-law behavior of user mobility in real scenarios without the need to extract parameters from real traces. In Section V, we exemplify SFSM's latter contribution by presenting an SFSM-based mobility regime.

[2]Since nodes in the Quinta trace visit a relatively small number of locations, the trace does not exhibit enough mobility to be statistically representative of node mobility degree. As such, we do not use the Quinta trace in our mobility degree characterization.



(a) Quinta Trace (b) Dartmouth Trace (c) SF taxi trace

(d) Dartmouth trace (e) SF Taxis trace

Fig. 1. CDFs of the spatial node density (a), (b), (c) and node mobility degree (d), (e) distributions for the traces.

A. Spatial Node Density as a Stochastic Process

Motivated by the empirical results presented in the previous section we now seek to model the spatial node density by means of a stochastic process. To this end, we divide the cells in groups such that cells with the same number of nodes belong to the same group. Then, we find the transition probabilities for a cell to migrate from its current group to another, either denser or sparser, group. These transition probabilities allow us to derive the node density distribution in the cells. In [22], a similar model was presented for modeling the income of people living in the UK in the early 50's.

We consider that the spatial node density distribution of countable groups of cells follow a stochastic process, and the stochastic matrix remains constant over time. In such context and provided certain specific conditions discussed below are satisfied, the distribution will tend towards an equilibrium distribution dependent on the stochastic matrix but not on the initial distribution. Table IV summarizes SFSM's notation.

We assume that cell density, i.e. the number of mobile users populating a cell, is divided into a number of proportionally distributed ranges. For example, we consider ranges per time interval to be $[1, 2)$ nodes, $[2, 4)$ nodes, $[4, 8)$ nodes, $[8, 16)$ nodes, and so forth.

We use smaller ranges for lower density values and larger ranges for higher density values, due to the fact that higher densities do not occur as frequently. This is a reasonable assumption since sparse cells occur in much greater numbers than dense cells, i.e. it is not uncommon for a small subset of the cells to account for most of the nodes in the entire network.

We then consider that the change in node density distribution in any individual cell in a given interval depends on its state in the previous interval and on a random process. In other words, we consider node density variation across these ranges as being a stochastic process. In fact, as users move, there are always new users coming into some cell and other users leaving. An acceptable assumption to make is that for each user leaving a cell, there is a cell welcoming that user in the next instant of time, and vice-versa. This assumption will imply that cell density is approximately constant over time and that each mobile node decides where and when to move. We also assume that the total number of cells in the system does not change with time as the region under study remains fixed.

Param.	Description
X_r	number of cells in each range R_r
X_s	number of cells in each range R_s
$p_{rs}(t)$	probability of cell in range R_r who shifts to range R_s
$p_{ru}(t)$	ratio of cells in range R_r that jumps u ranges
b	root of g(z)
N	total number of cells
y_{min}	lowest cell density
y_s	lower bound of the number of cells in range R_s
10^h	extent of each range
$\mathrm{F}(y_s)$	distribution of the number of cells exceeding y_s

TABLE IV
SUMMARY OF SFSM NOTATION.

Under such assumptions, to describe the spatial node density distribution, we first define $X_r(0)$ as the number of cells in each range R_r, $r = 0, 1, 2, ...$ at initial time T_0, and a series of matrices $p'_{rs}(t)$ as the probability of cells of R_r at time T_t who are shifted to range R_s in the next interval time T_{t+1}. Then, the density distribution $x_r(t)$ will be generated according to Equation (1).

$$X_s(t+1) = \sum_{r=0}^{\infty} X_r(t) p'_{rs}(t) \tag{1}$$

If we consider that the ranges are sorted by size, where the lowest cell density range is R_0, then we are able to define a new set of stochastic matrices

$$p_{ru}(t) = p^{'}_{r,r+u}(t) \tag{2}$$

and rewriting Equation (1) as

$$X_s(t+1) = \sum_{u=-\infty}^{s} X_{s-u}(t) p_{s-u,u}(t) \tag{3}$$

$p_{ru}(t)$ carries the information on the ratio of cells in range R_r which jumps a number u of ranges in T_t. As such, the frequency distribution of $p_{ru}(t)$ in u, is likely to be centered around $u = 0$.

In practice, this implies that the probability of cells shifting upwards and downwards across density ranges changes very little over time. We thus keep $p'_{r,r+u}(t) = p_{ru}(t)$ constant over time.

Given the discussion above, let us assume that, for all values of t and r, and for some fixed integer n, we have

$$p^{'}_{r,r+u}(t) = p_{r,u}(t) = 0 \quad \text{if} \quad u > 1 \quad \text{or} \quad u < -n \tag{4}$$

i.e., no cell can move upwards by more than one range or downwards by more than $n >= 1$ ranges at a time.

$$\begin{aligned} p^{'}_{r,r+u}(t) = p_{r,u}(t) = p_u > 0 \\ -n =< u =< 1 \quad \text{and} \quad u > -r \end{aligned} \tag{5}$$

Equation (5) is our basic postulate, which follows from our findings from Section III-A, that has tested the hypothesis that spatial node density follows a power law. What Equation (5) tells us is that the probabilities of a cell shifting up and down along the ranges of cell densities are distributed independently of the current cell density. This is true despite the imposed threshold forbidding that a cell descends below a given number of ranking positions and the frequency distribution of $p_{rs}(t)$ assumption discussed above. This will lead to a density distribution which obeys a Pareto's law, at least asymptotically, for high cell density values.

We also need to assume that for every value of r and t

$$\sum_{s=0}^{\infty} p'_{rs}(t) = \sum_{u=-r}^{\infty} p_{ru}(t) = 1 \tag{6}$$

which, according to (5), also implies

$$\sum_{u=-n}^{1} p_u = 1 \tag{7}$$

The assumption described by Equation (6) tells us that cell density preserve their identity over time, as described in Section IV-A above.

We also need to make sure that the cell density process is not dissipative. In other words, cell density does not increase indefinitely without reaching an equilibrium distribution. We can then denote

$$g(z) \equiv \sum_{u=-n}^{1} p_u z^{1-u} - z \tag{8}$$

Thus, our stability assumption is as follows:

$$g^{'}(1) \equiv -\sum_{u=-n}^{1} u p_u \quad \text{is positive.} \tag{9}$$

This means that for all cells, initially in any one of ranges $R_n, R_{n+1}, R_{n+2}...$, the average number of ranges shifted during the next time is negative.

