UNIVERSITY of CALIFORNIA
SANTA CRUZ

CHARACTERIZING USER MOBILITY IN WIRELESS NETWORKS

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

COMPUTER ENGINEERING

by

Bruno Astuto Arouche Nunes

September 2012

The dissertation of Bruno Astuto Arouche Nunes is
approved:

_____
Katia Obraczka, Chair

_____
Abel Rodrigues

_____
Brad Smith

_____
Tyrus Miller
Vice Provost and Dean of Graduate Studies

Contents

List of Figures

List of Tables

Abstract

Characterizing User Mobility in Wireless Networks

by

Bruno Astuto Arouche Nunes

In this work we study mobile wireless networks by looking at mobility management and analysis of human mobility, focusing on the main goal of understanding human mobility and applying our findings on developing new realistic mobility models for simulations. In our work, we start by analyzing Wireless Local Area Networks (WLAN) and GPS traces that record mobility in a variety of network environments. We observe that from a macroscopic level, human mobility is *symmetric*. We also study the direction of movement which also exhibits symmetric behavior in both real– as well as synthetic mobility. Moreover, motivated by the symmetric behavior identified, we continued our investigation on real mobility characteristics, by focusing on node spatial density in real applications. We show that human mobility exhibits "persistent" behavior in terms of the spatial density distribution of the mobile nodes over time. By using real mobility traces, we observe that the original non-homogeneous node spatial density distribution, where some regions may be quite dense while others may be completely deserted, is maintained at different instants of time.

We also show that mobility models that select the next node position based on the position of other nodes, a la "preferential attachment", do not preserve the original spatial node density distribution and lead to behavior similar to random mobility as exemplified by the Random Waypoint model. Based on these observations, we propose a simple mobility model that preserves the desired spatial density distribution. We found that performance results expressed by a number of network metrics also match closely results obtained under mobility governed by real traces. We also Introduce a modeling framework to analyze spatial node density in mobile networks under "waypoint"-like mobility regimes. The proposed framework is based on a set of first order ordinary differential equations (ODEs) that take as parameters (1) the probability of going from one subregion of the mobility domain to another and (2) the rate at which a node decides to leave a given subregion.

Since the social structure usually guides the choice of new destination in real human mobility, we introduce a user mobility modeling framework that accounts for both the users'

social structure as well as the geographic diversity of the region of interest. SAGA, or Socially– and Geography-Aware mobility model, captures social features through the use of *communities* which cluster users with similar features such as average time in a cell, average speed, and pause time. SAGA accounts for geographic diversity by considering that different communities exhibit different interests for different locales. Besides introducing SAGA, the contributions of this work includes a model calibration approach based on formal statistical procedures to extract social structures and geographical diversity from real traces and set SAGA's parameters.

To my wife:

Danielle Lopes Ferreira Astuto, first my great friend, now also my greatest love!

Te amo esposa! Chega de saudade. To voltando pra casa, doutor. :)

Acknowledgements

I want to thank my parents, Marcos Arouche Nunes and Miriam Cristina Astuto for their love and support. You gave me your names. Now I took the liberty of including three little letters in it... PhD, hope you don't mind. :) Could not have done it without you.

I would like to thank also my adviser Katia Obraczka for all the support and for believing in me.

Chapter 1

Introduction

Node mobility is a key factor in the design and performance evaluation of mobile networks and their protocols and mobility characterization has attracted considerable attention from the networking research community. Evaluation studies of early mobile networks and their protocols used most of the time "synthetic" mobility models such as random walk, Brownian motion, and the random way-point (RWP) model [15], just to mention a few. The RWP, in particular, has been one of the most used mobility models for evaluating mobile networks which motivated several studies that scrutinized its behavior, identified a number of undesirable features [69], as well as proposed variations to improve its behavior. Additionally, most network simulation platforms, which have been extensively used in the study of mobile networks, include mobility generators; i.e., they are able to generate mobility scenarios based on a specified mobility pattern (or mobility model) and a set of parameters (e.g., average speed, pause time, etc). The end result of these mobility generators is a sequence of positions for each node over the simulation time.

More recently, motivated in part by the problems associated with the RWP and recognizing the importance of employing more realistic mobility scenarios when designing and evaluating mobile networks, there has been considerable interest in using real mobility traces and developing models that reflect real mobility. Thus, network researchers and practitioners have been trying to use more realistic scenarios to drive the evaluation of wireless network protocols. This motivated initiatives such as the CRAWDAD [18] trace repository, which is an example of an initiative to make real mobility traces widely available to network researchers.

CRAWDAD stands for Community Resource for Archiving Wireless Data At Dartmouth, a wireless network data resource for the research community. This is a NSF-funded initiative to make real mobility, connectivity and traffic traces widely available to network researchers and practitioners and is a public repository that houses a wide variety of mobility traces. This archive has the capacity to store wireless trace data from many contributing locations, and staff to develop better tools for collecting, anonymizing, and analyzing the data.

Traces such as the ones found in repositories like this can then be used to run trace-driven simulations. However, even though initiatives like CRAWDAD have greatly increased availability of real traces, the number and variety of publicly-available mobility traces are 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.

In this context, since relying on real traces is not yet possible, mobility models are an indispensable tool in the design, testing, and evaluation of wireless networks and their protocols. This is the case for infrastructure-based networks (e.g., wireless LANs or WLANs), but even more so in infrastructure-less networks, a.k.a., wireless, self-organizing networks (WSONs) which include wireless mobile ad-hoc networks (MANETs), wireless sensor networks (WSNs) and disruption-tolerant networks (DTNs). In these networks, unlike their infrastructure-based counterparts where only the end user node is mobile, every node may have unlimited mobility. As an indication of the importance of mobility models to the study of wireless network protocols, most well-known network simulators include "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 [15].

To address this problem, a number of efforts have proposed mobility models based on realistic mobility patterns [49]. Notable examples include [7, 8, 60, 31]. Alternate approaches to developing realistic mobility generators try to use characteristics of real human mobility. For example, some efforts employ "transition probabilities", while others are based on maps of the area being simulated [70].

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. As mentioned in the related work section, the seminal work of Barabási and Albert [3] 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". As described in Section 2 which discusses related work in more detail, several recently proposed mobility models (e.g., [36, 37, 13, 47, 48, 41]), try to mimic real human mobility by following the preferential attachment principle: they define *attraction points*, whose probabilities of attracting other nodes increase as more nodes congregate around them.

In this work we focus on understanding human mobility and applying our findings on developing new realistic mobility models for simulations. We start by analyzing WLAN– and GPS– traces that record mobility in a variety of network environments. We observe that from a macroscopic level, human mobility is *symmetric*. We also study the direction of movement which also exhibits symmetric behavior in both real– as well as synthetic mobility.

Moreover, motivated by the symmetric behavior identified, we continued our investigation on real mobility characteristics, by focusing on node spatial density in real applications. We show that mobility models that select the next node position based on the position of other nodes, a la "preferential attachment", do not preserve the original spatial node density distribution and lead to behavior similar to random mobility as exemplified by the Random Waypoint model. We also show that this is an unwanted characteristic, since human mobility exhibits "persistent" behavior in terms of the spatial density distribution of the mobile nodes over time. By using real mobility traces, we observe that the original non-homogeneous node spatial density distribution, where some regions may be quite dense while others may be completely deserted, is maintained at different instants of time. To the best of our knowledge, this is the first time that these phenomena have been reported. Based on these observations, we propose a simple mobility model that preserves the desired spatial density distribution. We found that performance results expressed by a number of network metrics also match closely results obtained under mobility governed by real traces. We also compare our results to models whose steady-state do not preserve original non-homogeneous density distribution and show that network performance under such regimes deviates from performance under real trace mobility.

We continue our investigation to characterize and describe mobility through the *spatial density* of mobile nodes, given its significant impact on fundamental network properties, such as connectivity and capacity, as well as on core network functions, e.g., medium access

and routing. To date, only a few efforts have focused on modeling spatial density. Notable examples include [25, 9, 50]. However, most previous work have been focusing exclusively on the Random Waypoint (RWP) model [15]. In our work, we introduce an Ordinary Differential Equation (ODE) framework to mathematically model spatial node density under different "waypoint"-based mobility regimes. More specifically, our model describes node density's steady-state behavior under waypoint-based mobility which is characterized by having nodes probabilistically choose the next destination, or waypoint, based on some probability density function. We validate our model by using it to describe the steady-state behavior of real user mobility recorded by GPS traces in different scenarios and compare the results against the corresponding traces. Moreover, we present comparative results for steady-state spatial distribution analysis of a number of synthetic waypoint mobility regimes.

More recently, alternate approaches to developing realistic mobility regimes rely on social interactions between nodes to drive the selection of movement direction and destination [68, 72, 12, 27, 65, 21, 48]. These models usually construct a social network either using outside information unrelated to the mobility trace, or on the basis of geographical collocation.

We take a completely different approach to modeling human mobility, by proposing a framework that accounts for the social structure in the mobility traces through the inclusion of communities and the use of the preferential attachment principles to account for the interaction of nodes within and between communities. We thus refer to the proposed model as "socially-inspired". The proposed mobility modeling framework is also "geography-aware" as it accounts for the heterogeneity of the different locales in the region of interest. These two features allow our model to account for features observed in real life networks such as differential popularity, transitivity and clustering. Moreover, our modeling framework applies formal statistical learning techniques to extract, from real mobility traces, both the nodes' social structure as well as the region's geographical diversity for model calibration. We validate our work by applying it to real mobility scenarios and comparing our results against real measurements. We show that our model is able to preserve the desired non-uniform node spatial density, creating and maintaining clusters and accounting for differential node popularity and transitivity; it thus solves the problem exhibited by preferential attachment based mobility models.

The remainder of this document is organized as follows. Chapter 2 presents related work relevant to our goals, where we first discuss the study and statistical analysis of real live

traces and later on present some pertinent work on mobility modeling itself. In Chapter 3 we present our study on the symmetric behavior of human mobility. Further on, we discuss in Chapter 4 the invariance of spatial node density in real networks, where we propose a mobility model called Intensity WayPoint (IWP), that is able to mimic this invariance, and compare results on mobility and routing metrics to real traces and other synthetic models. We continue our study on spatial node density in Chapter 5, presenting our ODE framework for such purpose, validating it against real live traces and applying it to study node density in steady state in waypoint-based mobility. We present in Chapter 6 an extended and more complete version of the IWP model called SAGA, that is able to account for the social structure of mobile nodes differential popularity and transitivity, while maintaining the desired node spatial density. Finally, in Chapter 7 we present our final considerations.

Chapter 2

Previous Work

Over the past decade, mobility characterization has attracted considerable attention from the network research community. In this section, we briefly overview the efforts that are more closely related to our work. We start by revising several previous work on statistical analysis of real mobility traces and how researches have been using information from traces to build new and more realistic models. Later on in the following section, we discuss several previous work on mobility modeling itself focusing on research that is more closely related to our work.

2.1 Related Work on Analysis of Real Traces

Several empirical studies have focused on characterizing mobility behavior in WLANs. For example, the work reported in [7] studies a campus WLAN trace and extracts statistics such as the number of associations per time per AP as well as session duration. Their goal was to generate input for capacity planning. The MIT WLAN trace was used in [8], where statistics on traffic and AP associations were presented; the work also introduced important metrics such as *prevalence* and *persistence*. The study reported in [60] explored the Stanford traces but focused on traffic– and application-layer issues rather than mobility. Nonetheless, empirical distributions were derived for the number of users per AP and the maximum number of hand-offs per AP. Campus WLAN usage was further studied in [31] using the Dartmouth traces. Some of their findings include the fact that users tend to persist at a single location for longer and that different applications had different mobility characteristics. Dartmouth's, MIT's, and two other traces were studied further in [64] which presented ECDFs for time spent in each AP, number

of associations per user, and percentage of APs visited. Moreover, metrics such as clustering coefficient, degree of separation between nodes, and disconnection ratio were also reported.

Trace-driven approaches aimed at building realistic mobility models have been the focus of many research efforts. In [33], a weighted waypoint model is proposed based on a university campus mobile scenario containing 5 possible locations. The transition probabilities were set based on a survey conducted on site, where people were asked about their mobility patterns on campus. A transition matrix was built as a result of this survey; however, it did not exhibit any symmetry. We believe this was the case because the transitions did not take into consideration intermediate locations. In [70], a framework for building trace-driven mobility models was proposed. It uses a map-based model where transition probabilities are computed from route weights and depend on origin, destination, current and previous locations, all extracted from the Dartmouth trace. In a more recent study [47], a model where the next waypoint depends on cell popularity and distance from current position is proposed.

2.2 Related Work on Mobility Modeling

There has been considerable work on trying to depart from random mobility models towards more realistic ones applied to the evaluation of wireless networks and their protocols. In this section, we do not try to present a complete survey of the state-of-the-art on realistic mobility modeling; rather, our goal is to highlight some recent efforts that are more closely related to our work and thus motivated it.

Some existing approaches use RWP mobility as a starting point, especially due to its simplicity and ease of use and implementation. Another way to differentiate existing approaches is in terms of the type of network environment they target, namely infrastructure-based and infrastructure-less networks. An example of mobility model for infrastructure-based network, more specifically WLANs, is the work described in [64] which captures mobility as nodes move between access points. Parameters for the model are extracted from real WLAN traces. In [37], Kalman Filters (KF) are used to extract paths taken by nodes as they move from a hotspot to another in the Dartmouth campus. The waypoints in the path between hotspots are given by the KF for the map of that specific scenario. Hotspot regions are identified using a Gaussian distribution at each pause location, weighted by the time a node spends on each hotspot and then averaged over all nodes. Our mobility model, IWP, tries to yield high fidelity to real mobility

and can be applied to both types of networks.

As mentioned above, several efforts have focused on proposing variants of RWP mobility. For instance, to avoid RWP's "zig-zag" behavior, in [10], a model which avoids sudden accelerations and sharp turns during direction changes has been proposed. Spatial node distribution of the RWP model has been studied in [9], where analytical expressions of the expected node distribution in one dimension have been derived as well as an approximation for the two-dimensional case. Another study of node density has been presented in [55], where real traces of vehicular mobility in a city have been used to find correlation between node density and speed. A model called "heterogeneous random walk" has been proposed, and node clustering under such mobility regime has been studied.

The scale-free properties of real networks were investigated in several studies. An important milestone on the subject is the work by Barabási and Albert [3] where it has been demonstrated that many real large networks are scale free, that is, the node degree in the network graph follows a power law. The authors discuss the mechanism responsible for the emergence of scale-free networks and argue that understanding this problem will require a shift from modeling network topology to modeling "network assembly and evolution". To this end, they define the Barabási-Albert model based on *growth* and *preferential attachment*. Growth refers to the fact that the number of nodes in the network increases over time, where a new node is placed with m edges connecting it to other m nodes. Preferential attachment means that a node will choose to connect to another node i with probability $\Pi(k_i) = \frac{k_i}{\sum_j k_j}$ based on the degree k_i of node i and any node j connected to node i.

Several mobility models have been inspired by the Barabási-Albert preferential attachment principle. For example, in [13], a model based on preferential attachment has been proposed, where the choice of going towards an attraction region is weighted proportionally to the region's popularity (i.e., the number of other nodes that chose it) and inversely proportional to the distance to it. The work proposed in [41] is another example of a model that follows Barabási-Albert's growth and preferential attachment principles. The authors even show a figure where they present their initial (after growth) and steady-state spatial distribution. It is possible to see how clusters dissipate and fade away over time. The same concept is also used in [47] where nodes are also driven by identified social interactions. Authors validate their approach by means of showing the power-law exponential decay of inter-contact time among the

formed communities and compare it with measurements in real traces.

Bettstetter et al. point out that random mobility leads to homogeneous node distributions. In [11], they propose a method that creates initial non-homogeneous node distributions and in [25], a variant of RWP mobility that maintains the non-homogeneity of an original node distribution.

Map based approaches are interesting in the sense that they limit a node's roaming area, which is what happens in reality (i.e., vehicles in roads, people walking in trails in a park, campus streets and corridors in a conference or fair). For example, in [70], transitions between different locations in a campus map are modeled. The map is represented by a graph and users move between nodes of the graph, representing access points, buildings, and intersections in the map.

More recent approaches focus on the aspects of social interaction between nodes. Other examples of mobility modeling for social networks include the work in [28] and [48], where node movement is influenced by the strength of social ties and the choice of an attraction point is based on the history of visits of other nodes to that location.

Chapter 3

On the Symmetry of User Mobility in Wireless Networks

Motivated by the trend towards employing real mobility to design and evaluate wireless networks, we study different types of traces obtained by recording user mobility. Our goal is to identify patterns, extract features, and define metrics to characterize the *spatial* behavior of human mobility. As a result, we identify an interesting characteristic exhibited by human mobility that, to the best of our knowledge, have not yet been revealed in previous studies. We work with a number of real traces that record user mobility in infrastructure-based networked environments (i.e., wireless LANs) as well as GPS positioning traces. We show that, from a macroscopic level, human mobility is *symmetric*. In other words, the number of users that move from point A to point B approximates the number of users that go in the opposite direction, i.e., from B to A. We also show that this type of symmetry is more accentuated in synthetic mobility models, in particular, in random way-point mobility. Additionally, we study the direction of movement which also exhibit symmetric behavior in both real– as well as synthetic mobility. In order to quantify the degree of symmetry exhibited by a given mobility scenario, we define a new metric we call *coefficient of symmetry*. Finally, we hope that this work will generate input for the development of new and more realistic models, and that the new metrics presented here will aid in the evaluation of these models.

3.1 Mobility Traces

In this section we describe the mobility traces we used in our study. Here, we make a distinction between different types of traces and define *macro–* and *micro–mobility*. In *macro-mobility* traces, the resolution of node positions is coarser grained; this is the case of WLAN traces or cell phone traces, where the exact position of a node is unknown and is instead represented by the position of the access point with which the node is associated (in WLANs) or the cell in which the node is presently located (in cellular networks). In *micro-mobility* traces, a node's location is represented with finer grain resolution; this is the case of GPS traces, where positioning information is given by latitude and longitude, and typically has a resolution of meters.

We considered three types of mobility traces in our study, namely: infrastructure-based scenarios (WLANs); ad-hoc scenarios (using GPS positioning), and synthetic scenarios (using RWP mobility). Below we present the traces in these different categories in more detail.

