On the Heavy Tail Properties of Spatial Node Density for Realistic Mobility Modeling

Bruno A. A. Nunes[1] and Danielle Lopes Ferreira[2] and Katia Obraczka[2]
[1]INRIA, Sophia-Antipolis, France
[2]Department of Computer Engineering, Baskin School of Engineering
University of California, Santa Cruz, CA, USA
Email: llopes.danielle@gmail.com, bastuto@gmail.com, katia@soe.ucsc.edu

***Abstract*—In this paper, we show empirically that the spatial node density resulting from human mobility follows a power law. We also show that the number of locales visited by users also exhibit heavy-tail behavior. We develop a stochastic model that confirms our empirical observations by showing that node mobility resulting from our model closely approximates mobility recorded in real traces collected from a variety of scenarios. Besides corroborating our empirical observations, we showcase another application of our model by using it to generate mobility regimes whose spatial node density exhibit heavy-tail behavior. We validate the resulting mobility generator by comparing its output against real traces.**

I. Introduction

Understanding how users move is critical to the design of wireless networks and their protocols. So much so that, over the last decade, network researchers have dedicated considerable attention to user mobility modeling and characterization. The importance of node mobility in designing networks has motivated researchers and practitioners to try to use realistic scenarios to drive the design and evaluation of wireless network protocols. As a result, initiatives such as the CRAWDAD [1] trace repository have emerged. CRAWDAD makes available real mobility traces which can then be used by the networking community to test and evaluate mobile networks and their protocols. However, even though initiatives like CRAWDAD have greatly increased availability of real traces, the number and variety of publicly-available mobility traces is still quite limited. Consequently, relying exclusively on traces to design and evaluate network protocols would not allow a broad enough exploration of the design space.

The relatively limited availability of real traces has lead to the development of "mobility generators", which, following a pre-specified mobility regime, determine the position of network nodes over time during simulation runs. Synthetic mobility generators have been extensively used in the study of wireless networks [2]. One of the most widely used synthetic mobility regimes is the Random-Waypoint Mobility (RWP) model. However, RWP mobility has been known to exhibit some undesirable properties [3]. The work reported in [4] shows that random mobility leads to homogeneous spatial node distributions. This means that even if the initial spatial node distribution is spatially heterogeneous, node density will tend to "deteriorate" to a homogeneous distribution under RWP mobility.

978-1-4799-4657-0/14/$31.00

Indeed, one of the main challenges in constructing mobility generators is building models that can capture key features and the complexity of user mobility in real-live settings. As suggested before, one such key feature is node clustering, which can be defined as the tendency of nodes to agglomerate. In prior work [5], we show that, instead of being homogeneously distributed over an area, users tend to congregate and form clusters, where some regions may be quite dense while others completely deserted. In order to express this feature of user mobility, we define *spatial node density* as the number of nodes located in a given unit area. Our study of user mobility also revealed that the original non-homogeneous spatial node density distribution is maintained over time. Thus, according to our observations, human mobility is characterized by non-uniform, time-invariant spatial node density.

We contend that 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. To date, only a few efforts have focused on modeling spatial node density. Notable examples include [6] and [7], which propose analytical models to study spatial node density under RWP mobility. In our own prior work [8], we model spatial node density by developing a set of first order ordinary differential equations (ODEs) whose parameters are extracted from real mobility traces.

In this paper, we start by showing empirically (using real mobility traces collected in a variety of scenarios) that spatial node density observed in human mobility can be modeled by a *Power Law* (PL). We then propose a model to describe analytically the heavy tail behavior exhibited by spatial node density and verify that the proposed model closely approximates empirical spatial density distributions resulting from real mobility traces. As an example application, we use our model to build a waypoint-based mobility regime that is capable of generating mobility traces whose spatial node density distributions closely resemble the ones measured in real human mobility scenarios with the advantage of not requiring to extract model parameters from empirical datasets. We use our mobility regime to simulate node mobility in ad hoc network scenarios and show that the resulting average spatial node density closely resembles spatial density behavior observed in real mobility traces (the average absolute error in this case was less than 15%).

The remainder of this paper is organized as follows: The next section presents important definitions and the mobility datasets used in our work. In Section III we present our empirical study on the PL properties of real user spatial density. Section IV presents our analytical model for node spatial density and show a good matching between it and the same metric extracted from real mobility datasets. In Section V we present a waypoint-based mobility regime that is capable of generating simulated mobility traces in which the engendered spatial node density resembles closely the ones measured in real human mobility scenarios. Related work on spatial node density and realistic mobility models are presented in Section VI. Finally, Section VII concludes the paper.