Now we determine the equilibrium distribution corresponding to any matrix $p^{'}_{r,r+u}(t) = p_{r,u}(t)$ according to our assumptions. Owing to the uniqueness theorem mentioned above in Section IV-A, it will be sufficient to find any distribution which remains exactly unchanged under the action of the matrix $p'_{rs}(t)$ over time. Such distribution, when found, must be (apart from an arbitrary multiplying constant) the unique distribution which will be approached by all distributions under the repeated action of the matrix multiplier $p'_{rs}(t)$ over time.

If X_s is the desired equilibrium distribution, we need by (2), (4), (5)

$$X_s = \sum_{u=-n}^{1} p_u X_{s-u} \quad \text{for all} \quad s > 0 \tag{10}$$

and

$$X_0 = \sum_{u=-n}^{0} q_u X_{-u} \quad \text{where} \quad q_u = \sum_{v=-n}^{u} p_r \tag{11}$$

We need only satisfy (10), since (10), (4), (5) and (6) ensure the satisfaction of (11) as well. Now a solution of (10) is

$$X_s = b^s \tag{12}$$

where b is the real positive root other than unity. of the equation

$$g(z) \equiv \sum_{u=-n}^{1} p_u z^{1-u} - z = 0 \tag{13}$$

where $g(z)$ was already defined in (8). Descartes' rule of signs establishes that (13) has no more than two real positive roots: since unity is one root, and $g(0) = p_0 > 0$, and $g'(1) > 0$ by (9), the other real positive root must satisfy

$$0 < b < 1 \tag{14}$$

Hence (12) implies a total number of cells by

$$N' = \frac{1}{1-b} \tag{15}$$

and, to arrange for any other total number N, we need merely modify (12) to the form

$$X_s = N(1-b)b^s \tag{16}$$

We can now assume that the proportionate extent of each range is 10^h, and that the lowest cell density is y_{min}, then X_s is the number of cells in the range R_s whose lower bound is given by

$$y_s = 10^{sh} y_{min} \quad \text{from where} \quad log_{10} y_s = sh + log_{10} y_{min} \tag{17}$$

By summing a geometrical progression, using (16), we now find that in the equilibrium distribution of the number of cells exceeding y_s is given by

$$F(y_s) = N.b^s \quad \text{from where} \quad log_{10} F(y_s) = log_{10} N + s.log_{10} b \tag{18}$$

Now put

$$\alpha = log_{10} b^{-1/h} \quad \text{and} \quad \gamma = log_{10} N + \alpha log_{10} y_{min} \tag{19}$$

Then it follows from (17) and (18) that

$$log_{10} F(y_s) = \gamma - \alpha log_{10} y_s \tag{20}$$

This means that for $y = y_0, y_1, y_2...$, the logarithm of the number of cells exceeding y is a linear function of y. This states Pareto's law in its exact form [20].

Thus, if all ranges are equal proportionate extent, our simplifying assumptions ensure that any spatial node density initial distribution will, with time, approach the exact Pareto distribution given by Equations (19) and (20).

We validate the proposed SFSM model for spatial node density empirically by comparing it with mobility recorded in the Quinta, Dartmouth, and SF Taxis traces (summarized in Section II). The graphs in Figures 2(a), 2(b) and 2(c) show, for each trace, the probability of finding a cell that was visited by y or more mobile users. They were computed by extracting the number of users visiting each cell during a given interval, i.e. $[800s, 900s]$ for the Quinta trace, and a random non-interrupted 24 hour interval for the Dartmouth and SF traces. These intervals were chosen based on results presented in [8], which show that node density distribution does not change over time.

These figures also shows the graphs obtained by running SFSM for each trace. The coefficients of the stochastic matrix (i.e., the probability p_u of a cell changing u ranges between two consecutive time intervals) used to parameterize SFSM were extracted from the traces so that we could compare to the empirical density and validate our model. The SFSM curves start at $x_{min} = 4, 24, 247$ for Quinta, Dartmouth and SF Taxis traces, respectively, and are derived in Section III and shown in Figure 1. To quantify SFSM's fidelity to the empirical spatial node density for values of density greater than a y_{min}, we define the *modeling error* as a perceptual difference between the distribution obtained from the real traces and the one computed from SFSM. In other words, the modeling error is calculated as the absolute difference between SFSM-derived spatial node density distribution and the distribution computed for the real trace, taken at each point in the x-axis in the tail of the distribution (i.e., $(> y_{min})$, divided by the corresponding value from the real trace density distribution. We computed the mean error and confidence intervals with a 95% confidence level for the three traces studied. We average the errors computed for all points in the horizontal axis for values $> y_{min}$. The mean error and confidence interval for the Quinta trace shown in Figure 2(a) are $0.16\%[0.15\%, 0.19\%]$, respectively. Figure 2(b) shows the Dartmouth trace results, for which the mean error and confidence interval are $1.17\%[1.38\%, 0.96\%]$, respectively, and Figure 2(c) shows results for the San Francisco Taxi dataset with mean error and confidence interval of $0.43\%[0.47\%, 0.38\%]$, respectively.

B. Mobility Degree as a Stochastic Process

Following the observation that, similarly to the spatial node density, mobility degree also exhibits power law behavior (see Section III), we follow the same methodology used in Section IV-A to derive a stochastic model for user mobility degree.

Recall that mobility degree is defined as the *number of cells visited by a mobile user over a given period of time*. As such, a user with low mobility visits a small number of cells, while a very mobile user visits a larger number of cells. In order to describe the mobility degree distribution, we define $\Theta_d(0)$, as the number $\Theta_d(0)$ of users in each mobility degree range D_d, $d = 1, 2, ...$ at the initial time T_0, and a series of matrices $p'_{dv}(t)$ as the probability of users in the range D_d at time T_t who shifted to range D_v in the following interval time T_{t+1}. Then, the mobility degree distribution $\theta_d(t)$ will be generated according to

$$\Theta_v(t+1) = \sum_{d=0}^{\infty} \Theta_d(t) p'_{dv}(t) \tag{21}$$

Just as we did before, consider that the ranges are ordered by their size, where the lowest range of number of cells visited per user is C_0, then we can define a set of stochastic matrices such as