3.1.1 WLAN Traces

Table 3.1 summarizes the WLAN traces in terms of number of users, number of Access Points (APs), and duration of the trace. The first 3 WLAN traces were collected in university campus environments: Dartmouth and Stanford, available at [18] and the MIT trace [8].

Trace	# users	# APs	Duration
Dartmouth	9480	623	1 year (2003)
Stanford	74	21	12 weeks
MIT	1366	173	1 day
Rio	120	17	1 week

Table 3.1: WLAN traces description.

Since our goal is to study human mobility in general, not only in the context of a campus environment, we collected a fourth trace recording user mobility in a public WLAN network deployed in an urban environment. The *Rio* trace was collected during the first week of April 2010 recording user mobility in the *Digital Orla Project* [16] network deployed along Ipanema Beach, Rio de Janeiro, Brazil. At the time the trace was collected, the network consisted of 17 Cisco APs mounted on light poles along the beach's boardwalk, providing network

access in the boardwalk and nearby streets. The APs were configured to transmit a *syslog* message every time a client associated or disassociated. The resulting trace reflects one week of network usage; a total of 120 distinct clients who performed at least 1 transition between APs were recorded. Clients use a variety of devices to connect to the network including laptops and smart phones.

3.1.2 GPS Traces

The focus here was to study user mobility with finer-grained resolution. WLAN traces represent "on-off" type behavior, where users connect to an AP, stay connected for a given period of time, disconnects, stays disconnected for a period of time, reconnects again to some other AP, and so on. GPS traces, represent mobility in a more continuous fashion where nodes roam around a given area, and connection between nodes and an AP or another node is a function of the distance between the nodes and the transmission range of their communication devices. Table 6.2 summarizes the GPS traces we used in our study in terms of number of users, trace duration, and the period between GPS samples.

Trace	# users	Duration	Samples
Quinta	98	900 s	1 s
KAIST	78	5000 s	10 s

Table 3.2: Description of the GPS traces studied.

Quinta, refers to the "Quinta da Boa Vista Park" trace, presented in [6]. It is a GPS trace collected at a park in the city of Rio de Janeiro, Brazil, that has many trees, lakes, caves and trails, and holds the National Museum and the city Zoo. The KAIST trace, also available at [18], on the other hand, is a GPS trace collected on a university campus environment.

3.1.3 Simulated Traces

Two synthetic traces were generated using the RWP mobility model. We tried to simulate the two scenarios described by the traces in Section 6.4 using RWP mobility. Our goal here is to compare characteristics of synthetic mobility against real mobility, especially when the former is inspired by the latter. In our simulations, we set the velocity range in a way that the average velocities would match the ones measured in the Quinta and KAIST traces. The same

was done for average pause time and the dimensions of the area covered by the traces. Velocity and pause time were chosen according to a *uniform distribution* whose average is computed as described above [1]. Table 5.2 summarizes the parameters used to generate the synthetic traces.

Parameter	Quinta_RWP	KAIST_RWP
Avg. Speed	1.2 m/s	0.72 m/s
Avg. Pause	3.6 s	17 s
Area	840 m X 840 m	5000 m X 5000 m
Duration	900 s	5000 s
# nodes	98	78

Table 3.3: Simulation parameters.

In order to study transitions we divided the area into square cells. We assumed 100 meters of range for a Wi-Fi device and set the cell to be the square inscribed in a 200 meters diameter circle. That gives approximately a square cell of size 140 X 140 meters.

3.2 Dataset Analysis

In this section, we present the methodology we used to analyze the traces and the results of our study. We start by defining the metrics and parameters employed in our analysis.

3.2.1 Definitions

Transition is the event caused when a node moves from one cell to another. When analyzing the GPS traces, transitions are detected when nodes move between neighboring cells. In WLAN traces, transitions happen between any cells.

Transition Matrix (TM) A is a square matrix of N x N elements where N is the total number of cells in the system. Every element a_{ij}, denotes the number of transitions observed from cell i to cell j. Since we do not consider transitions when nodes move within the same cell, the main diagonal is zero.

Time-Aggregated Transition Matrix For a given window of time, we count how many transitions were observed for a given user. The larger this window, the more we aggregate transitions in time. The window can be *sliding* or *fixed* over time. If we choose a sliding window, we define a window size W and a starting time instant, say $t_{1,i}$, when the first transition for user i

[1] For more details on how these parameters are derived from real traces refer to [6].

occurred. We then count how many transitions occurred from $t_{1,i}$ to $t_{1,i} + W$. After that, the beginning of the window slides to the time of the next transition, say $t_{2,i}$ and the transitions between $t_{2,i}$ to $t_{2,i} + W$ are counted. This is done for all transitions for the duration of the trace. In the case of a fixed-window, the trace is divided in slots of size W and we count the transitions that happen inside every slot. We used the fixed-window approach to compute aggregated transitions in the results presented below.

User-Aggregated Transition Matrix The TM can be computed for each user in the system separately, called here individual transition matrix (iTM), or for a number of users n, an *n-aggregated transition matrix* (n-aTM). For a time window of fixed size, user aggregation refers to how many users we consider when computing the TM.

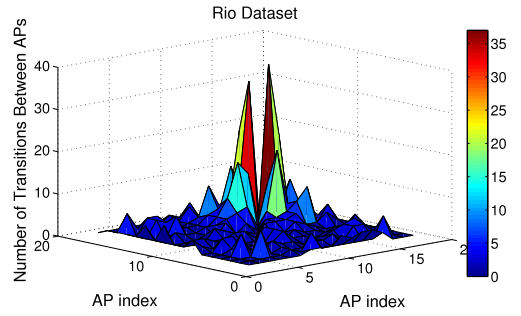
Symmetric Transition Matrix A is considered symmetric when $a_{ij} \approx a_{ji}, \forall i, j \in \{1, ..., N\}$.

Direction Angle is the counter-clockwise angle formed between the X-axis (0°) and the line connecting two consecutive location samples in the trace.

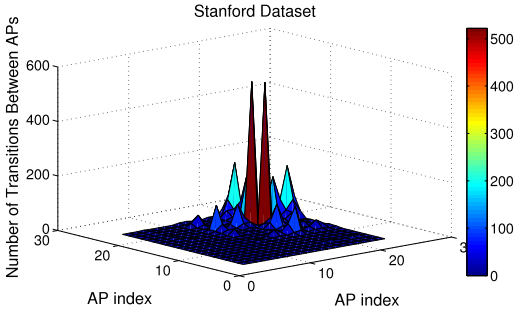
3.2.2 Transition Symmetry

Figure 3.1 gives the aggregated TMs. Figures 3.1(a), 3.1(b), and 3.1(d) show the aggregated TMs for the Rio, Stanford, and Dartmouth datasets, respectively. These TMs consider transitions within 24 hours and aggregate all users for the duration of the trace according to Table 3.1.

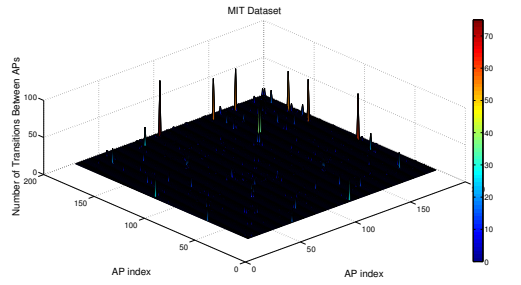
The aggregated TMs show a clear symmetry in relation to the main diagonal. One might argue that the symmetric behavior could be a consequence of the *ping-pong effect* reported in [70], commonly seen in WLAN traces. This phenomenon occurs when a mobile node is in range of two or more APs, and because of the variation in signal strength, this node might go back and forth associating and disassociating with the different APs in range. It is easy to identify this behavior if the session time is in the order of seconds, but when the gap between transitions is in the order of minutes, this task becomes non-trivial. In order to mitigate this behavior, we filtered high frequency transitions, only accounting for transitions that happened more than 1, 5, 10 and 15 minutes away from each other. As we increased the time gap, the number of transitions per node decreased as expected. However, it did not affect the symmetric behavior of the TM. For that reason, we set the *ping-pong filter* to 10 min for all the results



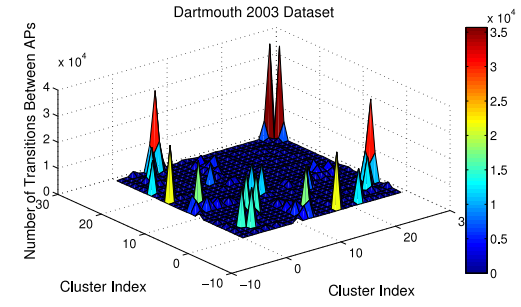
(a) Rio Dataset



(b) Stanford Dataset



(c) MIT Dataset



(d) Dartmouth Dataset.

Figure 3.1: Transition matrices for the WLAN traces.

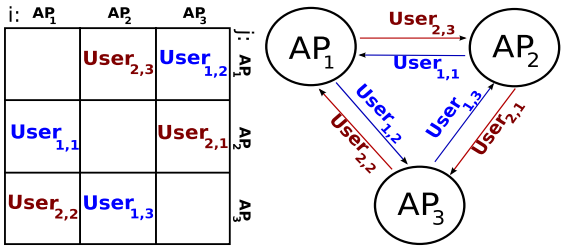


Figure 3.2: Transition matrix built for a network of 3 APs, after two users make 3 transitions each in opposite directions.

presented here.

In order to further explore the reasons for the TMs' symmetric behavior, we considered clusters of APs such that if a transition occurred between APs in the same cluster, either due to node movement or the *ping-pong* effect, that would not count as a transition. In other words, only transitions between clusters were considered. We use AP clustering in the Dartmouth trace, which was the one with the largest number of APs. A map of the Dartmouth campus was used to decide how to cluster the APs: we used the known locations of the APs to overlay them on the map; we then clustered the APs according to proximity and area of coverage (instead of using clusters of similar size and number of APs). The Dartmouth TM represented in Figure 3.1(d) is a result of organizing the campus APs in 30 clusters. Evidently, symmetry persisted even after filtering out high frequency transitions and using AP clusters.

The explanation for the symmetric behavior is illustrated in Figure 3.2. If $user_1$ takes the path from AP_2 to AP_1 to AP_3 and $user_2$ goes from AP_2 to AP_3 to AP_1, then the resulting transition matrix is symmetric. The more we aggregate users and/or the longer we observe their movement, the higher is the chance of observing movement in opposite directions.

In order to understand and be able to quantify symmetry as we observe user mobility collectively and over time, we define a metric called *coefficient of symmetry*, α, where $0 \leq \alpha \leq 1$, as defined in Equation 3.1:

$$\alpha = \frac{1}{T} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} |a_{ij} - a_{ji}| \tag{3.1}$$

where T is the total number of transitions in matrix A. Thus, the closer α is to 0, the higher the symmetry. $\alpha = 1$ means that A does not have any transitions in symmetric positions. This metric can also be seen as the percentage of transitions happening in symmetric positions.

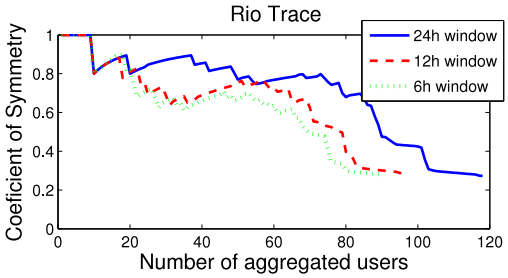
For example, if $\alpha = 0.2$, it means that 80% of the transitions are in symmetric positions. α can be computed for each node individually, or for the entire network from a TM aggregated over all users.

Figure 3.3 shows α for all 4 WLAN traces studied, varying the number of aggregated users and aggregating transitions in time for 3 different window sizes: 24–, 12–, and 6 hours. The X-axis is how many users were aggregated to compute α, and users are ordered in decreasing order, according to their individual α (computed using their iTM).

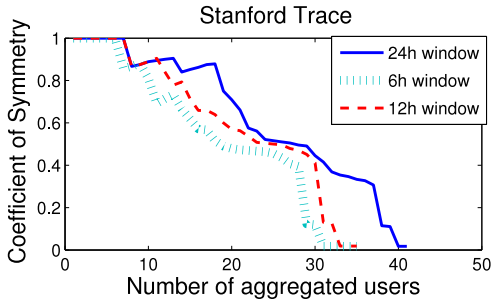
For the traces with larger number of users, when we decrease the aggregation time window, the TM becomes less symmetric (i.e., α increases), as we can observe from the Dartmouth TM (Figure 3.3(d)). This behavior is not as evident in the MIT trace (Figure 3.3(c)). This is due to the fact that all the APs are in the same building which greatly contributes to the ping-pong effect, making it much harder to filter. Therefore, decreasing the aggregation window does not affect the results as much. This happens because users' transitions, in this case, occur with great frequency, always falling inside the minimum window used (6 hours). To confirm this hypothesis we used an additional window of 30 minutes for the MIT dataset. Only then we were able to notice the difference, as the 30 minutes curve went over the other curves as we aggregated more users.

In the case of the traces with fewer users, time aggregation appears to have an opposite effect on the symmetry, as we can see in Figures 3.3(a) and 3.3(b). The smaller the time window is, the less transitions are accounted in the TM. In the case of those two traces, there is a preference for a few specific APs (more popular APs appear as the spikes in the TMs on Figures 3.1(a) and 3.1(b)). The transitions not accounted for due to the smaller time window are more likely to be taken from the most popular APs, what would decrease the difference amongst the cells in the matrix. This would increase the overall symmetric pattern, contributing to a lower α and making the α curve converge faster to lower values as we aggregate more users.

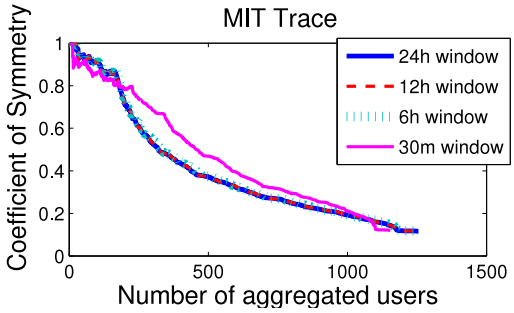
In Figure 3.4 we plot the coefficient of symmetry, α, for the GPS and RWP traces. For the RWP traces, we plot the average aggregated α for 15 different random seeds. The first interesting observation is that α for the synthetic traces is similar to α for the corresponding real traces. The differences are due to the fact that in real environments, users tend to favor some areas (e.g., they walk on roads), limiting the degree of freedom of their movement. On the other



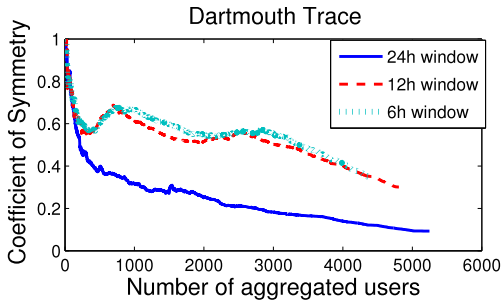
(a) Rio Dataset



(b) Stanford Dataset



(c) MIT Dataset



(d) Dartmouth Dataset

Figure 3.3: Metric of symmetry α for different number of aggregated users, under different time aggregation windows for (a) Rio, (b) Stanford, (c) MIT and (d) Dartmouth datasets.

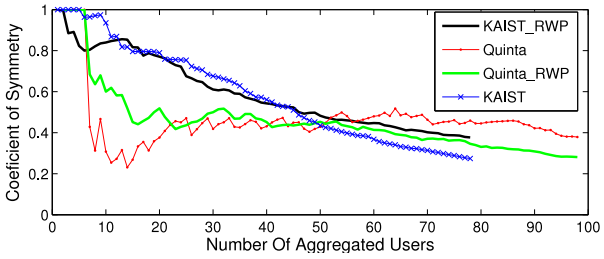


Figure 3.4: α per number of aggregated users, under a 24h aggregation window.

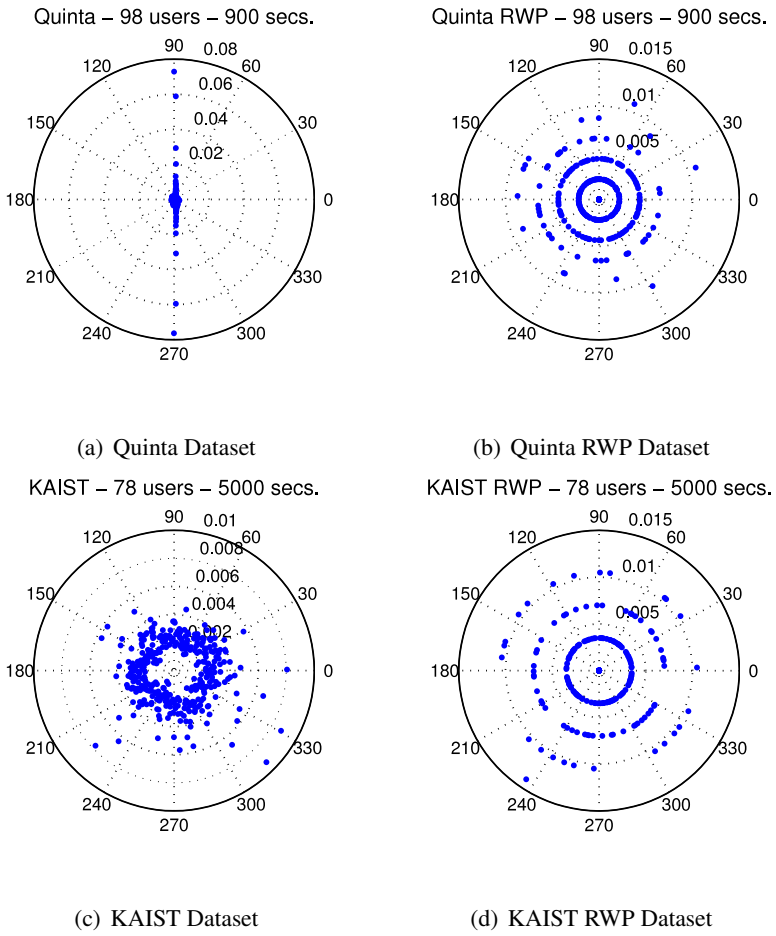
hand, for the synthetic traces, users tend to disperse more uniformly increasing the chances of the number of transitions between a cell and its neighbors to be similar for every cell. For the KAIST trace, in particular, we notice that nodes tend to concentrate even more than in the Quinta trace, visiting less cells overall. That explains KAIST's higher symmetry, even with less users. This limited number of visited cells, in the KAIST trace, increases the relation user/cells increasing the chances of random transitions falling in symmetric positions.