II. Data and Definitions

For our study, we use real traces collected in scenarios that are quite diverse, namely a park in the city of Rio de Janeiro, Brazil, a University campus, and taxis in San Francisco, California, USA. Some of the traces were collected using GPS devices, while others record associations and disassociations of users to Wi-Fi Access Points (APs). These traces are summarized in terms of number of users, duration of the trace, and the log sampling period in Table I.

For the GPS traces, we divide the mobility domain, i.e., the area around which nodes move, into equal sized squares, that we refer to as *cells*. To decide what cell size to use, we assume that a node's transmission range is 100 meters. If we use squared cells, they will be circumscribed in a 200m diameter circle. That results in 140 x 140 meter cells. To validate our choice of cell size, we ran experiments with different cell sizes and observed no significant change in the results obtained for reasonable cell sizes (i.e., not considering cells that are too small or too big). In the case of WLAN traces, their cells are defined by the area covered by the APs, where each cell corresponds to an AP.

Spatial Node Density

We define spatial node density as the number of nodes located in a given unit area. We extract spatial node densities from the traces using the trace's sampling period. We also use the following definitions:

- *Node Spatial density distribution:* the percentage of cells being visited by $\geq x$ nodes.
- *Node Mobility Degree:* the percentage of nodes that visit a number $\geq n$ of cells.

Intensity Map

We measure the *intensity* of a cell as the number of node visits a given cell receives during a given time interval. Each cell $i \in \{1..N\}$, where N is the total number of cells composing the physical mobility domain, is assigned an intensity μ_{i,T_t} at interval T_t given by an *Intensity Map (IM)*, for $t \in \{0, 1, 2, ...\}$. The IM is a vector composed by N elements, where each element has a value $\{\mu_{i,T_t} \in \mathbb{R} \mid \mu_{i,T_t} \geq 0\}$ that indicates how intense the activity in cell i is. The IM for interval T_t, will give us the spatial node density for that interval of time.

Analogously, we extract an *User Intensity Map (UIM)* composed by L elements, where each element has a value $\{\mu_{l,T_t} \in \mathbb{R} \mid \mu_{l,T_t} \geq 0\}$ that indicates how mobile a node $l \in \{1..L\}$ is. In other words, μ_{l,T_t} is the number of cells visited by node l during T_t.

Node Speed and Pause Time

We also extract from the traces the distributions of speed and pause time by using the trace's sampling period. For example, in the Quinta trace, the sampling period is $T = 1$ seconds. We define the node's 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 defined as $P = \Delta t$, if $d <$ *threshold*, or zero otherwise. The threshold is not zero here to account for GPS error. We set this threshold to be 0.5 meter for the Quinta trace due to jitter in the GPS update frequency.

Trace	# users	# Cells	Duration	Samples
Quinta [9]	97	16	900s	1s
SF Taxis [10]	483	1600	24 days	1 to 3 mins
Dartmouth [11]	6524	1776	60 days	Instant

TABLE I
Summary of the real traces studied.

III. Power Law and Node Density

In this section, we use well-known statistical techniques to show that spatial node density resulting from human mobility exhibit heavy tail behavior, i.e., it follows a Power Law distribution. Power law has been used to describe a number of phenomena in communication networks, such as, node inter-contact time [12], Internet topology [13], Internet traffic [14], traveled distance and trip displacement [15] etc. To the best of our knowledge, this is the first time that spatial node density resulting from real mobility has been shown to follow a power law.

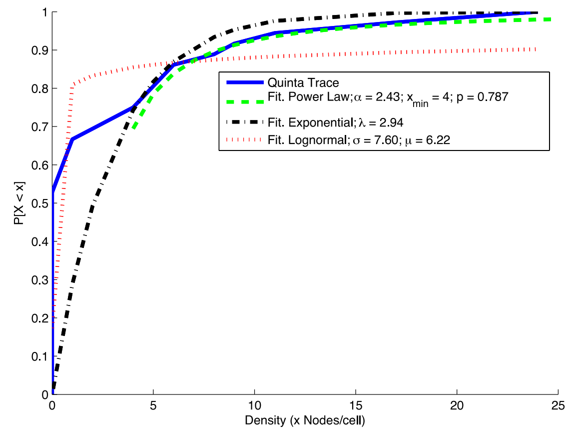
Fitting empirical data into a distribution that follows a power law is not trivial due to fluctuations that occur in the tail of the distribution representing rare events, and also due to the difficulty in identifying the part of the data distribution that actually follows the power law.

Power laws are expressions of the form $P(x) \propto x^{-a}$, where a is a constant parameter and x are the measurements of interest. Few physical phenomena follow a power law for all values of x [16]. 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.