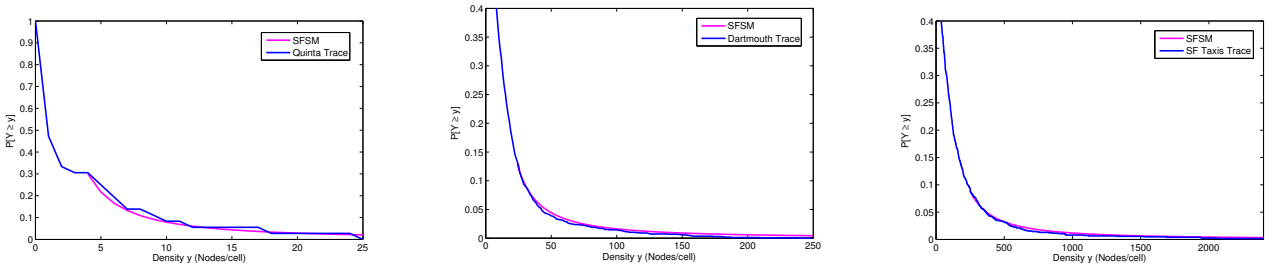
$$p_{df}(t) = p'_{d,d+f}(t) \tag{22}$$

where $p_{df}(t)$ indicates the ratio of users in D_d who jumps over a number f of ranges in T_t. Then, Equation (21) becomes:

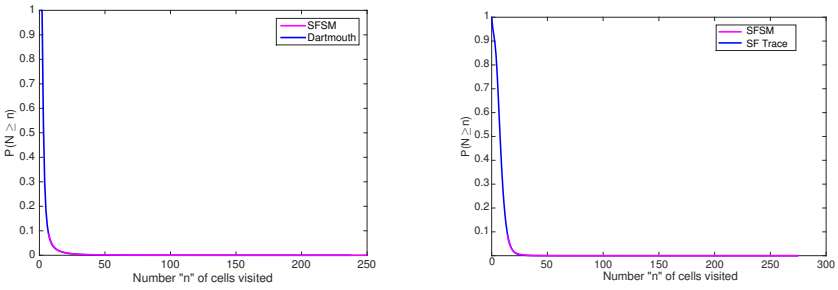
$$\Theta_v(t+1) = \sum_{f=-\infty}^{v} \Theta_{v-f}(t) p_{v-f,f}(t) \tag{23}$$

Following analogous derivations as in Section IV-A, we are able to find the equilibrium distribution $F(\omega_v)$ of the number of users whose number of visited cells exceeds ω_v.

We validate the proposed SFSM model for node degree distribution empirically by comparing it with mobility recorded in the Dartmouth, and SF Taxis traces. Figures 2(d) and 2(e) show the probability of a node visiting n or more cells in a single trip, and by running SFSM for each trace. They were computed by counting the number of cells each upropmted user visits during the trace duration. The coefficients of the stochastic matrix (i.e., the probability p_f of a user changing f ranges between two consecutive time intervals) used to parameterize SFSM were extracted from the traces so that we could compare to the empirical density and validate our model.



(a) Quinta trace and SFSM. (b) Dartmouth trace and SFSM. (c) SF Taxi trace and SFSM.



(d) Dartmouth trace and SFSM. (e) SF Taxis trace and SFSM.

Fig. 2. Spatial node density (a),(b),(c) and mobility degree (d),(e) distributions extracted from the mobility traces compared against distributions generated by SFSM.

V. Generating Scale-Free Mobility Regimes

Intelligent Transportation Systems have leveraged research and technology motivated by vehicular ad-hoc networks, or VANETs. In fact, many ITS services rely on the provision of an effective communication platform between vehicles, as well as between vehicles and road infrastructure (e.g., road-side units, sensors, etc). Also, communicating devices, such as laptops, smart phones, and even sensors now often carried by drivers and passengers can also be used to track vehicle mobility which is influenced by how humans move, their habits, social links, and locality [23]. It is known that in the real world, nodes present clustering behavior and community structure [7], with islands of connectivity and paths between clusters. For example, in VANETs, vehicles tend to group around traffic lights, junctions, toll, hazards, etc. The same behavior is also found in human mobility, where they tend to group in popular places, such as classrooms or cafeterias on campus, popular events, cafes, restaurants, etc.

As it is usually expensive and often logistically difficult to deploy and test ITS solutions in real world environments, network researchers and practitioners rely on simulation tools in order to develop and evaluate ITS services. Moreover, since we would like to be able to simulate realistic scenarios, mobility regimes that can closely represent real-world mobility are imperative in assessing the true impact and performance of ITS applications and protocols. In this section, we introduce the Scale-Free Mobility Regime (SFMR) that considers the previously discussed stochastic properties of node mobility, namely spatial node density and mobility degree, as well as nodes' geographical preferences. SFMR generates mobility regimes that reflect realistic human mobility behavior as characterized in Section III. Next, we show how to use the Scale-Free Stochastic Model (SFSM) proposed in Section IV to set SFMR's parameters.

In a nutshell, using SFMR to generate realistic mobility regimes works as follows: Before the simulation begins, cells with high node density (or clusters) are defined by specifying that the spatial node density in these cells is greater than a given threshold y_{min}; in other words, for these high density regions, we use the tail of the spatial density distribution to derive the probability that a node will choose a cell in the region. In the case of cells where density is below the y_{min} threshold, we apply an uniform spatial density distribution, for simplicity. As shown in Section VII, our results indicate that uniform spatial node density is a reasonable approximation for low density regions. As part of our ongoing work, we have been studying more closely the impact of different known distributions to model cell density bellow y_{min}.

As we have previously discussed, one of SFMR's benefits is the ability to generate mobility regimes that result in spatial density distributions similar to the ones found in real mobile applications (as exemplified by the traces presented in Section II) without the need to extract parameters from mobility traces. Below we provide a detailed description of SFMR, including how to set its parameters.

SFMR has two phases, namely initialization and movement. During the initialization phase (shown in Algorithm 1), nodes can be distributed in the geographic area according to an arbitrary‘ distribution. In the movement phase, for simplicity, we use a waypoint-based mobility regime, contending that simplicity is critical for wide adoption of

any mobility regime. As such, the steps involved in the movement phase, as shown in Algorithm 2.

During initialization, described in Algorithm 1, some node l may decide with probability $1 - P(\eta_l)$ if it will remain in the same cell, or if it will choose a destination with another cell with probability $P(\eta_l)$. The number of different cells η_l visited by node l is defined *a priori* by sampling from the computed distribution $F(\omega_v)$. $F(\omega_v)$ can be obtained as described in Section IV-B. The probability $P(\eta_l)$ that a user l will leave a cell is computed in Equation 24, and this value of $P(\eta_l)$ is kept constant for every node l during the simulation.

$$P(\eta_l) = \frac{\eta_l}{\sum_m \eta_m} \forall m \in \{1..L\} \tag{24}$$

When the simulation is in the movement phase, nodes behave as described in Algorithm 2. For every node, using a probability distribution given by $F(\omega_v)$, the node decides with probability $P(\eta_l)$ if it is going to move to another cell, as mentioned earlier. If the node decides to move, it chooses its next cell using a probability distribution given by $F(y_s)$. A (x, y) destination is picked randomly inside the chosen cell. Then the node moves to that destination at a randomly chosen speed, uniformly distributed between $[V_{min}, V_{max}]$. When the node reaches its destination it pauses for some time, and repeats. We discuss how the values for V_{min}, V_{max}, and pause time are chosen below.