3.2.3 Direction Angle Symmetry

Direction of movement provides additional information on the spatial behavior of human mobility. Figure 3.5 shows the distribution of direction angles which represent relative frequency of directions users take when moving as reported in the GPS– and synthetic traces. The bin size used in these plots is 1°. We observe a symmetric–, close to uniform distribution of direction angles for both of the synthetic traces (Figures 3.5(b) and 3.5(d)), where it is not possible to identify any significant preference for a given direction.

If instead of looking at transitions, we look at the direction nodes take when moving inside the network area, the distribution of angles taken by the aggregated nodes are also symmetric. Figure 5 shows this symmetry and the interesting patterns that arise when plotting the PDF of the angles users take while moving, weighted by the duration of the movement in each direction. The bin size used in this plots is 1°. This results show the symmetric, close to uniform distribution of the synthetic traces (Figures 3.5(b) and 3.5(d)), where it is not possible to identify any significant preference for a given direction.

For the KAIST trace, in Figure 3.5(c), it is possible to notice the concentration of points over a symmetric shape, like a circle. Despite some few outliers, it is also not possible to identify any significant preferred angle. That is not true for the Quinta trace in Figure 3.5(a).



(a) Quinta Dataset

(b) Quinta RWP Dataset

(c) KAIST Dataset

(d) KAIST RWP Dataset

Figure 3.5: Distribution of direction angles for (a) Quinta, (b) Quinta RWP, (c) KAIST and (d) KAIST RWP datasets.

This trace has a peculiar symmetric shape with respect to 180 degrees, for the direction nodes take. This can be explained by considering the physical terrain and the fact that users mostly move only along roads and tracks which limit the possible movement directions.

The KAIST trace, on the other hand, was collected on a much larger area which corresponds to the KAIST campus, with several roads and pathways. Because it was derived from the real KAIST trace, the KAIST_RWP trace also presents this characteristic. This is what is shown in Figure 3.6 which plots the Empirical Cumulative Distribution Function (ECDF) of the direction angles for the GPS– and synthetic traces. We observe that the KAIST_RWP trace is able to reflect closely the statistical properties of the trajectory angle. The differences between Quinta and Quinta_RWP traces are mainly due to movement restrictions (obstacles and

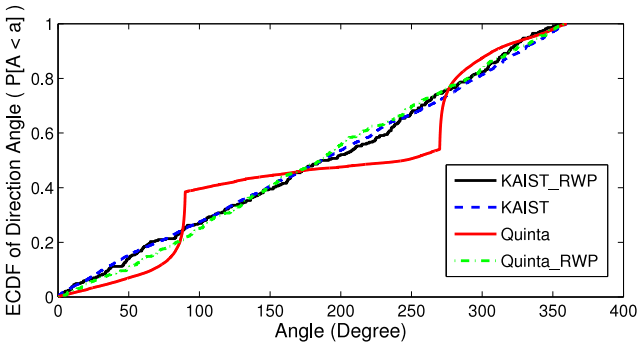


Figure 3.6: ECDF of the direction for GPS and RWP traces.

paths).

3.3 Discussion

Despite the large body of work that focus on studying real traces, to the best of our knowledge, our work is the first to observe and quantify spatial symmetry in user mobility in terms of cell transitions and angle of movement. The only other work that briefly alludes to direction angle symmetry is [37].

Using maps, social interaction,"hot spots", and other factors that limit and/or influence user mobility (e.g., what is the next destination) are effective strategies to "bias" synthetic mobility models in order to obtain behavior that is closer to reality. Along those lines, the proposed TM can be used to identify the "popularity" of the different regions of the network (including, e.g., hot spots) and based on this information specify the probability distributions used by nodes to select their next destination in synthetic mobility models.

Another interesting application of our results is the use of the direction angle symmetry as a way to compare mobility traces. The distributions of direction angles in Figures 3.5(a) and (b) show very different behavior when comparing the synthetic trace against the real one. By using strategies to approximate synthetic mobility models to real mobility, direction angle symmetry can be used to quantify how close the resulting models get to real mobility, and thus how effective these "approximation" strategies are. We are currently going a step further and trying to draw a transitive relationship between direction angle symmetry, spatial mobility behavior, and performance of core networking functions such as routing and forwarding.

Mobility or location prediction is yet another area in which our results can find interesting applications. For example, in [59], a Markovian predictor is used to predict user location based on a given mobility trace (in the paper, the Dartmouth trace was used). It was observed that for nodes that did not have enough historical information recorded in the trace, the performance of the predictor was not adequate. To mitigate this problem, we could use the coefficient of symmetry, α, of a given node to artificially "re-create" the node's past historical information.

3.4 Conclusions

We so far have explored the spatial behavior of human mobility through a variety of mobility traces collected in different network environments. As a result, we identify characteristics exhibited by human mobility that, to the best of our knowledge, have not yet been revealed in previous studies. We used a number of real traces that record user mobility in wireless LANs as well as GPS traces. Additionally, we also investigated synthetic mobility, in particular, through traces generated using the RWP model. We show that, from a macroscopic level, human mobility is *symmetric*. In other words, the number of users that move from point A to point B approximates the number of users that go in the opposite direction. We also show that this type of symmetry is more evident in synthetic models such as RWP. Additionally, we study the direction of movement which also exhibit symmetric behavior in both real– as well as synthetic mobility. In order to quantify the degree of symmetry exhibited by a given mobility scenario, we define a new metric we call *coefficient of symmetry*. We conclude by discussing the implications of our results in different areas of mobile networking.

In the next chapter we extend our study on human spatial behavior and, motivated by our findings about symmetric mobility, we focus on node spatial density. We theorize that the symmetric behavior of transitions may influence node density. We then present a study on node density in the following.

Acknowledgements

I would like to thank Bruno Sanso, Athanasios Kottas and Kumar Viswanath for their invaluable contribution to the work presented in this section.

Chapter 4

On the Invariance of Spatial Node Density for Realistic Mobility Modeling

In this chapter we show that human mobility exhibits "persistent" behavior in terms of the spatial density distribution of the mobile nodes over time. Using real mobility traces, we observe that the original non-homogeneous node spatial density distribution, where some regions may be quite dense while others may be completely deserted, is maintained at different instants of time. We also show that mobility models that select the next node position based on the position of other nodes, a la "preferential attachment", do not preserve the original spatial node density distribution and lead to behavior similar to random mobility as exemplified by the Random Waypoint model. To the best of our knowledge, this is the first time that these phenomena have been reported. Based on these observations, we propose a simple mobility model that preserves the desired spatial density distribution. Moreover, when simulating the operation of a network moving according to the proposed model, we found that performance results expressed by a number of network metrics also match closely results obtained under mobility governed by real traces. We also compare our results to models whose steady-state do not preserve the original non-homogeneous density distribution and show that network performance under such regimes deviates from performance under real trace mobility.

4.1 Introduction

In Section 2.2 we briefly presented the concept of preferential attachment, introduced by Barabási and Albert in their seminal work [3]. In this chapter we show that using the preferential attachment principle for modeling human mobility leads to undesirable steady-state behavior, which does not preserve the original spatial node density distribution and leads to behavior similar to random mobility as exemplified by the Random Waypoint model.

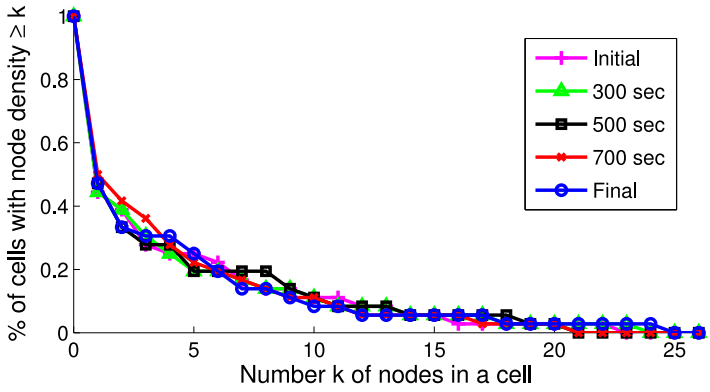


Figure 4.1: Node spatial density distribution at different trace collection times for mobility in a city park.

Instead, real human mobility exhibits "persistent" density heterogeneity as exemplified in Figure 4.1. This figure shows the spatial density distribution for one of the traces used in our work, which was collected in the Quinta da Boa Vista Park in Rio de Janeiro, Brazil. This distribution is defined as the percentage of cells containing $\geq k$ nodes. The 5 curves in the plot refer to the initial distribution and distributions at instants 300, 500, 700, and 900 seconds, which is the end of the trace collection interval. From the graph we observe that the density distribution does not vary much with time: the largest deviation from the initial distribution for any value of k at any instant is 8.3%; the average deviation from the initial distribution measured in all the instants for all values of k is 1.27%. Similar observations can be drawn from the other traces used in our work as reported in Section 6.4. We should point out that, from a survey we conducted on 82 papers published in various networking conferences between 1996 and 2010 which used some form of simulation study, only 57 papers listed simulation duration time. Out of those, 21 papers used 900s, 12 papers used 500s, 4 used 600s, and 3 used 300s. All others reported different times all lower than 300s. This means that the time windows we used here to

study human mobility are representative of– and go beyond the simulation times used to study wireless networks and their protocols.

Based on our observations of the persistent density distribution behavior in human mobility, we then propose a simple, yet realistic model of human mobility called *Intensity Waypoint*, or *IWP*. Besides preserving, by design, the invariance of the node density distribution over time, IWP is simple and easy to use and implement. We argue that these latter two features of IWP are critical for its widespread use and adoption, vis-a-vis the RWP model. Another important feature of IWP is its generality, i.e., it can be used to model mobility in any kind of networking environment including Wireless LANs or ad hoc networks.

We evaluate IWP in two different ways: first, we investigate its ability to adhere to the node density distribution invariance property. To this end, we define the *node density distribution* metric (mentioned above) and use it to compare IWP against RWP and the Natural Mobility Model [13], a representative of mobility models based on preferential attachment. In this study, we use three GPS-based mobility traces collected in different scenarios, namely at a park, a University campus, and a state fair. The traces were used to set the parameters of the models as well as the baseline for the comparative study. The second part of the evaluation focuses on studying the performance of core network functions, in this case routing, under IWP mobility. The same traces described above were used as baseline to compare IWP against RWP and the Natural Mobility Model. Our results show that, by preserving the original node density heterogeneity, IWP is able to closely approximate real network behavior according to a number of routing performance metrics.

In summary, to the best of our knowledge, this is the first time that the invariance of the non-homogeneous spatial density distribution of human mobility has been observed. This is also the first work to point out that mobility models based on the preferential attachment principle does not preserve the original spatial node density distribution and lead to behavior similar to random mobility, a la Random Waypoint model. Based on these observations, we propose a simple model that preserves the desired spatial density distribution, as well as show its impact on core network functions such as routing relative to real mobility traces, the RWP, and a preferential-attachment based model.

4.2 Intensity Waypoint Mobility

The main design principles guiding the proposed model are: (1) preserving the invariance of the original node density distribution and consequently its inherent non-homogeneity, (2) generality, i.e., applicability to different types of network scenarios, and (3) simplicity and ease of use in order to facilitate the model's adoption.

In our model, we divide the simulation area in equal sized squares, that we refer to as cells. Each cell $i \in \{1..N\}$, where N is the total number of cells composing the simulation area, is assigned an intensity $\mu_{i,t}$ at time t given by an *Intensity Map (IM)*. The IM is a vector composed of N elements, where each element has a value $\{\mu_{i,t} \in \mathbb{R} \mid \mu_{i,t} \geq 0\}$ that indicates how intense the activity in a cell is. We define the IM and how we build it in Section 4.5.

The probability $p_{i,t}$ that cell i is chosen as the next waypoint at time t is given by Equation 4.1 below:

$$p_{i,t} = \frac{\mu_{i,t}}{\sum_{j=1}^{N} \mu_{j,t}}, \forall i \in \{1..N\} \tag{4.1}$$

The evolution of the spatial node density distribution over time is an important aspect of the network, but as our results show, this distribution does not vary much with time. For that reason, we make $\mu_{i,t}$ and $p_{i,t}$ time-independent; thus, $p_{i,t} = p_{i,t'} = p_i$.

We define a step function as in Equation 4.2. This function will allow us to set up the attraction areas by simply setting the values of x_0 to the physical limits of a given attraction area or *cell* i: $x_{i,min}$, $x_{i,max}$, $y_{i,min}$ and $y_{i,max}$. The cell's area A_i is given by $(x_{i,max} - x_{i,min}) \times (y_{i,max} - y_{i,min})$.

$$u(x - x_0) = \begin{cases} 1 & \text{if } x \geq x_0 \\ 0 & \text{if } x < x_0 \end{cases} \tag{4.2}$$

Thus, the probability distribution of selecting the next destination (x_d, y_d) is defined by Equation 4.3 as follows:

$$f_{(x_d,y_d)} = \frac{1}{\sum_{i=1}^{N} p_i A_i}\Bigg[\sum_{i=1}^{N} p_i\Bigg[\Big(u(x_d - x_{i,min}) - u(x_d - x_{i,max})\Big) \Big(u(y_d - y_{i,min}) - u(y_d - y_{i,max})\Big)\Bigg]\Bigg] \tag{4.3}$$

According to Equation 4.3, in order to select a node's new destination, we need a mechanism that allows us to differentiate the choice of the next way-point based on the intensity of each cell. We turn to a *Fitness Proportionate Selection*-based scheme known as *roulette-wheel selection* for that. This strategy is commonly used in genetic algorithms [54].

Based on the values of p_i, a circular wheel is divided into N sectors, like a pie chart. The center angle of the i-th sector is $2\pi p_i$, which makes the area of each sector proportional to the probability p_i. The roulette wheel selection strategy can then be described in two steps, every time the next cell is to be selected. First, generate a random number n uniformly distributed, where $0 \leq n \leq 1$ and second, if $p_1 + p_2 + \ldots + p_i \leq n < p_1 + p_2 + \ldots + p_{i+1}$, then cell i is selected to be the next one visited by the node. This is equivalent to spinning the roulette and having it stop at sector i. The bigger p_i is, the bigger the sector and the bigger are the chances of sector i being chosen. Once the new cell i is known a coordinate (x, y) is then uniformly chosen inside of cell i, and $x_{i,min} \leq x \leq x_{i,max}$ and $y_{i,min} \leq y \leq y_{i,max}$. Hence, the cumulative probability q_i of choosing a cell is given by Equation 4.4 .

$$q_i = \sum_{j=1}^{i} p_j, \forall i \in \{1..N\} \tag{4.4}$$

The same mechanism described above can also be adopted for choosing nodes' velocity and pause time. In this case, instead of an IM we would use a velocity domain vector $V = \{\nu_1, .., \nu_{n_v}, .., \nu_{N_v}\}$ containing all the N_v instances ν_{n_v} of nodes' velocity measured in the traces and a velocity intensity vector $F_v = \{f_1, .., f_{n_v}, .., f_{N_v}\}$ containing the frequency f_{n_v} of all velocity instances ν_{n_v}. Analogously, we would also have a pause time domain vector $P = \{t_1, .., t_{n_p}, .., t_{N_p}\}$ containing all the N_p instances t_{n_p} of nodes' pause time and a pause time intensity vector $F_p = \{f_1, .., f_{n_p}, .., f_{N_p}\}$ containing the frequency f_{n_p} of all pause time instances t_{n_p}. The elements of V and P would identify the sectors of the roulette-wheel, where

velocity and pause time wheels would have N_v and N_p sectors respectively, and the elements of F_v and F_p would determine the size of each sector. However, we wanted to investigate the formation of clusters and the non-uniform nature of node density in real scenarios and its impacts on networking. Thus, we focus on the intensity of attraction areas and use uniform distributions to model node velocity and pause time.

4.3 Building the Intensity Map

Acknowledging the challenge in defining what are *real* mobility characteristics, we approach this subject by extrapolating reality from the observation of real human mobility. To that end, we extract node density characteristics form the traces and represent it in our model in the Intensity Map, or IM. The IM can be viewed as a "density map", i.e., a map of the region over which nodes move showing the node density at each cell. In this section, we describe how to build the IM from real traces. Note that once an initial IM is built for a specific mobility scenario, variants of that scenario can be generated by controlling the density distribution reported in the trace-based IM. For example, if one wants to further intensify certain cells or increase/decrease the number of "hotspots", that can be easily accomplished by adjusting the IM accordingly.

In this work, we use three GPS traces collected in scenarios that are quite diverse, namely a park, a University campus, and a fair. These traces are presented in detail below. Then, we describe how we extracted density information from the traces and use this information to drive IWP mobility.

4.4 Real Traces

Table 4.1 summarizes the GPS traces in terms of number of users, duration of the trace, and the GPS sampling period.

Trace	# users	Duration	Samples
Quinta [6]	98	900s	1s
KAIST [56]	78	5000s	10s
Statefair [56]	19	8000s	10s

Table 4.1: Summary of the GPS traces studied.

Quinta, refers to the "Quinta da Boa Vista Park" trace, first presented in [6]. It is a GPS trace collected at a park in the city of Rio de Janeiro, Brazil. The park has many trees, lakes, caves, and trails. It houses the National Museum of Natural History and the city Zoo. The *KAIST* trace [56], on the other hand, is a GPS trace collected at a University campus environment (KAIST) in Daejeon, South Korea. The *Statefair* trace, also available at [56], is yet another mobility scenario showing daily GPS track logs collected from the NC State Fair held in North Carolina, USA.

Given that the traces of some of the nodes presented some discontinuity, we selected sections of the raw traces where no discontinuity occurred. We selected only nodes where we could find a continuous sequence of GPS fixes that were as long as 900, 5000, and 8000 seconds for the Quinta, KAIST, and Statefair traces, respectively. These were the total durations of the traces.