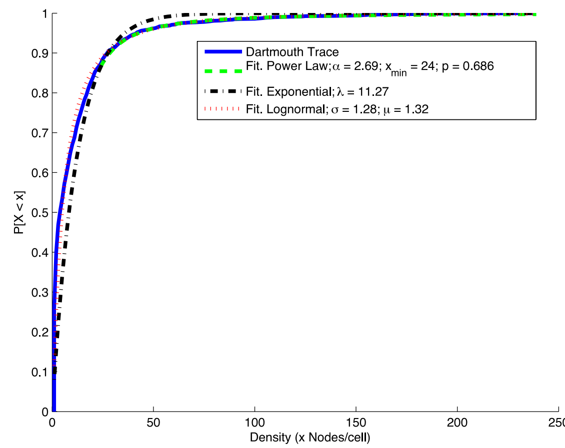
We employed the statistical techniques described in [16] to fit the empirical data into a power law and compute its parameters. Also in [16], *a goodness-of-fit* test was proposed which generates a p parameter to test whether a distribution follows a power law. In other words, it tests if a curve that follows a power law is a plausible fit for the empirical data. This test is based on the measurement of the distance between the distribution of the empirical data and the hypothesis of the model. This distance 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. Thus, if p is high (close to 1), than the differences between the empirical data and the model can be attributed to statistical fluctuations. The distance is computed through the statistical test of Kolmogorov-Smirnov (KS). In the case where p is small the model is not plausible. In [16], a reference value of $p < 0.1$ is used to reject the hypothesis that the empirical data follows a power law.

Figure 1 shows the cumulative distribution function (CDF) of the spatial node density distribution for the Quinta 1(a), Dartmouth 1(b), and San Francisco cabs 1(c) mobility traces. The plots show the fitting of this data according to three probability distributions: exponential, log-normal and a power law distribution (using the method described in [16] for the power law fitting). The first two distributions are shown here, due to the fact that even if the data is well fit by a power law it is still possible that another distribution, such as an exponential or a log-normal, might give a fit as good or better. Thus, for the sake of comparison, we also show the fitting of the same data using the exponential and log-normal distributions. It is possible to see from Figure 1 that the power law distribution is the best fit for all three scenarios studied.

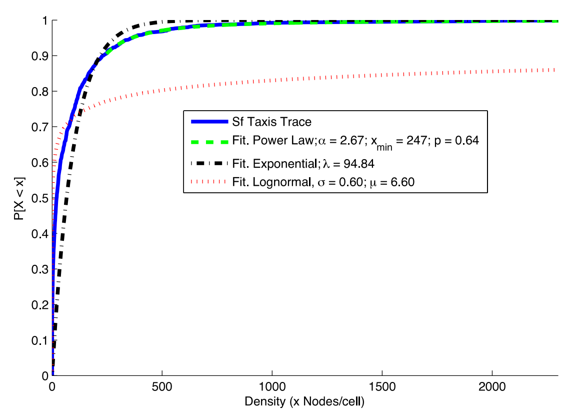
These figures also show the parameters of the fitted curves for each probability distribution. The power law distribution curves show the respective values of p for the three traces studied. Through the metric p, computed for the power law distribution, it is possible to verify if the hypothesis that the empirical data follows a power law is a plausible one for these scenarios. It is important to notice that for all the traces, p is well above the reference threshold of 0.1 defined in [16], validating the hypothesis as plausible that the spatial density distribution of mobile nodes follows a power law with parameters close to $\alpha = 2.5$ and x_{min} proportional to the number of observations. Here, α gives the slope of the curve and x_{min} depicts where the heavy tail behaviour begins, which for these traces ranges from 10% to 20% of the upper density.



(a) Spatial node density distribution from Quinta (city park)



(b) Spatial node density distribution from Dartmouth College (university campus)



(c) Spatial node density distribution from San Francisco taxis (vehicular network)

Fig. 1. CDF of the spatial node density distribution for the Quinta, Dartmouth and San Francisco cab traces.

Another important factor in mobile networks is how mobile a node is. In other words, the *number of different locations or cells that a mobile node visits*, defined here as *node mobility degree*. We have been able to verify that the hypothesis that the node mobility degree follows a power law distribution is also plausible. However, due to space constrains we do not show the statistical analysis here for this metric. From the point of view of DTNs or social networks, this information is quite useful since a mobile node that visits a large number of locations would potentially experience a greater number of opportunistic contacts.

IV. Spatial Node Density Stochastic Model

Having verified the power-law characteristics of spatial node density resulting from real user mobility motivated us to develop an analytical model for spatial node density and node mobility degree that would reflect their power law properties. A similar model was developed for the distribution of income of people living in the UK in the early 50's ([17]). In this section, we present a stochastic model for the spatial node density distribution and node mobility degree that reflects their power law behaviour.