The decision of which cell is going to be the next destination is made with probability $P(\mu_i)$. We assume that the probability $P(\mu_i)$ that a node would choose cell i as a next destination depends on the cell intensity μ_i, that can be obtained by sampling from the computed distribution $F(y_s)$, of every cell i. The probability $P(\mu_i)$ is computed as in Equation 25, and this value of $P(\mu_i)$ (i.e., the probability that cell i is chosen, given its intensity μ_i) is kept constant for each cell i during the simulation. Table V summarizes SFMR's notation.

$$P(\mu_i) = \frac{\mu_i}{\sum_j \mu_j}, \forall j \in \{1..N\} \tag{25}$$

Param.	Description
F_{y_s}	Distribution of the numb. of cells exceeding y_s
F_{ω_v}	Distribution of the numb. of cells visited by a user that exceeds ω_v
ν_l	Numb of different cells visited by node l
μ_i	Numb of mobile nodes that visits a cell i
P_{ν_l}	Prob. that node l chooses to leave a cell
P_{μ_i}	Prob. that a node chooses cell i as destination
y_{min}	Lowest cell density

TABLE V
SUMMARY OF SFMR NOTATION.

As discussed previously, it is worth pointing out that the parameters for the proposed mobility regime do not need necessarily to be extracted from real mobility traces. In fact, the model parameters can be set and tuned in order to generate a variety of mobility scenarios in terms of number of clusters, their size, as well as the nodes' mobility degree. In the proposed model we need to set only 4 parameters, namely the speed range, pause time range, y_{min}, and the set of coefficients for the generating function in Equation 8. The tuning of these parameters will depend on the parameters for the scenario itself (e.g. total area, cell size, number of nodes, cluster size, etc). For the simulation results presented in the next section, we extracted the parameters from the traces for the sake of having a baseline (i.e., a real trace scenario) for a fair comparison of all the mobility regimes considered in our evaluation. That also shows that it is possible to mimic specific real world scenarios.

Algorithm 1 SFMR: Initialization phase

1: *Distribute L nodes over the simulation area according to any given distribution*
2: **for** *each node* **do**
3: *Attribute the node degree probability $P(\eta_l)$, drawn from $F(\omega_v)$*
4: **end for**

From the statistical study presented in Section III, y_{min} was found to typically fall between 10% to 20% of the largest cluster (the highest node density). The coefficients of Equation 8 can be set according to the shape of the target density curve, considering: (1) the sizes of the clusters one wants to simulate and (2) the total population of nodes, which will provide an estimate of how many clusters of each size can be simulated. Equation 8 depends on the probability matrix of cells changing to another range (higher or lower). Depending on the scenario we would like to simulate, this probabilities can be set differently. For dense scenarios, where clusters are fewer and larger, such probabilities should be higher. For sparser scenarios, on the other hand the probability of choosing a given cell should vary little over the range of i.

Algorithm 2 SFMR: Movement phase

1: **for** *each node* **do**
2: **if** *node decides to move to another cell with probability $P(\eta_l)$* **then**
3: *Select next cell with probability, $P(\mu_i)$,* drawn from $F(y_s)$
4: *Moves to destination using randomly speed between* $[V_{min}, V_{max}]$
5: *pauses for a pause-time*
6: **end for**

VI. Evaluation Methodology

We evaluate the proposed Scale-Free Mobility Regime (SFMR) in terms of how accurately it reproduces real user mobility according to spatial density and mobility degree when compared against real mobility traces. In our study, we also compare SFMR against four well-known mobility regimes, namely: Random Waypoint mobility (RWP), Natural [14], Clustered Mobility Model (CMM) [15], and Self-similar Least Action Walk (SLAW) [16]. Our rationale for choosing these mobility models for our comparative performance study of SFMR is as follows. RWP, despite its limitations, has been widely used to evaluate wireless networks and their protocols. Natural and CMM were selected as representatives of the class of mobility regimes that follow the preferential attachment principle. More recently proposed models have extended CMM, e.g., HCMM [24] and ECMM [25] but preserve CMM's core preferential attachment based features; as such we use CMM, along with Natural, to represent preferential attachment based mobility regimes in our comparative analysis. Similarly,

SLAW is a well-known, widely cited mobility regime that accounts for social structure and social features. SLAW has inspired and has been extended by successors like SMOOTH [26] and MobHet [27] which prompted us to select SLAW to represent mobility models that consider social interactions.

Additionally, we evaluate SFMR's fidelity to real user mobility by investigating how it affects network routing behavior, and consequently the efficiency of message dissemination in ITS, when compared to real mobility traces as well as to the mobility regimes listed above, namely Random Way-Point (RWP) mobility, Natural [14], Clustered Mobility Model (CMM) [15], and Self-similar Least Action Walk (SLAW) [16].

We conducted two types of simulations: (1) first, we modified the Scengen [28] scenario simulator to generate traces according to RWP (already implemented), Natural and CMM (implemented at Scengen), SLAW (MATLAB implementation), and SFMR (also implemented in Scengen). Once the simulator was able to generate the mobility traces we computed the spatial node density distribution results presented in Section VII-A. (2) in the second type of simulation experiments, once the synthetic mobility traces were generated as described above, these and the real traces were fed to the Qualnet network simulator [29] in order to evaluate their impact to core network functions, such as routing and message dissemination for example.

For the first type of experiments, in order to compare synthetic traces generated with RWP, Natural, CMM, SLAW, and SFMR to real user mobility traces, we adjusted the Scengen simulation parameters according to information extracted from the real trace for all mobility models. For example, velocity range $[v_{min}, v_{max}]$ is set such that average node velocity (assuming that the velocity of each node is randomly chosen from a uniform distribution of values between $[v_{min}, v_{max}]$), matches the average node velocity extracted from the trace. In particular, for the RWP regime, in order to address the steady-state stationarity problem reported in [30], we followed the recommendations mentioned in that work. More specifically, the velocity range was set to be $\pm$, the standard deviation measured in the real traces, around the measured average velocity. Then, velocities were chosen *uniformly* within that range in which the lower limit was greater than zero and where the mean matches the one measured in the real trace.