4.5 Extracting Information from Traces

We extract from the traces the distributions of speed, pause time, and node density by using the trace's sampling period. For example, in the Quinta trace, the sampling period is $T = 1$ seconds. A histogram was built with this information, for speed and pause time, giving us the relative frequency at which each value occurs. 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 2 meters for KAIST and Statefair traces and 0.5 meter for the Quinta trace, due to jitter in GPS update frequency.

To extract spatial node density, the area is divided into squared cells of 50 x 50 meters. Using this particular cell size to generate the density map was decided empirically in our experiments: it provided both adequate resolution as well as node clustering. An alternative approach could be identifying specific attraction zones, as was done in [37]. This is one of the topics of future work we plan to address. For this first version of IWP, a fixed size cell was used.

With the area divided into cells, we took a snapshot of the nodes' spatial distribution in the area being simulated every T seconds. The value of $T = 10$ was used since, for the size of the cells and the speeds sampled from the traces, a node could not on average change between

more than two cells during T. For every cell, at every interval T we counted the number of nodes in each cell. We then averaged the number of nodes in each cell over the course of the whole duration of the trace, giving us what we will call here, a *Intensity Map (IM)*. The IM gives the popularity of each cell, i.e., how intense is the node mobility activity in a cell, or in other words, how interested nodes are in general in each cell.

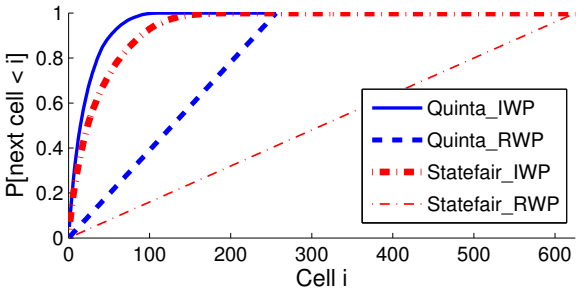


Figure 4.2: Cumulative cell selection distribution.

The cumulative probability of choosing a cell, given by Equation 4.4 can be visualized in Figure 4.2 for Quinta and Statefair scenarios and for RWP (uniform) and IWP (set by the IM extracted from the traces). The curves for the KAIST scenario are similar, but were omitted for the sake of clarity of the plot.

We acknowledge that always relying on extracting the IM from available traces can be somewhat constraining since there is no large and diverse enough set of publicly available traces. In general, the ones that exist represent a limited set of scenarios such as University or company campuses, parks, vehicles in city streets, conferences, etc. An alternative to using real traces to extract the IM is to use a technique like the one introduced in [11]. In this work, the authors propose a way to generate non-homogeneous spatial node distributions. Using such a technique, one could generate an initial non-homogeneous placement of nodes, which then could be used to generate the IM to feed our model by averaging the number of nodes in each cell. The fact that this technique does not allow averaging density over time does not pose as a big drawback: as we will show in Section 4.7, spatial node density distribution does not vary much in time.

Our model not only manages to preserve a desired level of heterogeneity over time [25], but allows for tuning the desired spatial density for evaluating network protocols under different levels of node density, according to the researchers needs. In case the objective is to simulate

some specific real scenario, information on trace sources and repositories can be found in [5] and can be used to generate the IM as described here.

4.6 Evaluation Methodology

In this section, we describe how we investigate IWP's behavior and evaluate its performance. The baselines used for our comparative study are the random waypoint mobility model, or RWP, and the natural mobility model. We then start by describing these two mobility models.

4.6.1 RWP Mobility

In the RWP model, mobile nodes are initially placed in the area being simulated according to a given distribution. Typically, a uniform distribution is used. Each node remains in its position for a given period of time, called *pause time* P, uniformly chosen in the interval $[0, P_{max}]$, where P_{max} is a parameter whose value needs to be pre-specified. After this period, the mobile node chooses a new destination (x_d, y_d) uniformly distributed in the simulation area, and a velocity, also uniformly distributed in the interval $[v_{min}, v_{max}]$, where both v_{min} and v_{max} are pre-specified parameters. Once the destination is reached, the node pauses again and chooses another destination and velocity, as described above.

4.6.2 Natural Mobility Model

We also compare IWP's behavior and performance against a model that follows the preferential attachment principle. As representative of this family of models we use the natural mobility model, or simply *Natural* [13].

As discussed in Section 2, Natural is based on attraction points, where the attractiveness of each point is proportional to its popularity (i.e., number of nodes at or going towards it) and inversely proportional to the distance to it. Thus, the probability $\Pi(a_i)$ that an node z_k chooses an attractor a_i among all possible attractors is proportional to the portion of the total attractiveness it carries: $\Pi(a_i) = \frac{\mathcal{A}_{a_i,z_k}}{\sum_j \mathcal{A}_{a_j,z_k}}$. The attractiveness of an attractor is then defined as:

$$\mathcal{A}_{a_i,z_k} = \frac{(1 + \sum_{z_j \in \mathbb{Z}, z_j \neq z_k} B(a_i, z_k))}{\sqrt{(X_{a_i} - X_{z_k})^2 + (Y_{a_i} - Y_{z_k})^2}} \tag{4.5}$$

where $B(a_i, z_k)$ is a Bernoulli variable, with $B = 1$ if the individual z_k is going toward or staying at attractor a_i and 0 otherwise, and X and Y are the coordinates of a node and an attractor. In our implementation, we divided the simulation area in equally sized squares, or cells, and consider each cell to be an attraction point. The coordinates (X_{a_i}, Y_{a_i}) mark the center of the i-th attraction point. Once the new destination is known, the node travels towards it with a velocity uniformly distributed in the interval $[v_{min}, v_{max}]$. A pause time is randomly selected once arriving at the destination before choosing another destination and beginning the process again.

4.6.3 Simulation Setup

We ran two types of simulations: first, we used a modified version of the Scengen [62] scenario simulator to generate traces according to RWP–, Natural–, and IWP mobility. With those traces we generate the node density distribution results presented in Section 4.7.2. In the second type of simulation, once the synthetic traces were generated, these and the GPS traces were fed to the Qualnet network simulator [58] in order to evaluate their impact on core network functions, like routing.

In the first type of experiment, three sets of synthetic traces were generated using the RWP, Natural and IWP mobility models. The velocity range was set in a way that the average velocities would match the ones measured in the GPS traces for Quinta, KAIST and Statefair. In order to address the decaying velocity problem reported in [69], we followed the recommendations mentioned in that work. To that end, the velocity range was set to be $\pm$ the standard deviation measured in the real traces around the measured average velocity. Thus, the velocities were chosen *uniformly* in a range in which the lower limit was greater than zero and where the mean matches the one measured in the real traces. This is not an optimum solution and it is the simplest solution mentioned in [69], but since our focus is not on evaluating the routing protocol itself and we rather want to compare the mobility models, we found that solution suitable for these purposes.

Pause time was chosen uniformly in the range $[0, P_{max}]$, where the value of P_{max} was set to an appropriate value, in a way 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. Moreover, in all simulation scenarios, Quinta, KAIST and Statefair, we used the same initial positions found in their respective real traces for the same number of users. For further discussions on the actual distributions for these traces' mobility parameters, please refer to [6, 56].

In the RWP simulations, the next destination (x_d, y_d) was chosen uniformly over the simulated area. For the IWP simulations, the choice of (x_d, y_d) was given by Equation 4.3, where the intensity values μ were measured from the real traces as described in Section 4.5. For Natural, the probability of choosing a given cell is computed "on-the-fly", based on the cell's attractiveness.

We of course do not attempt to simulate the real scenarios exactly with the RWP as it is. It was an attempt to approach the simulated mobility patterns to the real scenarios so that we could compare the simulation results driven by randomly generated mobility traces against simulations over a real trace. The goal is to evaluate how far apart the results using the RWP are form the ones using the real trace and how close we can get by using the proposed model. Table 4.2 summarizes the simulation parameters.

Parameter	**Quinta**	**KAIST**	**Statefair**
Avg. Speed ($\pm\sigma$)(m/s)	1.27($\pm$0.16)	0.72($\pm$0.68)	0.58($\pm$0.12)
Avg. Pause (sec)	3.6	86	72
Area (meters x meters)	840x840	5000x5000	1260x1260
Duration (sec)	900	5000	8000
# nodes	98	78	19
# CBR flows	20	20	10

Table 4.2: Simulation parameters.

The traffic scenarios used in our second type of simulations are composed by *20 CBR flows* with randomly chosen source and destination pairs, for the Quinta and KAIST scenarios. *10 CBR flows* were used in the Statefair simulations. Flows start at randomly chosen times and stay active during the course of the whole simulation, at a *rate of 4 packets per second*. The routing protocol used to compare the network metrics between the different mobility patterns is the Ad-Hoc on-demand distance vector (AODV) protocol. The *radio range* was set to 150m, for 802.11b data rate of 11.0 Mbps.

Results are reported here with a *confidence interval* of 90% and reflect *10 runs* of the simulations using the real GPS traces. Since we could not change the mobility for the

simulations with the GPS traces, we varied the traffic scenarios by changing the source and destination pairs of the flows in each of the 10 runs. The same traffic scenarios were used to feed the RWP, Natural, and IWP 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 one of the synthetic models.

4.7 Results

In this section we define our metrics and present our results for node spatial density and network routing.

4.7.1 Spatial Node Density

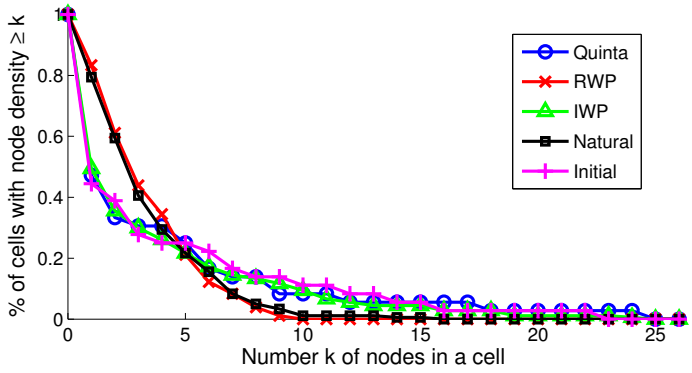
In order to study spatial node density behavior, we define the following metric:

- *Node density distribution:* the percentage of cells containing $\geq k$ nodes.

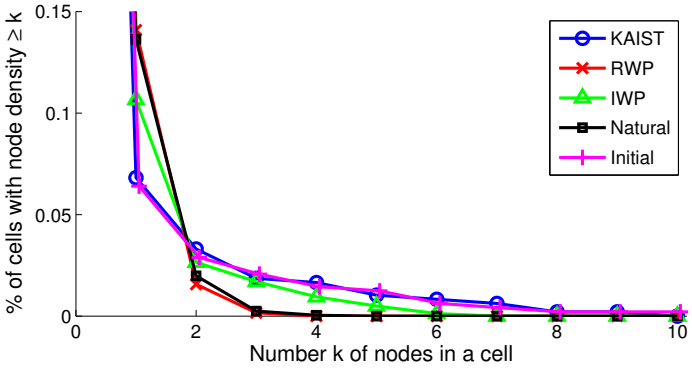
Each graph in Figure 4.3 shows the density distribution for one of the traces and each mobility model, namely IWP, RWP, and Natural. The curve labeled "*Initial*" shows the trace's density distribution at the beginning of the trace collection interval, while the curve labeled with the trace's name shows the distribution at the end of the trace collection interval, namely 900, 5000, and 8000 seconds for the Quinta, KAIST, and Statefair traces, respectively.

The data points shown for the RWP, Natural, and IWP models were averaged over 10 runs. The plots for the $KAIST$ and $Statefair$ traces are zoomed in to the region of interest. In those two plots, the only point not shown is $k = 0$, where the percentage of cells containing 0 or more nodes $P[k \geq 0]$ is the same for every curve and it is, of course, equal to 100%.

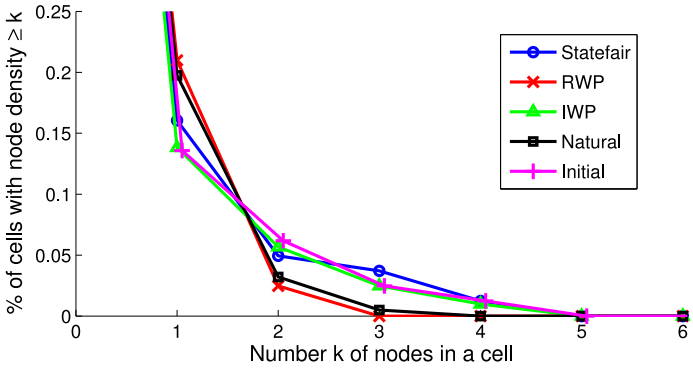
From these plots we observe that IWP's density distribution follows closely the distribution of the real trace, for all three traces. In the case of RWP, the majority of cells present similar number of nodes, and no cells contain significantly greater concentration of nodes. This is also the case for Natural. In the beginning of the simulation the initial distribution weighs on the computation of the attractiveness of the cells, but as the simulation goes on, more and more nodes may choose with small, but greater than zero probability, cells where there are no other nodes present. Over time, this behavior spreads the nodes around different cells, leveling the attractiveness of cells in general, approaching uniformity and RWP's density distribution.



(a) Quinta



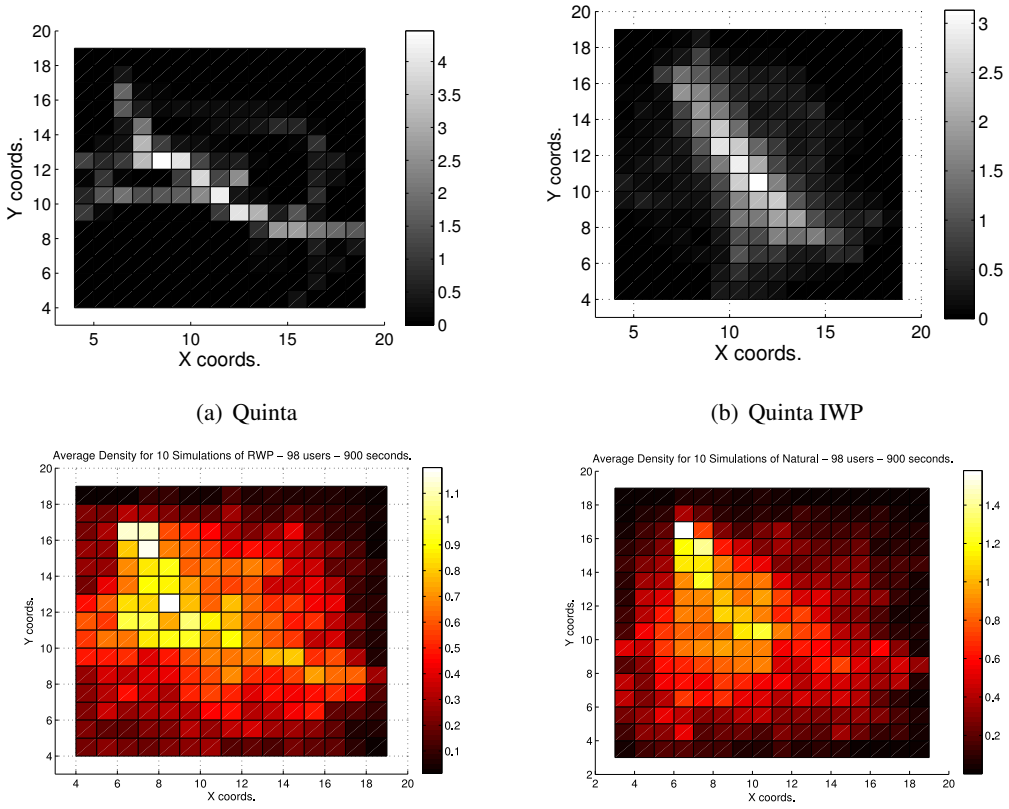
(b) KAIST



(c) Statefair

Figure 4.3: Node Density Distribution.

An important result here is to see how in real traces the spatial density distribution does not change much over time preserving the original non-homogeneous density distribution. Moreover, using IWP, we are able to preserve the initial spatial distribution and its non-homogeneity and thus yield very similar density distribution behavior when compared to the real trace. That holds true for all the three trace scenarios studied.



(a) Quinta

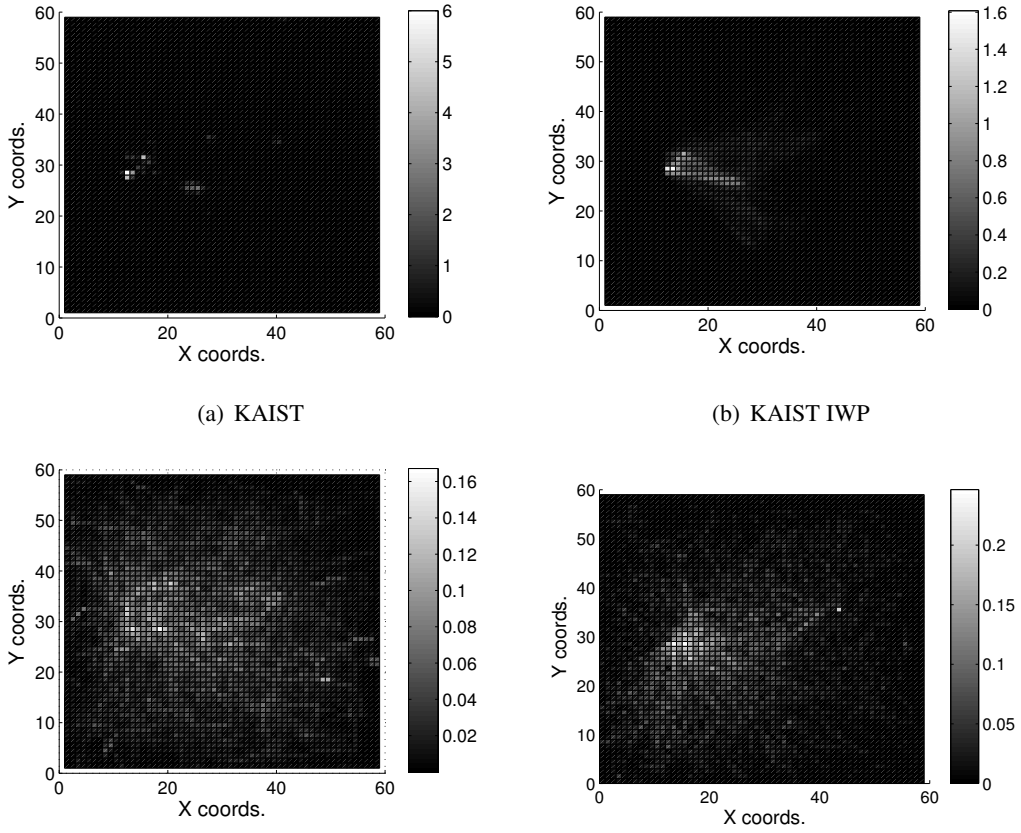
(b) Quinta IWP

(c) Quinta RWP

(d) Quinta Natural

Figure 4.4: Density matrices for the Quinta scenario.