A. Node Density Distribution as a Stochastic Process

In the model presented in this section the spatial node density distribution of some enumerable groups of cells is assumed to develop by means of a stochastic process. The stochastic matrix is assumed to remain constant through time. Under these circumstances and provided certain other conditions discussed below are satisfied, the distribution will tend towards a unique equilibrium distribution dependent upon the stochastic matrix but not on the initial distribution.

We suppose that the measured values of cell's density, or the measured number of mobile nodes visiting a cell, is divided into an enumerable infinity of ranges, which we assume to have uniform proportionate extend. For example, we might consider the ranges per interval time to be 1-2 nodes, 2-4 nodes, 4-8 nodes, 8-16 nodes, ... although different graduations could be considered. We shall regard the development through time of the density distribution between these ranges as being a stochastic process, so that the density of any individual cell in an interval time may depend on what it was in the previous interval time and on a random process. We are assuming, also, that the total number of cells in the system is invariant through time.

Under such assumptions, the node spatial density distribution can be described in terms of the following vectors and matrices. We first define $X_r(0)$, as the number $X_r(0)$ of cells in each range R_r, $r = 1, 2, ...$ in the initial time T_0 and a series of matrices $p'_{rs}(t)$ telling us in each time T_t the percentage of cells of R_r who are shifted to range R_s in the following interval time T_{t+1}. With these definitions the density distribution $x_r(t)$ in the successive time intervals will be generate according to Equation (1).

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

If we now suppose that the ranges are ordered by size (there being a lowest cell density range R_0), then we can define a new set of stochastic matrices

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

and rewriting (1) in the form

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

$p_{ru}(t)$ then tell us the proportion of cells in R_r who shift by a number u of ranges in T_t.

In order to make a simple model, we should assume that the matrix $p_{ru}(t)$ regarded as a frequency distribution in u differs very little in form for variations over a wide range of values of r and t. When considering the practical implications of this suggestion one can see that it means that the probability of shifts upwards and downwards along the ranges of cell's densities differ little as between the occupants of different ranges, and differ little over time.

Our other assumption that the function $p'_{r,r+u}(t) = p_{ru}(t)$ remains constant over time was shown in our previous work [5]. Examining the static equilibrium generated by a fixed set of functions $p'_{rs}(t)$ is an essential preliminary to the study of the dynamic equilibrium with moving $p'_{rs}(t)$.

It is known that under very general conditions the repeated application of the same set of cell-changes represented by an irreducible matrix $p'_{rs}(t)$ will make any initial density distribution eventually approach a unique equilibrium distribution, determined by the matrix $p'_{rs}(t)$ alone.

It would be a great advantage in constructing models of density distribution if we had empirical evidence about the matrix $p'_{rs}(t)$ describing actual movements of density distribution in real networks. Thus, in order to validate the proposed model, some such evidence was compiled from the CRAWDAD repository [1] through two of the studied traces, Dartmouth and SF Taxis, and the third trace, the Quinta trace [9].

B. A Model Generating a Pareto Distribution

Let us assume, for the sake of simplicity, that for every value of t and r, and for some fixed integer n, we have

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

This means that no cell move up by more than one range in a time, or down by more than n ranges at a time.

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

Equation (5) tells us that the probabilities of shifts upwards and downwards along the ranges of cell's densities are distributed in a manner independent of present cell, despite of the limitations imposed by the impossibility of descending below a given number of positions. This postulate always leads to a density distribution which obeys Pareto's law at least asymptotically for high density of cells.

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

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

which by (5) also implies

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

This assumption (6) expresses that all density of cells preserve their identity throughout time in the manner described in Section IV-A above.

One other assumption must be introduced in order to ensure that the process is not dissipative, i. e., that the density of cells do not go on increasing indefinitely without settling down to an equilibrium distribution. Let us denote

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

then our stability assumption is that

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

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 may determine the equilibrium distribution corresponding to any matrix $p^{'}_{r,r+u}(t) = p_{r,u}(t)$ conforming to our assumed rules. 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. This distribution must be, apart from a multiplying constant, the unique distribution which will be obtained by 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 Equaliton (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 the solution (12) implies a total number of cells given 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

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

Now we shall assume that the proportionate extent of each range is 10^h, and that the lowest density of cell 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

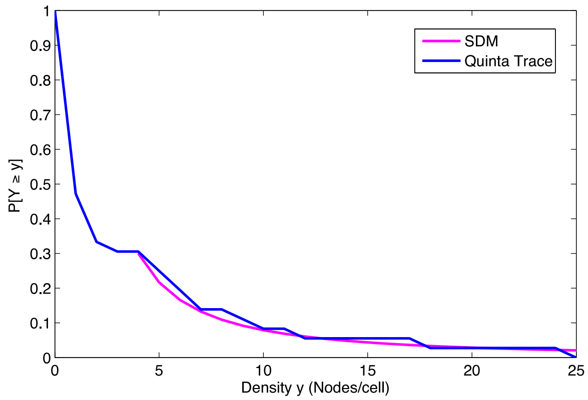
$$log_{10} F(y_s) = \gamma - \alpha log_{10} y_s \tag{20}$$

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.