Similarly, the pause time was chosen uniformly in the range $[0, P_{max}]$, where the value of P_{max} is such that the average pause time matches the one measured in the real traces. The dimensions of the rectangular simulation area are set to be the same as in the traces. Moreover, in our simulation scenarios, we use the same initial positions for the nodes found in the real traces, except for SLAW which has its own initialization procedure.

In the RWP simulations using Scengen, a node's next destination (x_d, y_d) is randomly chosen over the simulated area according to a uniform distribution. For SFMR, the choice of (x_d, y_d) is given by Equation 25, where the intensity values μ are set by the initialization procedure as described in Section V. For Natural and CMM, the probability of choosing the next destination is computed "on-the-fly", based on the destination's popularity as described in [14].

For the second type of experiments, synthetic mobility traces generated using Scengen as described above, as well as the real traces were fed to the Qualnet network simulator [29]. As previously pointed out, efficient message dissemination is critical to road safety and transportation efficiency in ITS. Thus, the goal of these experiments is to evaluate how close to the real trace are the synthetic mobility regimes as far as their impact on routing and data dissemination.

Data traffic scenarios used in these experiments try to simulate nodes communicating with one another in ITS scenarios (e.g., vehicle-to-vehicle, vehicle-to-infrastructure). We use 20 Constant Bit Rate (CBR) flows between randomly chosen source-destination node pairs. Flows start at randomly chosen times and stay active during the course of the whole simulation generating traffic at a rate of 4 packets per second. We use the Ad-hoc On-Demand Distance Vector (AODV) [31] routing protocol, an Internet standard for routing in wireless multi-hop ad-hoc networks, and the IEEE 802.11g data link layer protocol with radio range of 150m and data rate of 54.0 Mbps. Table VI summarizes other simulation parameters used in these experiments.

Parameter	**Quinta**
Average Velocity ($\pm\sigma$)(m/s)	1.2 ($\pm$0.53)
Average Pause Time Duration (sec)	3.6
Area Dimensions (meters x meters)	840 x 840
Duration of Simulation (sec)	900
Number of users	97
Number of CBR flows	20

TABLE VI
Simulation parameters.

VII. Results

Results are reported here for the Quinta trace with a 90% confidence interval over 10 runs. For the runs using the real trace, since we cannot vary mobility, we randomize the traffic scenarios by varying the source and destination pairs of the flows in each of the 10 runs. The same traffic patterns were used to feed the RWP, Natural, CMM, SLAW and SFMR simulations, but in these cases, we generated 5 mobility traces with each model, giving a total of $10 \times 5 = 50$ simulation runs for each synthetic mobility regime.

A. Spatial Node Density

In order to study spatial node density behavior, we define the *Node density distribution* metric as the ratio of cells containing $\geq n$ nodes. Each curve in Figure 3 shows the density distribution for the Quinta trace and each mobility model, namely SFMR, RWP, Natural, CMM, and SLAW. The curves shows the distribution at the end of the trace collection interval, which is at 900 seconds for Quinta.

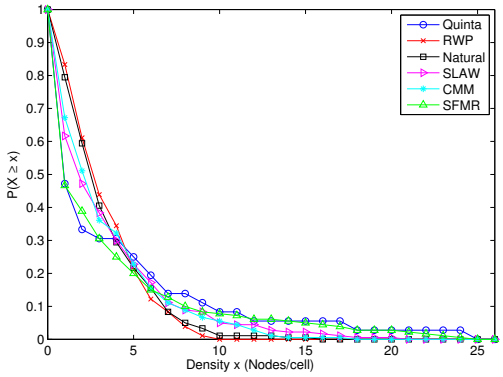


Fig. 3. Node Spatial Density Distribution.

From these plots we observe that SFMR's density distribution closely follows the distribution of the real trace. In the case of RWP, the majority of cells (i.e., more than 80%) present a similar number of nodes (i.e., one or more nodes), and no cells contain significantly greater concentration of nodes (i.e., no cell contains more than 9 nodes). This is also the case for Natural, CMM and SLAW. In order to quantitatively compare how close the node density distributions resulting from the synthetic mobility regimes are to the real trace, we compute the average normalized difference between the synthetic traces' spatial node density distribution and that of the real trace as follows: for each data point, we compute the absolute value of the difference between the density distribution resulting from the synthetic model and that of the real trace, divided by the latter. We average over all data points and Table VII reports these averages as well as lower and upper values of their 95% confidence interval. Table VII confirms that SFMR's spatial density distribution is the closest to the real trace's when compared to the other mobility regimes studied.

Mobility Model	Mean	Confidence Interval
SFMR	0.0161616	[0.00749751 0.0248257]
SLAW	0.0396465	[0.0234087 0.0558843]
CMM	0.0492424	[0.0282684 0.0702164]
Natural	0.070202	[0.0364066 0.103997]
RWP	0.813131	[0.0442555 0.118371]

TABLE VII
NORMALIZED DIFFERENCE BETWEEN THE SPATIAL DISTRIBUTION RESULTING FROM MOBILITY MODELS AND THE EMPIRICAL DISTRIBUTIONS COMPUTED FROM THE REAL TRACE: MEAN [LOWER UPPER] VALUES 95% CONFIDENCE INTERVAL.

B. Performance Evaluation of SFMR

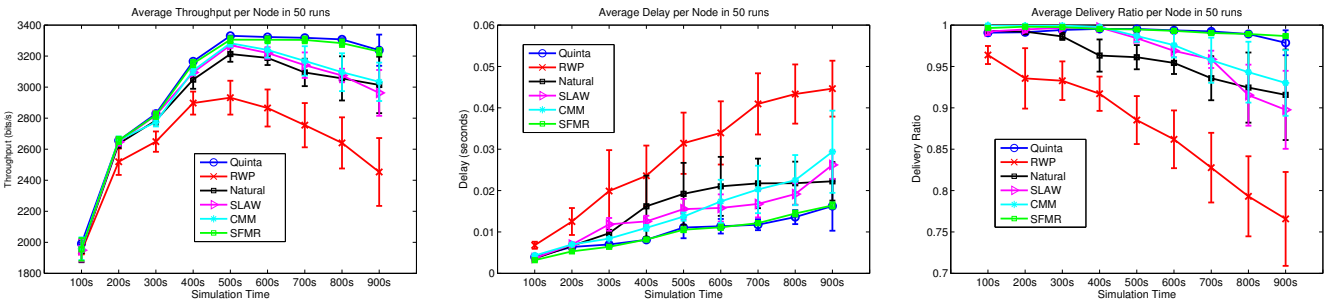
Mobility models are frequently used for simulation purposes when new communication-based vehicular and human mobile services are being investigated. One key factor researchers and developers must take into account when evaluating solutions through simulations of mobile scenarios such as V2V and V2I applications is realistic mobility patterns. In fact, mobility models play a vital role in determining the performance of various wireless mobile systems, such as Vehicular Ad-Hoc Network (VANET) [32], Wireless Sensor Network (WSN) AND Body Sensor Networks (BSNs) [33], etc. In ITS an efficient message dissemination scheme is critical to its applications, such as road safety and urban traffic status. Thus, in order to evaluate SFMR in such dynamic scenarios we focus on the study of the impact of different mobility models in an infrastructureless network, when compared to real mobility extracted from a real mobility trace.