When looking at the plots of density over time for the IWP model in Figures 4.4(b), 4.5(b) and 4.6(b) it is possible to notice how nodes appear to form clusters in the same places for the Quinta, KAIST and Statefair sites in Figures 4.4(a), 4.5(a) and 4.6(a), respectively. However, it is also possible to notice in the IWP plots that, over time, these clusters are not as concentrated as in the ones found in the real traces. The reason is that in the real trace, when going from one point to the other, nodes take different paths, sometimes avoiding obstacles,



(a) KAIST

(b) KAIST IWP

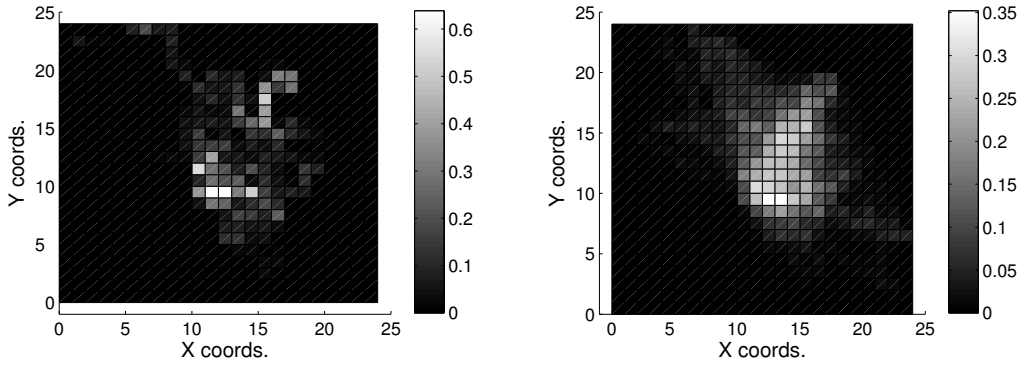
(c) KAIST RWP

(d) KAIST Natural

Figure 4.5: Density matrices for the KAIST scenario.

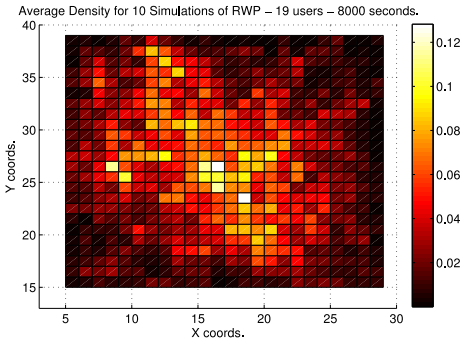
while in the IWP they still choose a path which is a straight line between origin and destination. Thus, despite of the similar concentrations in the same cells, over time this density is diluted because nodes, during their paths to the destination, may enter cells that were never, or very few times visited in the real trace.

In the same manner, when analyzing the plots of density over time for the RWP model in Figures 4.4(c), 4.5(c) and 4.6(c) it is possible to notice how nodes appear to form clusters in the same places for the Quinta, KAIST and Statefair sites in Figures 4.4(d), 4.5(d) and 4.6(d), respectively, we can see how close they are to each other. This illustrates the deficiency of preferential attachment based approaches and how close the density behavior of these approaches can get to total randomness.

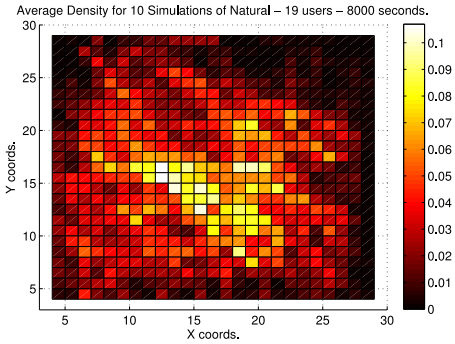


(a) Statefair

(b) Statefair IWP



(c) Statefair RWP



(d) Statefair Natural

Figure 4.6: Density matrices for the Statefair scenario.

4.7.2 Network Routing

The metrics used in our comparative study of network routing performance under the IWP mobility regime are:

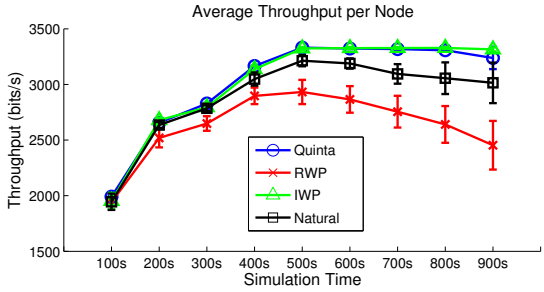
- *Throughput*: Total number of bits received at the destination divided by the time elapsed between the reception of the first data packet and the reception of the last data packet. This quantity is then averaged for all nodes.
- *End-to-End Delay*: Time elapsed between data transmission and reception then averaged for all nodes.
- *Delivery Ratio*: Ratio between the total number of packets received by all nodes and the total number of packets sent by all nodes.

- *Routing Overhead*: Total number of routing control packets transmitted by all nodes averaged over nodes.

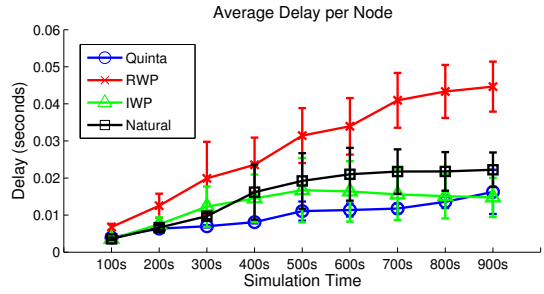
Figures 4.7(a), 4.7(b), 4.7(c), and 4.7(d) show, respectively, throughput, delay, delivery ratio, and routing overhead over time for the Quinta scenario. Notice the clear discrepancy between the real trace results and the RWP model. Also notice how the discrepancy widens over time. Routing performance suffers under RWP mobility once it spreads the nodes randomly over the simulation area. IWP, on the other hand, like the real trace, allows the formation and preservation of clusters, which, in the case of this scenario, ensures close to 100% delivery ratio.

In the case of Natural mobility model, specifically for the Quinta scenario, we notice that routing performance for this model stays close to the real trace curves up until around 300 seconds. Before that time, the probabilities of choosing each cell are differentiated by the initial non-uniform node positions, and the model is capable of maintaining some spatial node aggregation. However, after this point, nodes that were starting to roam away choose new cells where node concentration is not as high. As more and more nodes start to spread throughout the simulation area, the probability of choosing a new cell starts approaching a uniform distribution. This behavior causes the clusters to dissipate and the routing performance starts to diverge from the real traces.

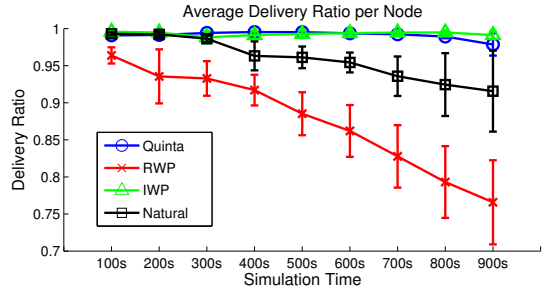
Figures 4.8 and 4.9 show routing performance metrics for the KAIST and Statefair scenarios, respectively. Once again, as in the Quinta scenario, we can see the IWP curves diverging from the RWP curves and approaching the real trace curves. In all the scenarios the routing overhead for RWP and Natural is much greater than for IWP and the real traces. This is mainly due to paths breaking caused by the random movement of RWP nodes and the increasing randomness towards the end of the simulations for Natural. Also, delivery ratio for these two scenarios is lower than for the Quinta scenario given the lower relative density caused by a larger simulation area and increased average distance between nodes. Nonetheless, the negative impact of a more sparse network in the case of KAIST and Statefair scenarios is greater for RWP and Natural. For these two scenarios, we were not able to see similar behavior to the real traces in the beginning of the simulations for Natural, as we saw in the Quinta scenario, due to the larger and lower resolution time scale. Nonetheless, the way Natural diverges from the real traces towards the end of the simulations remains.



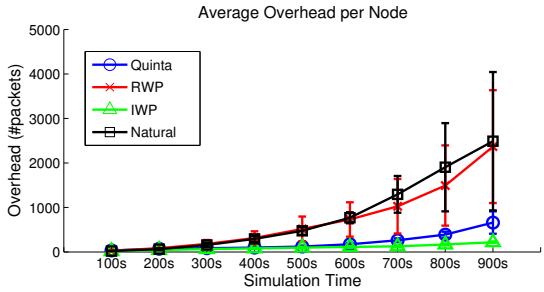
(a) Throughput



(b) Delay

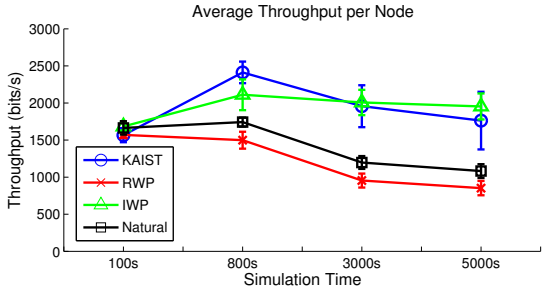


(c) Delivery Ratio

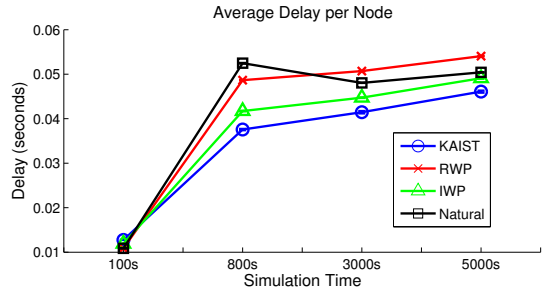


(d) Routing Overhead

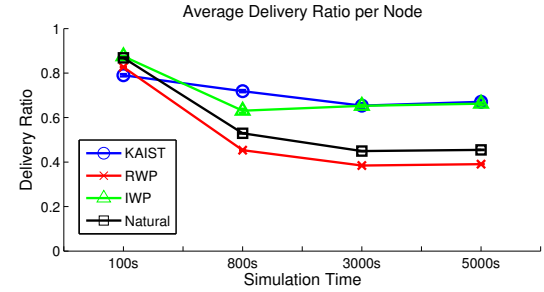
Figure 4.7: Network routing performance for the Quinta trace.



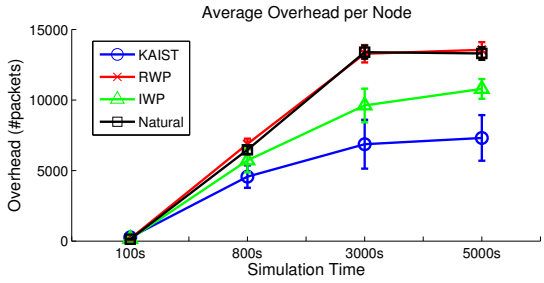
(a) Throughput



(b) Delay

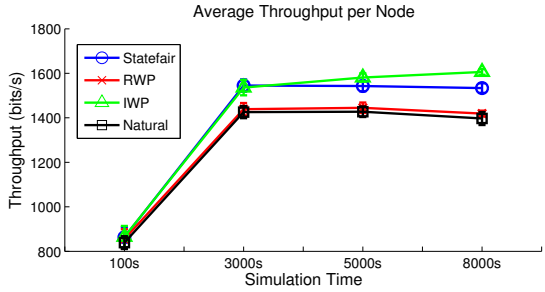


(c) Delivery Ratio

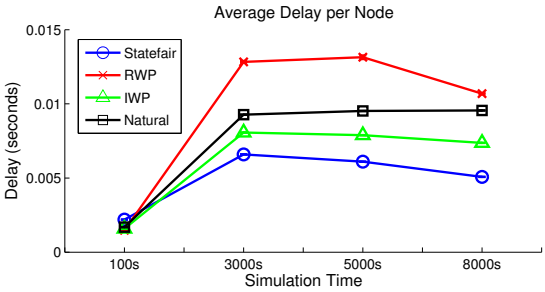


(d) Routing Overhead

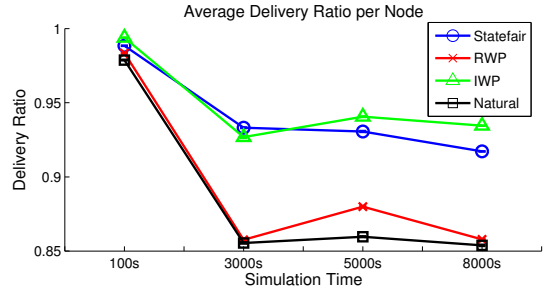
Figure 4.8: Network routing performance for KAIST trace.



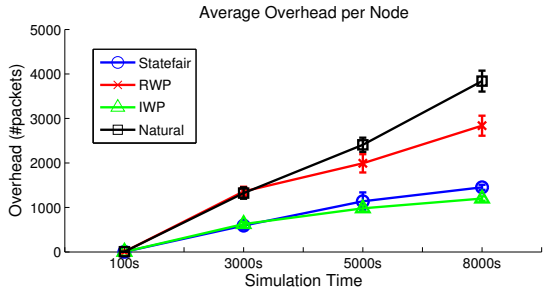
(a) Throughput



(b) Delay



(c) Delivery Ratio



(d) Routing Overhead

Figure 4.9: Network routing performance for Statefair trace.

Approaches like [36, 37, 13, 47, 48, 41], where the probability of choosing the new destination changes over time depending on nodes' positions, can present long transient phases and a stead state similar to a completely random behavior such as observed when applying the RWP model. We argue here that being able to set the spatial distribution in the begining of the simulation and being able to control it is important, and the proposed approach allows it.

4.8 Conclusions

We investigated in this section the spatial density distribution properties of real mobility using mobility traces collected in a diverse set of scenarios. More specifically, we chose three scenarios where human mobility was registered using GPS tracking devices carried by users as they move around a city park, a University campus, and a State Fair event. By defining a metric called spatial density distribution, we showed that, for all three scenarios, the original non-homogeneous node density distribution does not change over time. We also show that mobility models that rely on computing the probability of choosing a node's new destination dependent on the position of the other nodes can exhibit long transient phases and a steady state similar to completely random behavior as observed when applying the RWP model. To the best of our knowledge, this is the first time that these observations have been reported.

Acknowledging the importance of preserving and being able to control node density over time, we proposed a mobility model called Intensity Waypoint (IWP) that takes a density map as a parameter and where the mobility description is as simple as in the RWP model. By feeding the IWP model with a density map extracted from mobility traces, we were able to match closely the density distribution generated by the IWP to the one generated by the real traces, also preserving the original node density distribution. Moreover, when simulating the operation of a network moving according to the proposed model, we found that performance results expressed by a number of network metrics also match closely results obtained under mobility governed by the real traces. We also compared our results to models whose steady-state tend to random mobility and show that network performance under such regimes deviates from performance under real trace mobility.

Other features of our model are its generality as well as the user's ability to control node density distribution in order to simulate specific scenarios. Our conjecture is that, given the simplicity and ease of use of our approach, it has the potential of becoming adopted and

make its way to be part of well-known network simulation platforms.

As directions of future work, we plan to investigate human mobility behavior over extended periods of time as well as proving mathematically what we observed empirically, i.e., the steady-state behavior of preferential attachment based mobility models as well as IWP's.

Chapter 5

Modeling Spatial Node Density in Waypoint Mobility

This chapter introduces a modeling framework to analyze spatial node density in mobile networks under ”waypoint”-like mobility regimes. The proposed framework is based on a set of first order ordinary differential equations (ODEs) that take as parameters (1) the probability of going from one subregion of the mobility domain to another and (2) the rate at which a node decides to leave a given subregion.

We validate our model by using it to describe the steady-state behavior of real user mobility recorded by GPS traces in different scenarios. To the best of our knowledge, this is the first node density modeling framework generic enough that can be applied to any “waypoint”-based mobility regime. As an example, we use our framework to model the well-known Random Waypoint (RWP) mobility [15]. Our model confirms the well-known result showing that node density’s steady-state behavior under RWP mobility tends to homogeneity. We also use our framework to model mobility regimes based on the *preferential attachment* principle [3]. We show through the application of our proposed model that using preferential attachment to model human mobility leads to undesirable steady-state behavior, where the original spatial node density distribution is not preserved and exhibits behavior similar to random mobility a la RWP regime. Finally, as an example application of our framework, we discuss using it to generate steady-state node density distributions to prime mobile network simulations.

5.1 Introduction

When designing and evaluating wireless networks and their protocols, user mobility is a critical consideration. So much so that user mobility has inspired an extensive body of work both in infrastructure-based networks (e.g., wireless LANs or WLANs), as well as in infrastructure-less networks, a.k.a., wireless, self-organizing networks (WSONs). The latter include wireless mobile ad-hoc networks (MANETs), wireless sensor networks (WSNs), and disruption-tolerant networks (DTNs). Unlike their infrastructure-based counterparts where only end user nodes are mobile, in infrastructure-less networks, every node may move and thus mobility plays a considerable role in the performance of the network.

One way to characterize and describe mobility is through the *spatial density* of mobile nodes. Spatial node density can be defined as the number of nodes located in a given unit area and has significant impact on fundamental network properties, such as connectivity and capacity, as well as on core network functions, e.g., medium access and routing. To date, only a few efforts have focused on modeling spatial density. Notable examples include [25, 9, 50]. However, most previous work have been focusing exclusively on the Random Waypoint (RWP) model [15].