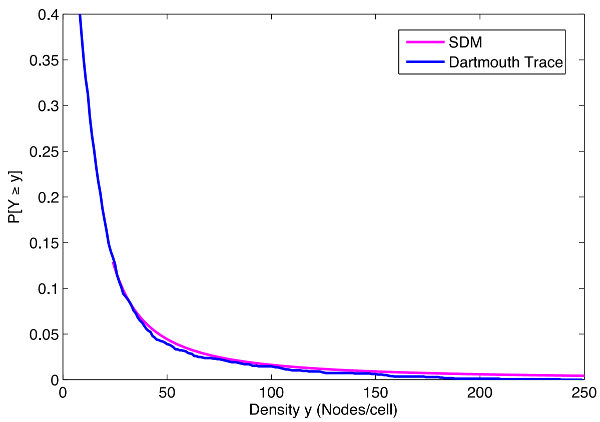
Thus if all ranges are equal proportionate extent, our simplifying assumptions ensure that any initial distribution of density cells will in the course of time approach the exact Pareto distribution given by (19) and (20).

The very simple model discussed in this section brings out clearly the tendency for Pareto's law to be obeyed in a "community of cells" where, above a certain minimum value for cell's spatial density, the chances of various amounts of percentage change for the cell's spatial node density are independent of the initial density distribution.

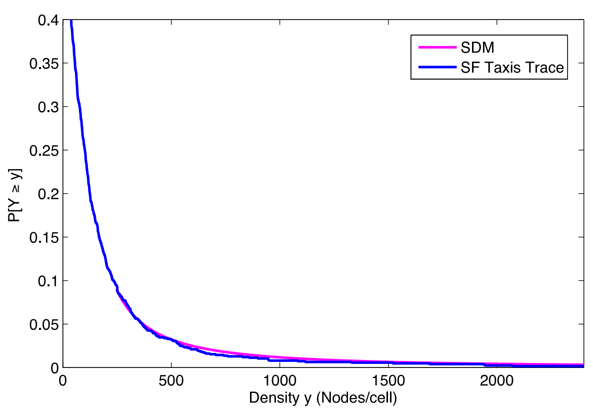
The spatial node density distribution can be observed in Figure 2 for the Quinta, Dartmouth and SF Taxis mobility traces respectively. These curves show the probability of finding a cell that was visited by y or more nodes at a given interval, and were computed by extracting the number of users visiting each cell at a given interval, i.e. $[800s, 900s]$



(a) City park - GPS.



(b) University campus - Wi-Fi



(c) Vehicular network in a city - GPS.

Fig. 2. Spatial node density distribution for the (a)Quinta, (b)Dartmouth and (c)SF Taxis mobility traces, real mobility trace distribution and the proposed analytical model.

for the Quinta trace, and a random non-interrupted time interval of 24hrs for Dartmouth and SF Taxis.

The aforementioned figures also show the curve for the model proposed in this section, that we call here the Stochastic Density Model (SDM). The coefficients of the stochastic matrix, used to parametrize our model were extracted from the traces so that we could compare to the empirical density and validate our model.

C. User Mobility as a Stochastic Process

Here, we define *user mobility* as the *number of cells visited by a mobile user over a given period of time.* A user with low mobility visits a small number of cells, while a very mobile user visits a larger number of cells. Following the assumption that user mobility follows a power law, as indicated by our findings in Section III, we then present in this section the user mobility distribution between an enumerable infinity groups of users is also assumed to develop by means of a stochastic process, in the same way as in Section IV-A.

Following similar assumptions, the user mobility distribution can be described in terms of the following vectors and matrices. We define $\Theta_d(0)$, as the number $\Theta_d(0)$ of users in each range D_d, $d = 1, 2, ...$ in the initial time T_0 and a series of matrices $p'_{dv}(t)$ telling us in each time T_t the percentage of users of D_d who are shifted to range D_v in the following interval time T_{t+1}. With these definitions the user mobility distribution $\theta_d(t)$ in the successive time intervals will be generate according to

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

As before we suppose that the ranges are ordered by size (there being a lowest range of number of cells visited per user C_0), then we can define a set of stochastic matrices

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

and rewriting (21) in the form

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

$p_{df}(t)$ then tell us the proportion of users in D_d who shift by various number f of ranges in T_t.

After some evaluations that follow similar assumptions as in Section IV-B, we are able to find the equilibrium distribution $F(\omega_v)$ of the number of users which the number of visited cells exceeds ω_v.

Taking into account the model derived in this section for the long tail distributions of number of visited cells and spatial node distribution, in the following section we present a new mobility regime that dictates how users move around the simulation area.