We report results comparing performance for the AODV wireless ad-hoc network routing protocol under our mobility regime, the Quinta mobility trace, as well as mobility regimes proposed in the literature and discussed in Section VI. The objective here is not to evaluate a proper ITS system or a real application, but rather evaluate the ability of our proposed model to deliver realistic node movement and how a network simulation can be affected by realistic and non realistic mobility. We compute the following metrics in our study:

- *Throughput*: is defined as the total number of bytes received at the destination node divided by the time elapsed between the reception of the first byte of the first data packet and the reception of the last byte from the last data packet. This quantity is measured at all nodes and averaged before reported.
- *End-to-End Delay*: is measured as the time elapsed between the moment a packet is sent and the instant it is received at the destination. This quantity is then averaged for all packets transmitted by all nodes in the network.
- *Delivery Ratio*: is computed as the ratio between the total number of packets received by all nodes and the total number of packets transmitted by these nodes.

The above described metrics for throughput, delay, and delivery ratio are reported in Figures 4(a), 4(b) and 4(c) respectively, over time for the Quinta scenarios. There is a notable discrepancy between the results for the real trace and results for RWP. Also noteworthy is how the discrepancy widens over time which can be explained by RWP's inability to maintain the trace's spatial node density distribution over time which directly impacts routing performance. SFMR, on the other hand, allows the formation and preservation of clusters of nodes, which, in the case of this scenario, resembles closely the real trace curves. As the clusters are bigger for the realistic scenarios and SFMR, information delivery is also more efficient, as more nodes are closer together in the clusters.

In the case of Natural, CMM, and SLAW, we notice that routing performance under these mobility regimes stay close to the real trace up until around 300s for Natural and around 500s for SLAW and CMM. Up until then, the probabilities of choosing each cell are based on the initial non-uniform spatial densities, and the mobility regimes are capable of maintaining some level of node clustering. However, later in the experiment, nodes start to spread out as the probability of choosing a new cell starts approaching a uniform distribution. This behavior causes the clusters to



(a) Throughput (b) Delay (c) Delivery Ratio

Fig. 4. Network routing performance for the Quinta trace.

dissipate and routing performance starts to diverge from the real traces.

VIII. Generating ITS-Inspired Traces with SFMR

As previously pointed out, one of the distinguishing features of SFMR is its ability to generate mobility traces without the need to prime its parameters using existing traces. In this section, we demonstrate this feature of SFMR by using it to generate mobility traces for ITS-inspired scenarios.

A. Mobility in Urban Scenarios

Suppose we want to simulate mobility in an urban scenario, such as the downtown area of a large metropolitan region. We could then consider two different types of mobility, namely pedestrian- and vehicle mobility.

Spatial Density: Pedestrians tend to congregate in locations like malls, markets, cafes, schools, etc. Since pedestrian density tends to be relatively high in most downtown areas (e.g., compared to rural or even suburban areas), the mobility model used to represent spatial node density of pedestrians in urban centers could then be assigned a lower value for α. This means that the power-law curve representing spatial node density of pedestrians in downtown areas would have a longer tail to indicate that a relatively higher percentage of cells have higher concentration of nodes.On the other hand, if we are now interested in simulating vehicle mobility in a city center, we could consider fewer nodes (e.g., in some cities, only public transportation is allowed to circulate in the city's downtown area) compared to pedestrians. Assuming that public transportation vehicles are moving most of the time, except for high traffic congestion spots or bus depots, most cells would have lower concentration of nodes. As such, we could use a power-law distribution with longer tail, i.e., a higher value of α, to represent spatial density of vehicles in a city center.

Mobility Degree: To model pedestrian mobility degree, we would assume that most pedestrians would typically visit less cells due to their limited mobility and thus exhibit lower mobility degree relative to vehicles. This means that pedestrian's mobility degree would follow a power law that decays quickly, i.e., with higher α.

For vehicles, since they can cover longer distances and, as a result, visit more cells, the tail of the power law describing their mobility degree distribution would be longer when compared to pedestrians'.

B. Mobility in Suburban Areas

In the case of suburbs, we could still consider different mobility regimes for pedestrians and vehicles. However, unlike urban scenarios, suburbs are typically less densely populated and there are less people walking than driving.

Spatial Density: For pedestrians, there would likely be only a few areas with higher pedestrian density like parks and street malls, while most everywhere else would present low densities. As such, we could use a higher α value to simulate spatial density of pedestrian mobility in suburban areas.

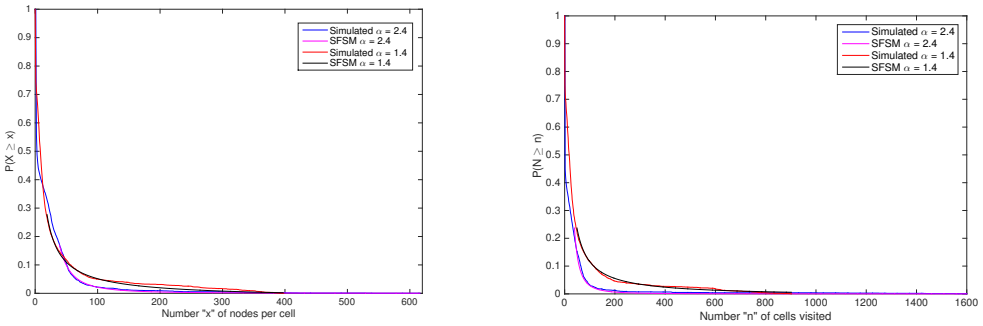
We could also envision similar behavior for the spatial density of vehicle mobility in suburban settings, i.e., that most cells will exhibit low vehicle density. As such, we could use higher α values to model vehicle spatial density in suburban scenarios.