In this work, we introduce an Ordinary Differential Equation (ODE) framework to mathematically model spatial node density under different "waypoint"-based mobility regimes. More specifically, our model describes node density's steady-state behavior under waypoint-based mobility which is characterized by having nodes probabilistically choose the next destination, or waypoint', based on some probability density function. We contend that waypoint mobility is one way to describe forms of human mobility. Therefore, we validate our model by using it to describe the steady-state behavior of real user mobility recorded by GPS traces in different scenarios and compare the results against the corresponding traces. Moreover, we present comparative results for steady-state spatial distribution analysis of a number of synthetic waypoint mobility regimes. To the best of our knowledge, this is the first node density modeling framework generic enough that it can be applied to any waypoint-based mobility regime. As an example, we use our framework to model the well-known RWP mobility regime. Our model confirms the well-known result showing that node density's steady-state behavior under RWP mobility tends to homogeneity, as defined in [11][1].

[1]The use of the term "homogeneous node distribution" refers here to the fact that there is no significant concen-

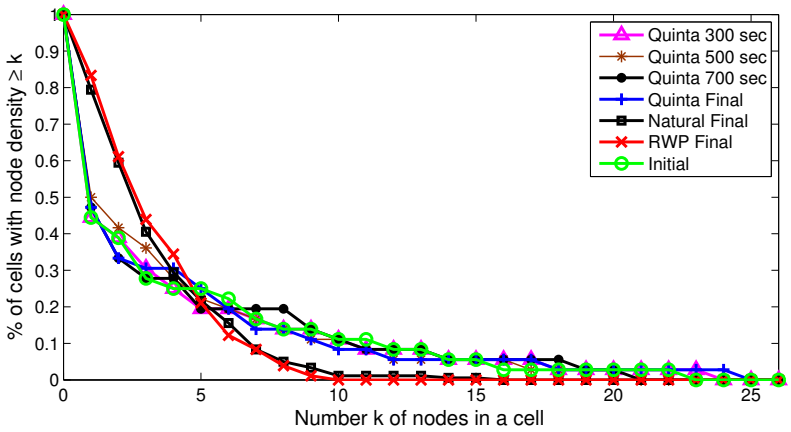


Figure 5.1: Node spatial density distribution at different trace collection times for mobility in a city park.

Furthermore, we also use our framework to model mobility regimes based on the *preferential attachment* principle [3]. We show through the application of our proposed model that using preferential attachment to model human mobility leads to undesirable steady-state behavior. More specifically, our model shows that, at steady state, the original spatial node density distribution is not preserved and exhibits behavior similar to random mobility a la Random Waypoint regime. This behavior has been observed empirically in [52]. Instead, real human mobility exhibits "persistent" density heterogeneity as illustrated in Figure 5.1. This figure shows the spatial density distribution for one of the traces used by us which was collected in the Quinta da Boa Vista Park in Rio de Janeiro, Brazil.

This distribution is defined as the percentage of cells containing $\geq k$ nodes. The first 4 curves in the plot refer to the distributions at instants 300, 500, 700, and final (900 seconds), which is the end of the trace collection interval. The two other curves correspond to the node density distributions measured after 900 seconds of simulations of two synthetic waypoint-like mobility regimes, namely RWP and Natural [13]. Each of these curves reflect the final node distribution averaged over 10 runs of simulations. Both mobility models and the experiments that generated these curves are discussed in detail in Section 5.5. The last curve shown in the graph is the initial node distribution measured from the Quinta trace and also used as the initial distribution in all simulation runs of both synthetic mobility regimes studied.

From Figure 5.1, we observe that the density distribution of the real trace does not

tration of nodes (clusters), and should not be mistaken with uniform distribution normally used to model the choice of next destination, speed and pause time in random mobility models.

vary much with time: the largest deviation from the initial distribution for any value of k at any instant is 8.3%; the average deviation from the initial distribution measured in all the instants for all values of k is 1.27%. Similar observations can be drawn from the other traces used in our work as reported in Section 5.4.1. Moreover, we also observe a very different behavior when applying either one of the synthetic regimes to the same scenario. They deviate significantly from the initial conditions.

However, the most interesting observation here is the fact that using a preferential attachment based regime such as Natural, does not preserve the original clustering of the nodes. In fact, spatial density resulting from preferential-attachment based waypoint mobility "deteriorates", at steady state, to behavior similar to random mobility. In Section 5.5, we present more details on these results and apply our proposed model to study the steady-state behavior of spatial density of these mobility regimes.

Overall, the contributions of our work are many-fold: (1) we introduce the first spatial node density modeling framework for waypoint mobility regimes, (2) we validate our framework by applying it to real user mobility traces as well as synthetic waypoint mobility regimes, (3) we use our model to show analytically that the steady-state behavior of node density under preferential-attachment based mobility does not preserve node density's original distribution and exhibits behavior similar to random mobility, and (4) as an example application of our framework, we discuss using it to generate steady-state node density distributions to prime mobile network simulations.

The remainder of this chapter is organized as follows. Section 5.2 places our work in perspective by presenting related work in mobility modeling and characterization. Our ODE model is presented in detail in Section 6.3, how its parameters are set, and our implementation. Section 5.4 show the validation of our proposed framework towards modeling real human mobility, while Section 5.5 presents the applications of our work on modeling spatial density of synthetic waypoint mobility regimes. Finally, Section 6.7 concludes this topic.

5.2 Related Work

Mobility models are vital in the design, testing, and evaluation of wireless networks and their protocols. As an indication of the importance of mobility models to the study of wireless network protocols, most well-known network simulators include "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 [15]. A notable example is the Random-Waypoint Mobility (RWP) regime.

Recently, however, network researchers and practitioners have been trying to use more realistic scenarios to drive the evaluation of wireless network protocols. This motivated initiatives such as the CRAWDAD [18] trace repository, which makes real traces available to the networking community. These traces can then be used to run trace-driven simulations. Even though initiatives like CRAWDAD have greatly increased availability of real traces, relying exclusively on traces to design and evaluate network protocols would not allow a broad enough exploration of the design space.

To address this problem, a number of efforts have proposed mobility models based on realistic mobility patterns [49]. Notable examples include [7, 8, 60, 31]. 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 [3] proposes a model that generates scale-free networks, i.e., networks whose node degrees follow a power law distribution, and demonstrate that many real large networks are scale free, that is, the node degree in the network graph follows a power law. The authors discuss the mechanism responsible for the emergence of scale-free networks and argue that understanding this problem will require a shift from modeling network topology to modeling "network assembly and evolution". To this end, they define the Barabási-Albert model based on *growth* and *preferential attachment*. Growth refers to the fact that the number of nodes in the network increases over time, where a new node is placed with m edges connecting it to other m nodes. Preferential attachment means that a node will choose to connect to another node i with probability $\Pi(k_i) = \frac{k_i}{\sum_j k_j}$ based on the degree k_i of node i and any node j connected to node i.

The work proposed in [41] is another example of a model that follows Barabási-Albert's growth and preferential attachment principles. The authors even show a figure where they present their initial (after growth) and steady-state spatial distribution. It is possible to see how clusters dissipate and fade away over time. The same concept is also used in [47] where nodes are also driven by identified social interactions. Authors validate their approach by means of showing the power-law exponential decay of inter-contact time among the formed

communities and compare it with measurements in real traces.

We apply our proposed model in the evaluation of an example of a mobility regime that uses preferential attachment and use it to demonstrate that approaches that take advantage of this principle lead to a node spatial density in stationary regime similar to a total random approach, like the Random Waypoint.

In this work, the main focus is over spatial node density. Most previous work on modeling node spatial density have been focusing on the RWP model, rather then trying to characterize it in real applications. In [9], for example, analytical expressions are derived for the spatial density distribution that results from using the RWP model on simulations. One-dimensional case is analyzed and an approximation for the two-dimensional case is also given. They also analyze the concept of attraction areas in a modified version of the RWP regime. One other example of analytical work towards modeling steady-state behavior of the RWP can be found in [50]. In that work, authors derive stationary analytical expressions for node density and node velocity.

The proposed first order ODE framework described here differs from previous work on modeling spatial distribution, first in its nature, the fact that it uses ODEs as a tool for studying density distributions (never used before). Secondly, also by the fact that our approach is the first that is generic enough that can be applied in any mobility regime that bases its behavior on a waypoint-like movement. By waypoint-like movement we mean, "pause - choose the next destination following some given probability distribution - move to next destination in a straight line and constant speed - repeat". It is yet, capable of faithfully replicate the stead-state behavior of spatial distributions in real scenarios, as we show through our results, validating our model against real live traces.

ODEs are a powerful tool that have been used on modeling in many areas in networking. To mention just a couple, in [71] authors use a ODE framework to derive a set of equations that model the performance of epidemic routing in a DTN environment, and in [20] an ODE model for performance analysis of a self-limiting epidemic forwarding mechanism is proposed.

Moreover, in [23] Partial Differential Equations (PDEs) are used to model spatial node density of the RWP and Random Direction mobility regimes. They present an analysis of the transient behavior of the spatial density under these two mobility regimes. While this work focus on the RWP and Random Direction mobility, our approach is generic enough that

can be used to study any waypoint-based mobility regime (including random approaches and preferential-attachment based ones). The work described in [29], proposes a Markovian based mobility model with the purpose of forming and dissolving clusters of nodes. They analyze analytically the node spatial distribution, presenting results, specifically for their mobility model.

5.3 Proposed Model and Framework

Our objective is to model the spatial node density of a mobile network. We assume a waypoint-based mobility pattern, where nodes stay in a given location i for a given period of time and choose to leave i towards another location j with probability p_{ij}. Once the node arrives at j, the process restarts.

5.3.1 ODE Framework

Assume a mobile network composed of m mobile nodes, where all nodes are capable of moving around inside a delimited area a. Now assume this area is divided into equally sized square subareas of size $l \times l$, defined here as *cells*. The mobile nodes can then choose to move from cell to cell with a given probability. Let $X(t)$ be the stochastic process that determines which cell a mobile node chooses at time t. We can write then $p_{ij} = P\{X(t) = i \mid X(t+\gamma) = j\}$, as the *transition probability*, which is the probability that a node in cell i, at time t, is going to choose to go to cell j at time $(t + \gamma)$, after some time step γ.

Thus, we are interested in the average number of nodes in each cell i, represented by the component $N_i(t)\ \forall i \in \{1, ..., n\}$ of the state vector $N(t) \in \mathbb{R}^{n\times 1}$, where n is the total number of cells for the desired scenario.

The variation in the number of nodes at each cell $\dot{N}_i(t) = \frac{dN_i(t)}{dt}$ is simply the difference between nodes arriving in cell i and the ones departing from the same cell at time t, as expressed in Equation 5.1.

$$\dot{N}_i(t) = \underbrace{\lambda_0 + \sum_j p_{ji}\mu_j N_j(t)}_{\text{Arriving at cell i}} - \underbrace{\Big(\sum_j p_{ij}\mu_i N_i(t) + \mu_0 N_i(t)\Big)}_{\text{Departing from cell i}}, \tag{5.1}$$

where λ_0 is the rate at which new nodes arrive in cell i from outside the system and μ_0 the rate at which nodes decide to get disconnected and leave the system, given that they are at

cell i. Also, μ_i is the rate at which nodes decide to leave cell i towards another cell, which allow us to write $\mu_{ij} = p_{ij}\mu_i$ as the rate at which nodes in cell i decide to leave this cell towards cell j. We can also define the arrival rate in cell i as the sum of the departing rates of all nodes going from cell j to cell i, over all possible values of j, including $j = i$, since we allow transitions from a cell to another position in itself. The arrival rate is given by Equation 5.2.

$$\lambda_i = \sum_j p_{ji}\mu_j. \tag{5.2}$$

5.3.2 Parameters Choice, Discussion and Simplifications

In reality we observe that nodes prefer some cells over others and some transitions over others. The probability of choosing a destination and the rate at which nodes depart from that destination depends on how popular that destination is and what are the nodes' interests in each destination. For example, nodes moving around on a campus environment may go very often from the cafeteria to the classroom, but not so often from the cafeteria to the library. This means that $p_{\text{cafeteria,classroom}} > p_{\text{cafeteria,library}}$. Moreover, since people might tend to stay inside the library for longer than in the cafeteria, the relationship between the departure rate from this two locations might be such as $\mu_{\text{cafeteria}} > \mu_{\text{library}}$.

In order to simplify our model, more specifically the choice of the parameters (departure rates and transition probabilities), we define the rate μ_i as the inverse of the average time spent by the nodes in cell i. We also considered the transition probabilities independent of where the transition originated. This means that the probability of going from cell j to cell i is the same probability of simply choosing cell i as the next destination for all j. We then make $p_{ji} = P\{X(t) = j \mid X(t+\gamma) = i\} = P\{X(t+\gamma) = i\} = p_i$.

Moreover, in order to validate our model we have chosen to extract the model parameters from— and compare our results with— real live GPS traces, where the number of nodes in the system remains constant during the whole duration of the trace. For that reason, in the results we present in Section 5.4.3 we used a slightly simplified version of our model, where $\lambda_0 = \mu_0 = 0$. Equation 5.3 gives this version of our ODE model.

$$\dot{N}_i(t) = \underbrace{\sum_j p_i \mu_j N_j(t)}_{\text{Arriving at cell i}} - \underbrace{\sum_j p_j \mu_i N_i(t)}_{\text{Departing from cell i}}, \tag{5.3}$$

5.3.3 Implementation

In this section we present a vectorized version of Equation 5.3, so that we could implement it on MATLAB [45]. We used a 4th order Runge-Kutta ODE solver, native to the platform, to do so.

We start by defining a matrix $A \in \mathbb{R}^{n \times n}$ as a parameter matrix given by $A = P \times M$. $P \in \mathbb{R}^{n \times 1}$ is a column vector containing in every ith position the probability p_i of a node choosing cell i as the next destination, and $M \in \mathbb{R}^{1 \times n}$ a row vector containing in every ith position the rate μ_i at which nodes choose to leave cell i. The components of matrix A, resulting from this multiplication are $a_{ij} = p_i \mu_j$.

Thus, it is possible to write Equation 5.3 for $\dot{N}(t) \in \mathbb{R}^{n \times 1}$ in its equivalent vectorized form as follows:

$$\dot{N}(t) = \underbrace{A \times N(t)}_{\text{Arriving}} - \underbrace{\Big((A^T \times \vec{\mathbb{1}}) \cdot N(t)\Big)}_{\text{Departing}}, \tag{5.4}$$

where A^T is the transpose of matrix A, that we multiply by $\vec{\mathbb{1}} \in \mathbb{R}^{n \times 1}$, a column vector of ones, to give us a resulting $n \times 1$ column vector in which every component i represents the summation of all the components of the ith row of matrix A^T. After that, we perform a component wise multiplication with the state vector $N(t)$, which gives us the number of nodes departing from a given cell. That represents the second summation in the right-handed side of Equation 5.3.

5.4 Spatial Node Density of Human Mobility

We validate our model using real mobility traces; in other words, we show how the model can be applied to describe the steady-state behavior of spatial node density associated with human mobility. Three real GPS traces were used in our validation. These traces were

collected in scenarios that are quite diverse, namely a city park, a university campus, and a state fair. We describe these traces in detail below as well as how we use information from the traces to estimate the parameters of our ODE framework.

5.4.1 Mobility Traces

Table 5.1 summarizes the GPS traces in terms of number of users, duration of the trace, and GPS sampling period.

Trace	# users	Duration	Samples
Quinta [6]	98	900s	1s
KAIST [56]	78	5000s	10s
Statefair [56]	19	8000s	10s

Table 5.1: Summary of the GPS traces studied.

Quinta, refers to the "Quinta da Boa Vista Park" trace, first presented in [6]. It is a GPS trace collected at a park in the city of Rio de Janeiro, Brazil. The park has many trees, lakes, caves, and trails. It houses the National Museum of Natural History and the city Zoo. The *KAIST* trace [56], on the other hand, is a GPS trace collected at the KAIST University campus in Daejeon, South Korea. The *Statefair* trace, also available at [56], is yet another mobility scenario showing daily GPS track logs collected from the NC State Fair held in North Carolina, USA.

5.4.2 Parameter Estimation

We extract from the traces the distributions of *speed*, *pause time*, and *node density*. We use the trace's sampling period, e.g., for example, in the *Quinta* trace, the sampling period is $T = 1$ seconds. Node speed is defined 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 used to account for GPS error. We set this threshold to be 2 meters for KAIST and Statefair traces and 0.5 meter for the Quinta trace, due to jitter in GPS update frequency.

To extract spatial node density, the area covered in the trace is divided into squared cells of 140 x 140 meters. The choice of cell size was based on empirical observations, i.e., we picked a cell size that provided both adequate resolution as well as clustering. An alternate approach could be based on identifying "attraction zones", as was done in [37]. This is one

of the topics of future work we plan to address. At the limit, i.e., where the cell is either infinitesimal (lower limit) or the size of the whole area (upper limit), all the traces and synthetic mobility regimes would have the same relative spatial density, namely one or zero nodes per cell for the lower limit and all the nodes in the same (unique) cell for the upper limit.

After dividing the area into cells, we took a snapshot of the number of nodes at every cell every T seconds. The value of $T = 10$ was used since, for the size of the cells and the speeds sampled from the traces, a node could not on average change between more than two cells during T. For every cell, at every interval T we counted the number of nodes in each cell. We then averaged the number of nodes in each cell over the course of the whole duration of the trace. The result is what we refer to as *Intensity Map (IM)* which we use to estimate the probability a node will choose a given cell as its next destination.

In the case of real mobility, e.g., as described by GPS traces, we set the *probabilities of choosing a given cell*, p_i of our ODE model to be the normalized value of the IM for cell i, such that $p_i = \frac{IM(i)}{\sum_j IM(j)}$, where IM($i$) is the intensity in cell i.

The rate μ_i, as mentioned before, is computed as the inverse of the average time spent by the nodes in cell i. This time has two components. The time spent by the node moving towards or from a given point in the cell, and the time spent in pause at this point, which reflect both main basic parameters of human mobility, speed and pause time. This two components were empirically measured from the GPS traces and used to compute μ_i.

5.4.3 Results

As highlighted in previous sections, the goal of our model is to describe the steady-state behavior of spatial node density in *waypoint*-like mobility regimes in which: (1) a node chooses its next destination following some given probability distribution, (2) moves to that destination, (3) pauses for some time, and (4) repeats from step (1).