V. Stochastic Density Mobility Model

In Section III, we show, based on empirical evidence, that spatial node density resulting from human mobility follows a power law. Besides validating our observations analytically, another application of the Stochastic Density

Model (SDM) we developed in Section IV above is to generate mobility regimes that result in spatial node density that exhibits heavy-tail behaviour.

We call the resulting mobility regime Stochastic Density Mobility Model (SDMM) which works as follows. We first define "cluster" regions (high density cells) using SDM. These regions are cells where the concentration of nodes visiting them is greater than a given threshold y_{min}; or, in other words, we use the tail of the density distribution to derive the probability of a node choosing a cell, thereby generating cluster regions. In the case of non-cluster regions (low density cells), where the concentration of node visits is below the y_{min} threshold, for the sake of simplicity, we apply a uniform spatial density distribution. The results shown here for one of the mobility traces indicate that the uniform distribution is a reasonable approximation for the low density distribution. However, it is our ongoing work to study more closely the impact of different known distributions to model cell density bellow y_{min}.

Below we provide a more detailed description of our proposed model and describe how its parameters are set. We highlight that one of the benefits of our model is to generate node spatial distributions that are similar to the ones found in real live mobile applications (e.g., the traces presented in Section II) without the need to extract parameters from mobility traces.

A. Proposed Mobility Model

During the initialization phase, nodes can be distributed in the geographic area according to any given distribution. For simplicity, we then use a waypoint-based mobility regime, contending that a mobility model should be simple in order to be widely adopted. Thus, the steps involved in the the movement are as follows. In Step (1), using a probability distribution given by $F(\omega_v)$, a node decides if it is going to move to another cell or not. In Step (2) the node chooses its next cell using a probability distribution given by $F(y_s)$. A (x, y) destination is chosen randomly inside the chosen cell. In Step (3) the node moves to that destination at a randomly chosen speed, uniformly distributed between [Vmin, Vmax]. In Step (4) the node pauses for some time, and repeats from Step (1).

More specifically, in Step (1), a node l decides with probability $P(\mu_l)$ if it will remain in the same cell, or if it will choose a new destination in a different cell, with probability $1 - P(\mu_l)$. The number of different cells μ_l visited by node l is defined *a priori* by sampling from the computed distribution $F(\omega_v)$. The probability $P(\mu_l)$ that a user l will remain in the same cell is computed as in Equation 24, and this value of $P(\mu_l)$ is kept constant for each node l during the simulation.

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

In Step 2 a node decides with probability $P(\mu_i)$ its next destination cell. We assume that the probability $P(\mu_i)$ that a node would choose cell i as a next destination depends on the 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 $P(\mu_i)$ is kept constant for each cell i during the simulation.

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

As we previously pointed out, it is important to note that the parameters for the proposed mobility regime do not need to be extracted from real mobility traces. 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). In the simulation results presented in the next section, we extracted the parameters from the traces for a fair comparison between the proposed model and the real trace. From the statistical study presented in Section III, y_{min} was found to normally 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.

B. Simulation and Results

We validate our mobility regime by utilizing a modified version of the Scengen [18] scenario simulator to generate traces according to our proposed stochastic density mobility regime. It is worth noting that the analytical model presented in Section IV is generic enough that could be applied to compute the probability of choosing a node's next destination not only in simulations in ad-hoc scenarios, but also in other mobility scenarios (e.g., WLAN scenarios). Moreover, the parameters of our model can be adjusted to represent different and arbitrary scenarios, without the need to rely on real mobility traces. The results in terms of spatial density distribution and user mobility, enabled by our model, should fit the spatial characteristics found in real human mobility or, at least, present the same power law behaviour. Nonetheless, the proposed model also allows, as shown in this section, that it is possible to apply parameters extracted from real mobility traces, if the goal is to represent some specific existing real human mobility scenario, from where a mobility trace dataset was made available.

In these experiments, a set of 10 synthetic traces were generated using our implementation of SDMM in Scengen. The results shown in Figure 3 refer to the Quinta trace. In order to compare our model with an existing real human mobility scenario we adjusted the parameters of the simulation according to data extracted from a real mobility trace. The speed range was set such that the average speed would match the ones measured in the GPS traces for Quinta. Pause times were chosen uniformly in the range $[0, P_{max}]$, where the value of P_{max} was set so that the average pause time would match the one measured in the real traces. The same was done for the dimensions of the rectangular simulation area, set to be the same as in the GPS traces. Also, we used the same initial positions found in the real trace for the same number of users. Table II summarizes the simulation parameters.