Mobility Degree: In suburban areas, we could envision scenarios where a reasonable number of vehicles circulate only locally but a good number travels longer distances, e.g. when people commute to work. As such, we would use lower α values for the mobility degree power law distribution.

In the case of pedestrians, we may consider people spending most of their time inside their property and going out to move around the streets for a few sporadic activities, e.g., jogging, walking the dog, go to the playground or store close by. For that reason, we would recommend using a higher value of α for simulating pedestrians in this conditions.

C. Sample ITS Mobility Regime

Here we use a sample ITS-inspired mobility scenario to illustrate how SFMR can be used to generate synthetic mobility traces without the need to extract parameters from existing traces. The goal is to show how to use SFMR to simulate a given ITS scenario and validate the resulting spatial density and mobility degree distributions



(a) Spatial Density. (b) Mobility Degree.

Fig. 5. Mobility degree and spatial density distribution for ITS-inspired mobility regime generated by SFMR and SFSM.

by comparing them to the ones obtained using our analytical model SFSM derived in Section IV. We use our implementation of SFMR on the Scengen [28] simulator to generate SFMR mobility traces.

In particular, this example simulates 3,000 vehicles moving around a large metropolitan region of size 8km-by-6km. Vehicle speeds vary uniformly over a range of 15 to 40 km/h [3]. The duration of the simulation is set to 100.000 seconds (i.e., around 27 hours, or a little more than a day). We wanted to keep the network always mobile and for that reason we set pause time to be 0 at all times. Table VIII summarizes the simulation parameters and their values. We simulated two scenarios by essentially changing the value of α. We first use an alpha of 1.4 for both mobility degree and density, and then increase α to 2.4.

Parameter	**Value**
Velocity Range (km/h)	[15 40] uniform
Average Pause Time Duration (sec)	0
Area Dimensions (meters x meters)	8000 x 6300
Duration of Simulation (sec)	100000
Number of nodes	3000
α for Mobility Degree	1.4 and 2.4
α for Spatial Density	1.4 and 2.4

TABLE VIII
SFMR PARAMETERS AND THEIR VALUES FOR SAMPLE ITS MOBILITY REGIMES.

The data points in the SFMR curves in Figure 5 are averaged over 20 simulation runs; the graphs also show the SFSM with the previously mentioned values of α. Figure 5(a) shows the spatial density distribution for two values of α. In the example scenario described in Section VIII-B above, most cells present low density of vehicles with a small number of cells exhibiting high vehicle densities (e.g., shopping malls, supermarkets, school campuses, etc); we would use $\alpha = 2.4$ in this case. The value of x_{min} was set to 45 for SFSM with $\alpha = 2.4$. The value of x_{min} was then set to 25 in the case of SFSM with $\alpha = 1.4$. We observe that both curves match closely the SFSM curves.

One of the curves in Figure 5(b), i.e., the one with $\alpha = 2.4$, shows an example of high mobility degree where few mobile nodes visit > 1500 cells. This mobility degree behavior can mimic the behavior of vehicles in a city center as described in Section VIII-A. When $\alpha = 1.4$ the decay of the curve is slower and more nodes have lower and more uniform mobility degrees, meaning that 25% of the nodes visit from 85 (x_{min}) to 900 cells. This could be true if we wanted to simulate for example, vehicles moving on the suburban neighborhood scenario mentioned before in Section VIII-B.

[3]These parameter values were set based on real scenarios as reported in "http://infinitemonkeycorps.net/projects/cityspeed/"

IX. Conclusion

In this paper, we showed the scale-free properties of some important human mobility characteristics, namely spatial node density and mobility degree. In our study we analyzed a set of real mobility traces collected in diverse scenarios motivated by ITS, namely a city park, a University campus, and taxis in the downtown area of a major city. We demonstrated that both spatial node density and mobility degree exhibit power law behavior which then allowed us to derive analytical models for these two mobility features. We showed that the proposed analytical model closely matches the empirical data extracted from the real mobility traces. Another contribution of our work was to use the proposed analytical models for spatial node density and mobility degree to build a waypoint-based mobility regime capable of generating synthetic mobility traces whose spatial node density and mobility degree closely resembles the ones measured in real human mobility scenarios. As such, the proposed mobility regime can be employed to test and evaluate ITS services and protocols. Finally, using a network simulator, we evaluated a wireless ad-hoc network routing protocol and showed that its performance under our mobility regime and under the real trace is very similar.

References

[1] "Directive 2010/40/eu of the european parliament and of the council of 7 july 2010 on the framework for the deployment of intelligent transport systems," http://data.europa.eu/eli/dir/2010/40/oj.
[2] M. Conti and S. Giordano, "Mobile ad hoc networking: milestones, challenges, and new research directions," *IEEE Communications Magazine*, vol. 52, pp. 85–96, January 2014.
[3] CRAWDAD, "http://crawdad.cs.dartmouth.edu/."

[4] D. Karamshuk, C. Boldrini, M. Conti, and A. Passarella, "Human mobility models for opportunistic networks," *IEEE Communications Magazine*, vol. 49, pp. 157–165, December 2011.

[5] V. F. Mota, F. D. Cunha, D. F. Macedo, J. M. Nogueira, and A. A. Loureiro, "Protocols, mobility models and tools in opportunistic networks: A survey," *Computer Communications*, vol. 48, no. Supplement C, pp. 5 – 19, 2014.

[6] M. Lin and W.-J. Hsu, "Mining gps data for mobility patterns: A survey," *Pervasive and Mobile Computing*, vol. 12, no. Supplement C, pp. 1 – 16, 2014.

[7] M. Newman, "Detecting community structure in networks," *The European Physical Journal B - Condensed Matter and Complex Systems*, vol. 38, no. 2, pp. 321–330, 2004.

[8] B. A. A. Nunes and K. Obraczka, "On the invariance of spatial node density for realistic mobility modeling," in *Proceedings of the 2011 IEEE 8th International Conference on Mobile Ad-Hoc and Sensor Systems*, pp. 322–331, 2011.

[9] E. Hyytia, P. Lassila, and J. Virtamo, "Spatial node distribution of the random waypoint mobility model with applications," *IEEE Transactions on Mobile Computing*, vol. 5, pp. 680–694, June 2006.

[10] D. Mitsche, G. Resta, and P. Santi, "The random waypoint mobility model with uniform node spatial distribution," *Wireless networks*, vol. 20, no. 5, pp. 1053–1066, 2014.