Spatial node density is defined as the percentage of cells containing $\geq k$ nodes. It can also be expressed as the probability of finding a cell containing $\geq k$ nodes. It describes the degree of "clustering" exhibited by mobility regimes and can be used to evaluate how close to reality a given synthetic mobility regime is as far as its ability to mimic the degree of clustering exhibited by real mobility.

We followed the guidelines presented in Section 5.4.2 to estimate the parameters of

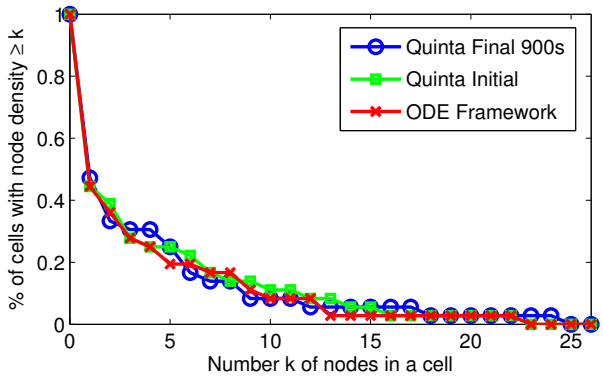


Figure 5.2: Initial and final spatial node density distribution for the Quinta trace, and the respective steady-state density distribution using the proposed ODE framework.

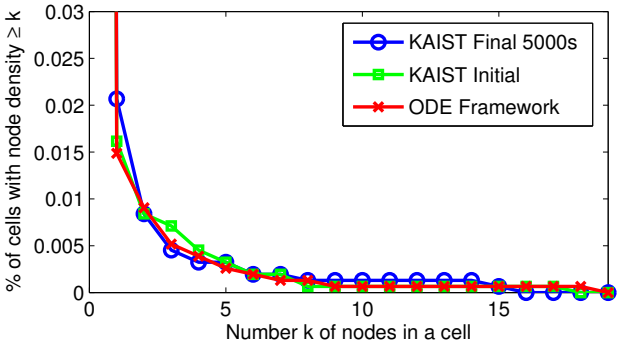


Figure 5.3: Initial and final spatial node density distribution for the KAIST trace, and the respective steady-state density distribution using the proposed ODE framework.

our model for each of the traces studied. Figures 5.2, 5.3 and 5.4 plot spatial node density in the Quinta, KAIST, and Statefair scenarios, respectively. Each figure shows three curves plotting the spatial density: (1) at the beginning of the trace, (2) at the end of the trace, i.e., at 900 seconds for the Quinta Trace, 500 seconds for KAIST, and 8000 seconds for the Statefair, and (3) by applying our ODE framework. Note that the plots for the $KAIST$ and $Statefair$ traces are zoomed into the region of interest. In those two plots, the only point not shown is $k = 0$, where the percentage of cells containing 0 or more nodes $P[k \geq 0]$ is the same for every curve and it is, of course, equal to 100%.

The *largest* deviation of our ODE model from the final density distribution measured from the traces, for any value of k at any instant is 5.36%, 0.58% and 7.52%; the average deviation from the initial distribution measured in all the instants for all values of k is 1.45%, 0.06% and 2.02% for the Quinta, KAIST and Statefair traces respectively.

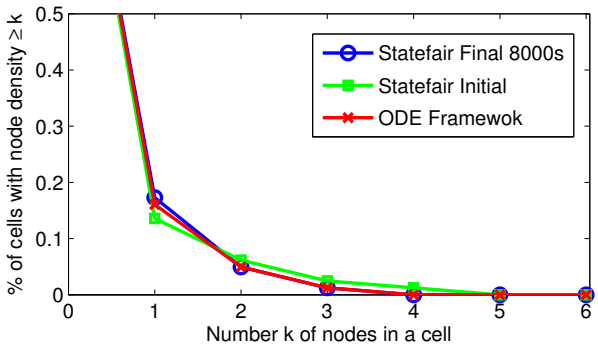


Figure 5.4: Initial and final spatial node density distribution for the Quinta trace, and the respective steady-state density distribution using the proposed ODE framework.

5.5 Node Density in Synthetic Waypoint Mobility

Here we show our framework's ability to closely describe the steady-state behavior of spatial node density resulting from synthetic *waypoint*-like mobility. We apply our model to two different regimes, namely Random Waypoint (RWP) [15] and the Natural [13] mobility. We start by briefly describing these two mobility regimes and then present the setup we used to generate mobility traces under them. Subsequently, we explain how we estimated the parameters for the model and present results comparing spatial node density distributions resulting from the synthetic mobility regimes and our model.

5.5.1 RWP Mobility

Random Waypoint (RWP) mobility, an example of waypoint-like mobility regime, has been widely used in the study of multi-hop ad-hoc wireless networks (MANETs). Under RWP mobility, mobile nodes are initially placed in the area being simulated according to a given distribution. Typically, a uniform distribution is used. Each node remains in its position for a given period of time, called *pause time* P uniformly chosen in the interval $[0, P_{max}]$, where P_{max} is a pre-specified parameter. After this period, the mobile node chooses a new destination uniformly distributed in the simulation area, and a velocity, also uniformly distributed in the interval $[v_{min}, v_{max}]$, where both v_{min} and v_{max} are pre-specified parameters. Once the destination is reached, the node pauses again and chooses another destination and velocity, as described above.

5.5.2 The Natural Mobility Regime

We also compare our results against mobility regimes that follow the *preferential attachment* principle. As representative of this family of mobility regimes, we use the Natural mobility model, or simply *Natural* [13].

As discussed in Section 5.2, *Natural* is based on attraction points, where the attractiveness of each point is proportional to the *attractor*'s popularity given by the number of nodes at or going towards it and inversely proportional to the distance to it. Thus, the probability $\Pi(a_i)$ that a node z_k chooses an attractor a_i among all possible attractors is proportional to the portion of the total attractiveness it carries: $\Pi(a_i) = \frac{\mathcal{A}_{a_i,z_k}}{\sum_j \mathcal{A}_{a_j,z_k}}$. The attractiveness of an attractor is then defined as:

$$\mathcal{A}_{a_i,z_k} = \frac{(1 + \sum_{z_j \in \mathbb{Z}, z_j \neq z_k} B(a_i, z_k))}{\sqrt{(X_{a_i} - X_{z_k})^2 + (Y_{a_i} - Y_{z_k})^2}} \tag{5.5}$$

where $B(a_i, z_k)$ is a Bernoulli variable, with $B = 1$ if the individual z_k is going toward or staying at attractor a_i and 0 otherwise, and X and Y are the coordinates of a node and an attractor. In our implementation, we divided the simulation area in equally sized squares, or cells, and consider each cell to be an attraction point. The coordinates (X_{a_i}, Y_{a_i}) mark the center of the i-th attraction point. Once the new destination is known, the node travels towards it with a velocity uniformly distributed in the interval $[v_{min}, v_{max}]$. A pause time is randomly selected once arriving at the destination before choosing another destination and beginning the process again.

5.5.3 Generating Synthetic Waypoint Mobility Traces

Using a modified version of the Scengen [62] scenario simulator we generated mobility traces according to the RWP and the *Natural* mobility regimes. We setup the simulations trying to mimic the three real scenarios described here, for Quinta, KAIST and Statefair. Three sets of synthetic traces were generated using the RWP and Natural mobility models. The velocity range was set in a way that the average velocities would match the ones measured in the GPS traces.

In order to address the decaying velocity problem reported in [69], we followed the recommendations mentioned in that work. The velocity range was thus set to be $\pm$ the standard

deviation measured in the real traces around the measured average velocity. Thus, the velocities were chosen *uniformly* in a range in which the lower limit was greater than zero and where the mean matches the one measured in the real traces. This is the simplest though not the optimal solution mentioned in [69]. However, since the focus of our work is not network performance evaluation itself, we found this solution to be suitable for our purposes.

Pause time was chosen uniformly in the range $[0, P_{max}]$, where the value of P_{max} was set to an appropriate value, in a way 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. Moreover, in all simulation scenarios, we used the same initial positions found in the respective real traces for the same number of users. For further discussions on the actual distributions for these traces' mobility parameters, please refer to [6, 56].

When applying the ODE framework to describe spatial density behavior in synthetic mobility, p_i follows the probability distribution particular to the specific mobility regime used. For example, in RWP mobility, p_i is the same for every value of i, since the probability of choosing the next waypoint follows a uniform distribution. For Natural, the probability of choosing a given cell is computed "on-the-fly", based on the cell's attractiveness, defined by Equation 6.4.

The rate μ_i is computed as the inverse of the average time spent by the nodes in cell i. This time has two components. The time spent by the node moving towards or from a given point in the cell, and the time spent in pause at this point. This two components were empirically measured from long simulations (10^5 seconds), using the same parameters for each scenario, and used to compute μ_i. A more generic approach to determine the value of this parameters for a given generic scenario is the subject of our future work.

Reported simulation results on density, comparing our ODE framework with simulated traces, reflect *10 runs* of the simulations using each mobility regime at each scenario. Table 5.2 summarizes the simulation parameters.

5.5.4 Results

Figures 5.5 and 5.6 show the results for spatial node density distribution at the Quinta scenario, for the RWP and Natural mobility regimes respectively. The plots show three curves

Parameter	Quinta	KAIST	Statefair
Avg. Speed ($\pm\sigma$)(m/s)	1.27($\pm$0.16)	0.72($\pm$0.68)	0.58($\pm$0.12)
Avg. Pause (sec)	4.61	86.00	7.20
Area (meters x meters)	840x840	5000x5000	1260x1260
Duration (sec)	900	5000	8000
# nodes	98	78	19

Table 5.2: Simulation parameters.

corresponding to: 1) the initial density distribution taken from the trace and used to feed all the simulations for both mobility models, 2) the steady-state density distribution using the proposed ODE framework applied to the RWP and Natural mobility regimes, and 3) the final density distribution measured and averaged at the end of the simulations for the synthetic mobility regimes.

The first obvious observation analyzing these plots is that the synthetic mobility regimes are unable to follow the long tail behavior of the density metric presented by the distribution measured in the real traces. Moreover, as we demonstrated in the previous section, the presented ODE model is able to follow this long tail characteristic when applying the parameters extracted from the real traces. In the case of the synthetic models, when applying the exit rate and probability of choosing the next cell, characteristic to each mobility regime, the proposed framework now behaves as the synthetic models do and is able to describe density distribution curves very similar to the RWP-like regimes studied.

Figures 5.7 and 5.8 show the results for spatial node density distributions at the KAIST scenario, for the RWP and Natural mobility regimes respectively. Similar results can be observed in Figures 5.9 and 5.10 for the Statefair scenario. Once again, the plots for the $KAIST$ and $Statefair$ traces are zoomed in to the region of interest.

These show not only the flexibility of the proposed ODE framework and that it is able to describe the steady-state behavior of the density distribution of real and synthetic (RWP-like) mobility, but it also show the accuracy of the proposed approach.

5.5.5 Application

In addition to its applications in the study of the steady-state behavior of spatial node density of waypoint-like mobility regimes, we find another very interesting application of our framework. Recall that the output of our model is a vector where each position contains the ex-

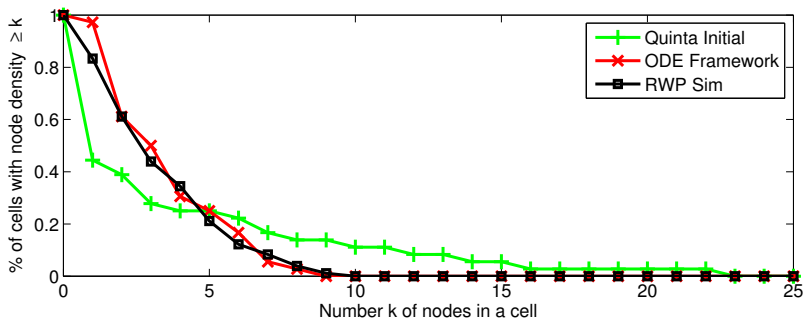


Figure 5.5: Initial spatial node density distribution for the Quinta trace, simulated final density distribution using the RWP mobility regime and the respective steady-state density distribution using the proposed ODE framework applied to the RWP mobility regime.

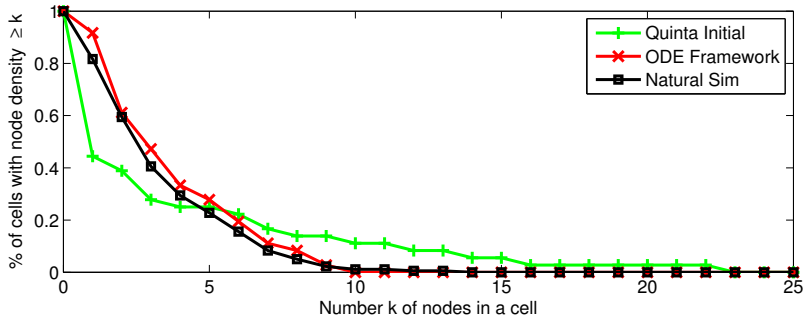


Figure 5.6: Initial spatial node density distribution for the Quinta trace, simulated final density distribution using the Natural mobility regime and the respective steady-state density distribution using the proposed ODE framework applied to the Natural mobility regime.

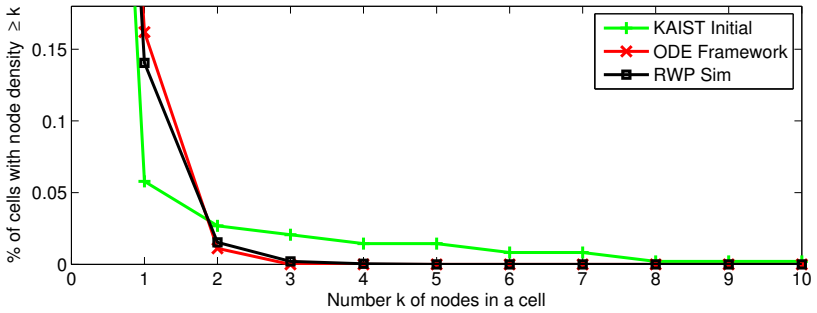


Figure 5.7: Initial spatial node density distribution for the KAIST trace, simulated final density distribution using the RWP mobility regime and the respective steady-state density distribution using the proposed ODE framework applied to the RWP mobility regime.

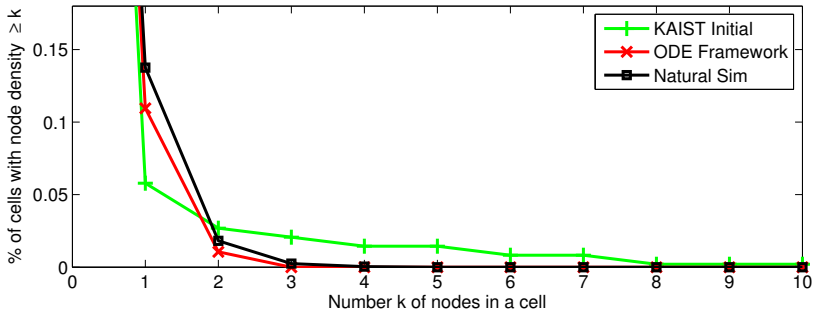


Figure 5.8: Initial spatial node density distribution for the KAIST trace, simulated final density distribution using the Natural mobility regime and the respective steady-state density distribution using the proposed ODE framework applied to the Natural mobility regime.

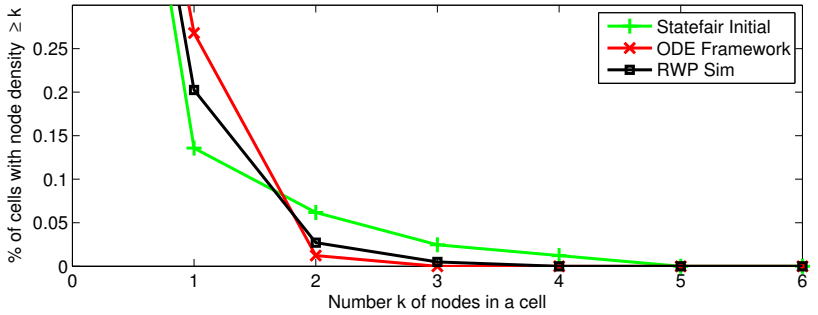


Figure 5.9: Initial spatial node density distribution for the Statefair trace, simulated final density distribution using the RWP mobility regime and the respective steady-state density distribution using the proposed ODE framework applied to the RWP mobility regime.

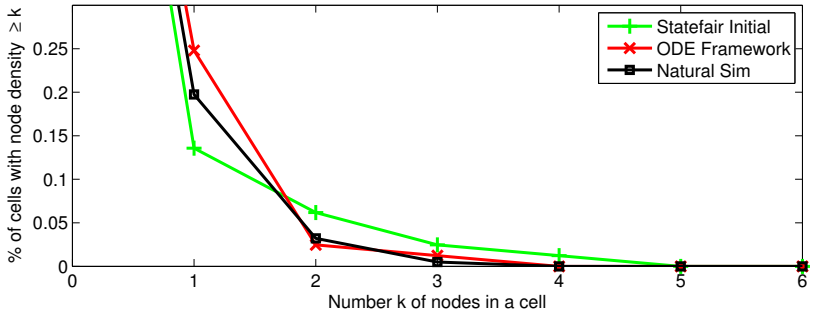


Figure 5.10: Initial spatial node density distribution for the Statefair trace, simulated final density distribution using the Natural mobility regime and the respective steady-state density distribution using the proposed ODE framework applied to the Natural mobility regime.