Parameter	**Quinta GPS Trace**
Avg. Speed ($\pm\sigma$)(m/s)	1.2($\pm$0.53)
Avg. Pause (sec)	3.6
Total Area (meters x meters)	840x840
Cell Size (meters x meters)	140x140
Duration (sec)	900
# nodes	97

TABLE II
SIMULATION PARAMETERS.

Figure 3 shows the mean spatial node density distribution for SDMM averaged over 10 simulation runs. The error bars in this curve show the confidence interval for a confidence level of 95%. The other two curves represent the proposed analytical model presented in Section IV and the empirical spatial node density distribution at end of the trace collection interval, i.e., 900 seconds.

From these plots we observe that the density distribution for SDMM regime follows closely the distribution of the Quinta trace. We also observe that the difference between those two curves and the curve for the SDM analytical model is quite small. Quantitatively, we define these errors as a percentage difference between the distribution for the real trace and the curve for the proposed mobility regime taken at each point in the x-axis. In other words, the error is the absolute difference between the spatial distribution for the proposed model and the distribution computed for the real trace, taken at each point in the x-axis, divided by the corresponding value of the density distribution given by the real trace. The mean error in this case, averaging the errors computed for all points in the x-axis is 14.82%, where the confidence interval is equal to $[11.72\%, 17.91\%]$ for a 95% confidence level.

VI. Related Work

A key aspect of human mobility is user's, or node's spatial density. There has been some effort in trying to model user spatial distribution for mobile networks, however, these efforts focus on modeling density of known to be unrealistic random mobility models. To name a few, authors in [6] and [7], propose analytical models for spatial node density in simulations driven by the RWP mobility regime. On the other hand, in a more recent effort [8] a set of first order ODEs was used to model spatial density where the model parameters are extracted from real mobility traces. In the present work we are able to closely model spatial density without the need for dataset input. As an application use case, we apply our analytical model in defining the probabilities of choosing a node's next destination in a new mobility model that is also able to generate realistic node spatial density distributions.

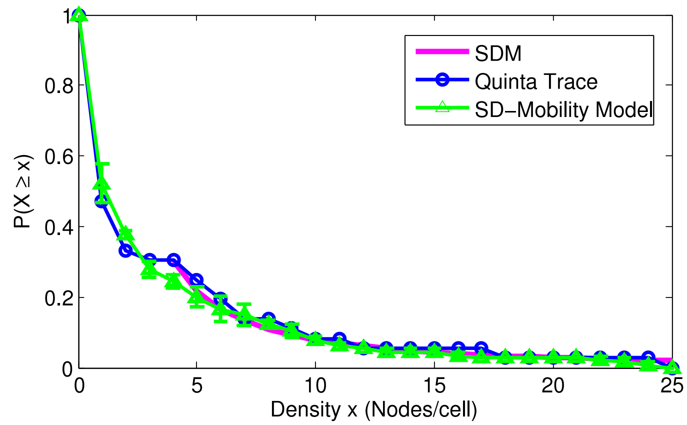


Fig. 3. Node Density Distribution.

Mobility models are indeed an indispensable tool in the design, testing, and evaluation of wireless networks and their protocols. To address the challenge of realistic mobility simulation, a number of efforts have proposed mobility models based on realistic mobility patterns [19]. Notable examples include [20, 21, 22, 23]. However, this approach may limit their practical application in scenarios where there isn't any mobility data available.

More recent work focuses on the "scale-free" properties observed in many real networks like the Internet, the Web, and some social networks, to name a few. The seminal work of Barabási and Albert [24] proposes a model that generates scale-free networks, i.e., networks whose node degrees follow a power law distribution. One key concept underpinning the Barabási-Albert model is referred to as the *preferential attachment* principle which states that "the more connected a node is, the more likely it is to receive new links". Several recently proposed mobility models (e.g., [25, 26, 27, 28, 29]), try to mimic real human mobility by following the preferential attachment principle and aim on generating spatial node density that resembles the one found in real mobility scenarios.

However, it was shown in previous work that using the preferential attachment principle to model human mobility leads to undesirable long-term behavior [5]. More specifically, preferential attachment based mobility regimes do not preserve the original spatial node density distribution and lead to steady-state behavior similar to random mobility as exemplified by the RWP model. Instead, real human mobility exhibits persistent density heterogeneity as shown in [5]. In our work we take a different approach, where we propose a model that may receive input parameters extracted from real mobility records but is also able to faithfully represent spatial node density that resembles the one found in real-live mobility without the need for extracting parameters from mobility datasets.