[11] C. Bettstetter, G. Resta, and P. Santi, "The node distribution of the random waypoint mobility model for wireless ad hoc networks," *Mobile Computing, IEEE Transactions on*, vol. 2, no. 3, pp. 257 – 269, 2003.

[12] B. A. A. Nunes and K. Obraczka, "A framework for modeling spatial node density in waypoint-based mobility," *Wireless Networks*, vol. 20, no. 4, pp. 775–786, 2014.

[13] D. L. Ferreira, B. A. A. Nunes, and K. Obraczka, "On the heavy tail properties of spatial node density for realistic mobility modeling," in *Sensing, Communication, and Networking (SECON), 2014 Eleventh Annual IEEE International Conference on*, pp. 504–512, June 2014.

[14] V. Borrel, M. D. de Amorim, and S. Fdida, "On natural mobility models," in *WAC*, 2005.

[15] S. Lim, C. Yu, and C. Das, "Clustered mobility model for scale-free wireless networks," in *Local Computer Networks, Proceedings 2006 31st IEEE Conference on*, 2006.

[16] K. Lee, S. Hong, S. J. Kim, I. Rhee, and S. Chong, "Slaw: A mobility model for human walks," in *Proceedings of IEEE INFOCOM*, 2009.

[17] C. Campos, T. Azevedo, R. Bezerra, and L. de Moraes, "An analysis of human mobility using real traces," in *Proceedings of the 2009 IEEE WCNC*, 2009.

[18] D. Kotz, T. Henderson, I. Abyzov, and J. Yeo, "CRAWDAD data set dartmouth/campus (v. 2009-09-09)." Downloaded from http://crawdad.cs.dartmouth.edu/dartmouth/campus, Sept. 2009.

[19] M. Piorkowski, N. Sarafijanovic-Djukic, and M. Grossglauser, "CRAWDAD data set epfl/mobility (v. 2009-02-24)." Downloaded from http://crawdad.cs.dartmouth.edu/epfl/mobility, Feb. 2009.

[20] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data," *SIAM Rev.*, vol. 51, pp. 661–703, Nov. 2009.

[21] M. E. Newman, "Power laws, pareto distributions and zipf's law," *Contemporary physics*, vol. 46, no. 5, pp. 323–351, 2005.

[22] D. Champernowne, "A model for income distribution," *Economic Journal*, vol. 63, pp. 318–351, 1953.

[23] T. Hossmann, T. Spyropoulos, and F. Legendre, "Putting contacts into context: Mobility modeling beyond inter-contact times," in *Proceedings of the Twelfth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, MobiHoc '11, (New York, NY, USA), pp. 18:1–18:11, ACM, 2011.

[24] C. Boldrini and A. Passarella, "Hcmm: Modelling spatial and temporal properties of human mobility driven by users' social relationships," *Computer Communications*, vol. 33, pp. 1056 – 1074, June 2010.

[25] N. Vastardis and K. Yang, "An enhanced community-based mobility model for distributed mobile social networks," *Journal of Ambient Intelligence and Humanized Computing*, vol. 5, pp. 65–75, Apr. 2014.

[26] A. Munjal, T. Camp, and W. C. Navidi, "Smooth: A simple way to model human mobility," in *Proceedings of the 14th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, MSWiM '11, (New York, NY, USA), pp. 351–360, ACM, 2011.

[27] L. M. Silveira, J. M. de Almeida, H. T. Marques-Neto, C. Sarraute, and A. Ziviani, "Mobhet: Predicting human mobility using heterogeneous data sources," *Computer Communications*, vol. 95, pp. 54 – 68, 2016. Mobile Traffic Analytics.

[28] The Scenario Generator, "http://isis.poly.edu/ qiming/scengen/index.html."

[29] Scalable Network Technologies, "Qualnet 4.0."

[30] J. Yoon, M. Liu, and B. Noble, "Random waypoint considered harmful," in *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, vol. 2, pp. 1312–1321 vol.2, March 2003.

[31] C. Perkins, E. Belding-Royer, and S. Das, "Ad hoc on-demand distance vector (aodv) routing," 2003.

[32] X. Hou, Y. Li, D. Jin, D. O. Wu, and S. Chen, "Modeling the impact of mobility on the connectivity of vehicular networks in large-scale urban environments," *IEEE Transactions on Vehicular Technology*, vol. 65, pp. 2753–2758, April 2016.

[33] B. O. Sadiq, A. E. Adedokun, and Z. M. Abubakar, "The impact of mobility model in the optimal placement of sensor nodes in wireless body sensor network," *CoRR*, vol. abs/1801.01435, 2018.

Danielle Ferreira received her B.S. in engineering from State University of Rio de Janeiro (UERJ) and her M.S. degrees in computer science at the Federal University of Rio de Janeiro (UFRJ), Brazil. She is currently a Ph.D. candidate in the computer science graduate program at Federal University of the State of Rio de Janeiro (UNIRIO), Brazil. Since August 2018 she has joined the Internetwork-Research Group (i-NRG) at University of California in Santa Cruz's Computer Engineering Department as a visiting Ph.D. researcher. Her research interests include wireless network, mobile computing, opportunistic network, software defined networking, Internet of thing and intelligent transportation system.

Bruno Astuto A. Nunes is a researcher at UCSF, CA, USA. Before joining UCSF he worked as a researcher at GE Global Research in Rio de Janeiro, Brazil for 4 years. He worked as a Pos-Doc fellow researcher at INRIA Sophia Antipolis, France. He received his B.Sc. in Electronic Engineering at the Federal University of Rio de Janeiro (UFRJ), Brazil, where he also completed his M.Sc. degree in Computer Engineering. He received his PhD degree in Computer Engineering from UC Santa Cruz, USA.

Katia Obraczka is Professor of Computer Engineering at UC Santa Cruz. Before joining UCSC, she was a research Scientis at USC's Information Sciences Institute (ISI) and had a joint research faculty appointment at USC's Computer Science Department. Prof. Obraczka's research interests span the areas of computer networks, distributed systems, and Internet information systems. Her lab, the Internetwork Research Group (i-NRG) at UCSC, conducts research on designing and developing protocol architectures motivated by the internets of the future. She has been a PI and a co-PI in a number of projects sponsored by government agencies (e.g., NSF, DARPA, NASA, ARO, DoE, AFOSR) as well as industry (e.g., Cisco, Google, Nokia). She is currently serving as Associate Editor for the IEEE Transactions on Mobile Computing as well as IEEE Letters of the Computer Society. She is a Fellow of the IEEE.