VII. Conclusions

We investigated in this paper human mobility and the heavy tail characteristics of some important mobility metrics, such as spatial node density. In our study we analyze a set of three very diverse real mobility traces. Those mobility traces include human mobility in a park, a university campus WiFi network and a vehicular mobility trace on a city center. We applied sound statistical methods and demonstrated that a power law distribution is a plausible model for real human mobility and spatial node density in the diverse set of scenarios considered. Moreover, we took advantage of the power law properties of such important mobility metrics and proposed an analytical model, regarding the spatial node density distribution and the user mobility in a mobile scenario as a stochastic process. We showed that our model approaches closely the empirical data measured from real mobility traces. As our final contribution we applied the equations derived from the proposed analytical model to build a waypoint-based mobility regime that is capable of generating simulated mobility traces in which the engendered spatial node density resembles closely the ones measured in real human mobility scenarios.

In this work, the evaluation of our SD-Mobility Model was performed in order to show-case a practical application of our analytical model. Further evaluating the proposed mobility regime and comparing it with spatial and temporal characteristics found in real traces, extracted from other real mobility scenarios, is part of our on-going work. Moreover, we plan on studying the impact of our mobility model and the spatial and temporal characteristics mentioned, on core network functions, such as routing, for example. This evaluation would involve simulating mobile ad-hoc networks that evolve according to our mobility model and comparing it against other mobility regimes designed to account for the scale-free characteristics found in real human mobility scenarios.

References

[1] CRAWDAD, "http://crawdad.cs.dartmouth.edu/."

[2] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," *Wireless Communications and Mobile Computing*, vol. 2, pp. 483–502, 2002.

[3] 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.

[4] C. Bettstetter, M. Gyarmati, and U. Schilcher, "An inhomogeneous spatial node distribution and its stochastic properties," MSWiM '07, pp. 400–404, 2007.

[5] B. A. A. Nunes and K. Obraczka, "On the invariance of spatial node density for realistic mobility modeling," MASS, pp. 322–331, 2011.

[6] W. Navidi and T. Camp, "Stationary distributions for the random waypoint mobility model," *IEEE Transactions on Mobile Computing*, vol. 3, pp. 99–108, 2003.

[7] 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.

[8] B. A. A. Nunes and K. Obraczka, "Modeling spatial node density in waypoint mobility," in *Proceedings of the 2012 IEEE Ninth International Conference on Mobile Ad-Hoc and Sensor Systems*, MASS, 2012.

[9] 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.

[10] 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.

[11] 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.

[12] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on opportunistic forwarding algorithms," *Mobile Computing, IEEE Transactions on*, vol. 6, no. 6, pp. 606–620, 2007.

[13] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the internet topology," *SIGCOMM Comput. Commun. Rev.*, vol. 29, pp. 251–262, Aug. 1999.

[14] M. Crovella and A. Bestavros, "Self-similarity in world wide web traffic: evidence and possible causes," *Networking, IEEE/ACM Transactions on*, vol. 5, no. 6, pp. 835–846, 1997.

[15] M. Zhao, L. Mason, and W. Wang, "Empirical study on human mobility for mobile wireless networks," in *Military Communications Conference, 2008. MILCOM 2008. IEEE*, pp. 1–7, Nov 2008.

[16] 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.

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

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

[19] M. Musolesi and C. Mascolo, "Mobility models for systems evaluation," in *Middleware for Network Eccentric and Mobile Applications*, pp. 43–62, Springer, 2009.

[20] A. Balachandran, G. M. Voelker, P. Bahl, and P. V. Rangan, "Characterizing user behavior and network performance in a public wireless lan," in *Proceedings of the 2002 ACM SIGMETRICS*, pp. 195–205, 2002.

[21] M. Balazinska and P. Castro, "Characterizing mobility and network usage in a corporate wireless local-area network," in *Proceedings of MobiSys 2003*, pp. 303–316, 2003.

[22] D. Tang and M. Baker, "Analysis of a local-area wireless network," in *Proceedings of the MobiCom 2000*, pp. 1–10, 2000.

[23] T. Henderson, D. Kotz, and I. Abyzov, "The changing usage of a mature campus-wide wireless network," in *In Proceedings of ACM MOBICOM 2004*, pp. 187–201, 2004.

[24] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, vol. 74, no. 1, pp. 47–97, 2002.

[25] A. Jardosh, E. Belding-Royer, K. Almeroth, and S. Suri, "Real-world environment models for mobile network evaluation," *Selected Areas in Communications, IEEE Journal on*, vol. 23, no. 3, pp. 622 – 632, 2005.

[26] M. Kim, D. Kotz, and S. Kim, "Extracting a mobility model from real user traces," in *INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings*, pp. 1–13, April 2006.

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

[28] A. Mei and J. Stefa, "Swim: A simple model to generate small mobile worlds," in *INFOCOM 2009, IEEE*, pp. 2106–2113, April 2009.

[29] M. Musolesi and C. Mascolo, "Designing mobility models based on social network theory," *ACM SIGMOBILE Mobile Computing and Communication Review*, vol. 11, pp. 59–70, 2007.