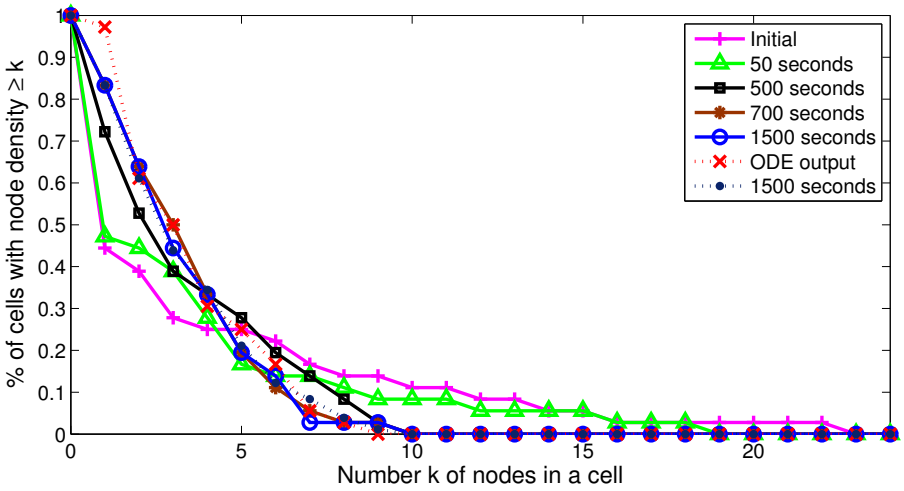


Figure 5.11: RWP simulation using initial density conditions from the trace (solid lines), and simulation using the output of the ODE framework as the initial density conditions (dashed lines).

pected number of nodes encountered in the cell corresponding to that position. This vector provides node density distribution at steady-state. Consequently, we can interpret the normalized value in each cell as the probability of placing a node in that cell. In other words, the normalized output vector can be seen as the steady-state spatial distribution for the waypoint-based mobility regime of interest. In this context, following the steady-state distribution given by our model, it is possible to use a *Fitness Proportionate Selection* scheme, such as the *Roulette-Wheel* described in [52] and commonly used in genetic algorithms [54] to perform initial node placement when setting up wireless network simulations.

Figure 5.11 illustrates this usage of our model. The solid lines in this plot reflect one simulation of the RWP regime in the Quinta scenario at instants 0 seconds (initial placement using the initial positions from the Quinta trace), 50, 500, 700 and 1500 seconds of simulation. It is possible to see how the solid lines "move away" from the initial conditions and converge to the steady-state. Another simulation with the same seed and parameters was run, using the initial node placement given by the normalized output of the ODE framework. The dashed lines correspond to this initial distribution, and the measured final distribution at 1500 seconds of the same simulation run. As we can see, these two curves are very similar, showing that the output of our model gives an accurate steady-state distribution for the studied mobility regime, where the final distribution measured at the end of the simulation does not deviate form the initial placement provided by the ODE model. These two curves (the dashed ones) are a very close

match also to the final distribution of the simulation that uses the trace's initial node placement (solid line at 1500 seconds). Density results of simulations using the Natural mobility regime were also generated, with the exact same parameters and scenario. The same behavior was observed.

In conclusion, the main advantages of using the output of our framework for initial node placement in wireless network simulations are two-fold. First, as illustrated by Figure 5.11, by defining what the steady-state node density distribution is, the model saves considerable simulation time, which is the time it takes to get to the steady-state behavior. For instance, in the case of the RWP simulation, for this particular scenario, steady state is achieved after 700 seconds, since the 700 seconds curve and the 1500 second are almost indistinguishable. Second, this result is important, once it is very difficult to determine when a model is going to reach a steady state, since this time varies from scenario to scenario and according to the mobility model and parameters used. Using our ODE model's output to establish what the steady-state behavior is constitutes a technique that is much more scientifically sound than figuring out when steady state is achieved by inspecting the mobility model's behavior over time.

5.6 Conclusions

We investigated the spatial density properties of waypoint-like mobility, which can describe some instances of human mobility. To this end, we developed an Ordinary Differential Equations (ODEs) framework to model spatial node density. With our model, it is possible to study the steady-state behavior of node density for waypoint-based mobility regimes as well as real mobility described by GPS traces. We validate our approach by comparing its results against GPS traces, and present steady-state spatial distribution for two synthetic mobility regimes in three different scenarios. To the best of our knowledge, this is the first approach to spatial node density modeling that is generic enough that can be applied to any waypoint-like mobility regime.

We also use the proposed model to show the inability of mobility regimes that apply the preferential attachment principle to maintain node clusters and preserve the desired node density distribution. In the case of these type of mobility models, we show that in steady-state, their density distribution approaches the distribution of a totally random mobility regime, such

as the RWP mobility model.

We conclude by discussing how our modeling framework can be applied to derive spatial density steady-state conditions. The resulting distribution can then be used to perform initial node placement when setting up mobile network simulations.

Chapter 6

Accounting for Social Structure in Mobility Modeling

In the previous chapters we studied node spatial density and proposed a simple mobility model capable of forming and maintaining clusters of nodes, the IWP mobility model. However, IWP can only account for clustering (i.e., the tendency of individuals to form communities beyond what can be explained by transitivity) and is not able to account for differential popularity, where a specific node cannot choose to stay connected for a prolonged period of time with another node. It also fails to represent transitivity (i.e., two individuals that have a tie to a third one are more likely to be tied together). Those are important characteristics that should be present and accounted during the simulations, and the chosen mobility regime should be capable of representing them.

For these reasons, in this chapter, we introduce a user mobility modeling framework that accounts for both the users' social structure as well as the geographic diversity of the region of interest. SAGA, or Socially– and Geography-Aware mobility model, captures social features through the use of *communities* which cluster users with similar features such as average time in a cell, average speed, and pause time. SAGA accounts for geographic diversity by considering that different communities exhibit different interests for different locales; therefore, different communities are attracted to certain physical locations with different intensities. Besides introducing SAGA, the contributions of this work include: a model calibration approach based on formal statistical procedures to extract social structures and geographical diversity

from real traces and set SAGA's parameters; and validation of SAGA by applying it to real mobility traces. Our experimental results show that, when compared to existing mobility regimes such as Random-Waypoint and Preferential-Attachment based mobility, SAGA is able to preserve the desired non-uniform node spatial density present in real user mobility, creating and maintaining clusters and accounting for differential node popularity and transitivity.

6.1 Introduction

Mobility models are an indispensable tool in the design and evaluation of wireless networks and their protocols. Therefore, defining realistic mobility models is critical when simulating applications and protocols for mobile networked environments. This is the case of infrastructure-based networks (e.g., wireless LANs or WLANs), but even more so in the case of infrastructure-less networks, a.k.a., wireless, self-organizing networks (WSONs), which include wireless mobile ad-hoc networks (MANETs), wireless sensor networks (WSNs), and disruption-tolerant networks (DTNs). In these networks, unlike their infrastructure-based counterparts where only the end user node is mobile, every node may have unlimited mobility. Indeed, most network simulators include "mobility generators", which, following a pre-specified mobility regime, determine the position of network nodes over time during simulation runs. These synthetic mobility generators have been extensively used in the study of wireless networks [15].

An example of a widely used synthetic mobility model is the Random Waypoint (RWP) mobility regime. However, RWP mobility has been known to exhibit some undesirable properties. For example, Bettstetter et al. [11] show that random mobility leads to "homogeneous" steady-state spatial node distribution [1]. This means that if there were any spatial density heterogeneity, or node clustering, in the initial spatial distribution of the nodes, random mobility does not preserve it.

To address the problem of unrealistic mobility regimes, a number of models that are based on real traces and mobility patterns have been proposed; notable examples include [49, 7, 8, 60, 31]. Additionally, several efforts have focused on the "scale-free" properties observed in many real networks such as the Internet, the Web, and some social networks, to name a few. In

[1] The use of the term "homogeneous node distribution" refers to the definition of homogeneity as given in [11] and reflects the fact that there is no significant concentration of nodes (clusters).

particular, the seminal work of Barabási and Albert [3] 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 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., [36, 37, 13, 47, 48, 41]) try to mimic real human mobility by following the preferential attachment principle. This is done by defining *attraction points*, whose probabilities of attracting other nodes increase as more nodes congregate around them.

In our previous work, we have shown that using the preferential attachment principle to model human mobility leads to undesirable long-term behavior [52]. 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 Random Waypoint model. Instead, real human mobility exhibits "persistent" density heterogeneity as shown in [52].

More recently, alternate approaches to developing realistic mobility regimes rely on social interactions between nodes to drive the selection of movement direction and destination [68, 72, 12, 27, 65, 21, 48]. These models usually construct a social network either using outside information unrelated to the mobility trace, or on the basis of geographical collocation. As will become clear in the remainder of this chapter, we take a completely different approach to modeling human mobility.

Our proposed framework accounts for the social structure in the data through the inclusion of communities and the use of the preferential attachment principles to account for the interaction of nodes within and between communities. We thus refer to our models as "socially-inspired". The proposed mobility modeling framework is also "geography-aware" as it accounts for the heterogeneity of the different locales in the region of interest. These two features allow our model to account for features observed in real life networks such as differential popularity, transitivity and clustering. Moreover, our modeling framework applies formal statistical learning techniques to extract, from real mobility traces, both the nodes' social structure as well as the region's geographical diversity for model calibration.

Overall, the contributions of our work are four-fold: (1) we introduce a novel mobility modeling framework that is both socially-inspired and geography-aware, (2) we propose an effi-

cient method to extract social structures and geographical diversity from real traces to calibrate and set the model's parameters based on formal statistical procedures, (3) we validate our work by applying it to real mobility scenarios and comparing our results against real measurements, (4) we show that our model is able to preserve the desired non-uniform node spatial density, creating and maintaining clusters and accounting for differential node popularity and transitivity; it thus solves the problem exhibited by preferential attachment based mobility models.

The remainder of this chapter is organized as follows. Section 6.2 places our work in perspective by presenting related work in mobility modeling and characterization. We present our mobility modeling framework in detail in Section 6.3. Section 6.4 we present the GPS traces used for model calibration and validation, while Section 6.5 describes how we calibrate our model parameters . Section 6.6 presents the validation of our work towards modeling real human mobility. Finally, Section 6.7 concludes the chapter with a discussion of future work.

6.2 Related Work

Motivated by the growing need to evaluate mobile networks and their protocols under realistic conditions, there has been considerable work in trying to depart from random– towards more realistic mobility models. In this section, we do not attempt to present an exhaustive survey of the state-of-the-art in mobility modeling but, rather, highlight efforts that are more closely related to our work on socially-inspired mobility modeling.

The scale-free properties of real networks were investigated in several studies. As pointed out in Section 6.1, an important milestone on the subject is the work by Barabási and Albert [3] which demonstrated that many real large networks are scale free, that is, the node degree in the network graph follows a power law. Several mobility models have been inspired by the Barabási-Albert preferential attachment principle. For example, in [13], a model based on preferential attachment has been proposed, where the choice of going towards an attraction region is weighted proportionally to the region's popularity (i.e., the number of other nodes that chose it) and inversely proportional to the distance to it. The work proposed in [41] is another example of a model that follows Barabási-Albert's growth and preferential attachment principles. The same concept is also used in [47] where nodes are also driven by identified social interactions, following homofilly by attributes.

More recent approaches try to account for social interaction between nodes. Some

examples of mobility modeling for social networks include the work in [28] and [48]. In these models, node movement is influenced by the strength of social ties and the choice of an attraction point is based on the history of visits of other nodes to that location.

The model proposed in [21] defines a set of locations called anchors; dictated by an anchor association function, nodes are associated with specific anchors at specific times during the simulation. Also, nodes are associated with each other by another function that determines their social acquaintances. These values change over time given that, e.g., coworkers are likely to meet during working hours and family members more likely to meet during free time. They also define the concepts of attraction and repulsion, and evaluate their approach comparing the trace generated by their approach against the social graph used to feed the model.

Two basic behaviors, namely *socialize* and *isolate* are defined in [12]. The strength of social interactions are represented by an *acquaintance matrix* and its complement, the *alienness matrix*. Then, these are used as input to define the attraction and repulsion forces acting on each node, which define the node's direction while moving, as well as its acceleration and velocity. The work in [72] also computes acting forces of attraction and repulsion over the nodes that are given by a social interaction matrix. Here, in addition to the social forces between nodes, they also take into account the forces exerted by locations (attraction points).

In [65] authors also use the notion of communities and incorporate assortative/dissortative behavior, that is given by an *Attractor Matrix*. This matrix describes the attributes of each node, such as gender, age and field of work, for example; these attributes are matched through an *Attribute Matching Matrix*. Nodes then move through communities looking for communities where they find other nodes that are a better match to their attributes.

Another mobility model based on the formation of communities is introduced in [68]. It builds a k-clique overlapping community structure which satisfies the common statistical features observed from real social networks. Functions that describe the attraction of a node to a group and a location are defined and these graph connections change over time based on the defined overlapping graph.

The social mobility regimes presented here construct a social network using outside information unrelated to the mobility itself, such as a "social trace" (a priori description of the relationship between nodes or its attributes). Our proposed framework accounts for the social structure in the data through the inclusion of communities and the use of the preferential

attachment principles to account for the interaction of nodes within and between communities. Our proposal also accounts for the heterogeneity of the different locales in the region of interest. These two features allow our model to account for features observed in real life networks such as differential popularity, transitivity and clustering. Moreover, our modeling framework applies formal statistical learning techniques to extract, from real mobility traces, both the nodes' social structure as well as the region's geographical diversity for model calibration.

6.3 Socially- And Geography-Aware (SAGA) Mobility Model

Our model, SAGA for Socially- And Geography-Aware mobility model, captures both social interactions among individuals, as well as the geographical diversity of the region of interest. At the social level, our model captures *transitivity* in the relationship among subjects and could easily be adapted to capture *homophily* by attributes, i.e., the tendency of individuals with similar or completely opposite features to have a higher probability of being linked [27]. As described in detail below, SAGA captures these features through the use of *communities* or *clusters*. SAGA also accounts for geographical diversity: it considers that different node communities do not have the same interests and thus are "attracted" to certain locales more than others. For example, in a University campus, students are "attracted" to the library, classrooms, or cafeteria in a different way than faculty or staff are. Similarly, it is often the case that students and faculty spend time together in the same physical location (e.g., classrooms). However, it is not as common for staff members and students to be collocated the same way as faculty and students tend to be.

We now proceed to provide a more detailed description of SAGA. We use a *waypoint*-based mobility regime in which: (1) a node chooses its next destination following some given probability distribution, then (2) moves to that destination at a randomly chosen speed uniformly distributed between [V_{min},V_{max}], (3) pauses for some time, and (4) repeats from step (1).

In our model, the area of interest is divided in equally sized squares, or *cells*. Every mobile node independently chooses as their next destination a cell $i \in \{1, \ldots, I\}$, where I is the total number of cells in the simulation area according to $p_{i,k,t}$, the probability that destination i is chosen by a node belonging to community k at time t. Here $p_{i,k,t}$ is constructed on the basis of the intensity $\{\mu_{i,k,t} \in \mathbb{R} \mid \mu_{i,k,t} > 0\}$, which represents the attractiveness of destination i to community k at time t. In other words, $\mu_{i,k,t}$ captures geographic diversity in terms of the

different levels of interests each community has in particular regions at different points in time, along with the preferences of individuals to being in the same locations with members of the same or other communities.

We employ a set of community indicators $\xi_1, ..., \xi_n, ..., \xi_N$ for each user n at all points in time, where N is the total number of users. Community indicator ξ_n indicates which community user n belongs to. Also, $\Pr(\xi_n = k) = w_k$ is the probability that user n belongs to community k, where $\sum_{k=1}^{K} w_k = 1$ and K is the total number of communities that exist in the user population. In other words, w_k defines the portion of the population that belongs to community k. We can now write $p_{i,k,t}$, the probability that a node n, belonging to community k, chooses cell i as its next destination at time t as:

$$p_{i,k,t} = \frac{\mu_{i,k,t}}{\sum_{j=1}^{I} \mu_{j,k,t}} \quad \text{for all} \quad i, j \in \{1, \ldots, I\}. \tag{6.1}$$

where the intensity $\mu_{i,k,t}$ of destination i towards nodes belonging to community k at time t is defined as:

$$\mu_{i,k,t} = \exp\left\{ c_{i,k} + \sum_{l=1}^{K} \theta_{k,l} \log(1 + a_{i,l,t-1}) \right\} \tag{6.2}$$

In Equation 6.2, $c_{i,k} \in \mathbb{R}$ is a constant that reflects the *intrinsic attraction* that cell i exerts on members of community k, independent of the number of other nodes present in the cell. On the other hand, $a_{i,l,t-1}$ is the total number of nodes in cell i that belong to community l at time $t-1$ and can be written as:

$$a_{i,l,t-1} = \sum_{m=1}^{N} \mathbb{1}_{\{m \in \mathbb{N}_{i,t-1}\}}, \tag{6.3}$$

where $\mathbb{N}_{i,t-1}$ is the set of nodes currently present or going towards a location inside cell i at time $t-1$ and $\mathbb{1}_{\{m \in \mathbb{N}_{i,t}\}}$ is an indicator function equal to 1 if node m belongs to the set $\mathbb{N}_{i,t}$ ($m \in \mathbb{N}_{i,t}$) or 0 otherwise. Finally, $\theta_{k,l} = \theta_{l,k}$ expresses the *affinity* between nodes in community k and l. Positive values $\theta_{k,l}$ imply that nodes belonging to community k are attracted to nodes belonging to community l (and vice-versa). On the other hand, negative values imply that nodes belonging to community k are repelled by those in community l.

Note that we express $\mu_{i,k,t}$ as an exponential in (6.2) rather than the more traditional linear structure for the intensity. We do it for two reasons. Firstly, the exponential transformation automatically ensures that intensities are positives without any need to impose constrains

in the space of parameters, and allows for both assortative and dissortative patterns of behavior. Secondly, the representation agrees with the structure of a multivariate logistic regression model (see Section 6.5.2), which simplifies the estimation of parameters from the data.

Table 6.1 summarizes SAGA's parameters.

Param.	Description
N,**I**, **K**	Numb. of nodes, cells and communities respectively.
$a_{i,l,t-1}$	Numb. of nodes in cell i belonging to community l at time $t-1$.
$\theta_{k,l}$	Affinity between nodes from communities k and l.
$c_{i,k}$	Intrinsic attraction of node n towards cell i.
$\mathbb{N}_{i,t-1}$	Set of nodes present or going to cell i at time $t-1$.
$p_{i,k,t}$	Prob. that node n, from community k, chooses cell i at time t.
$\mu_{i,k,t}$	Intensity of nodes from community k towards cell i at time t.

Table 6.1: Summary of SAGA parameters: The methodology used to extract these parameters from real traces is described in Sections 6.5.1 and 6.5.2

Collecting the affinity parameters into a symmetric affinity matrix Θ, as illustrated in Table 6.4, facilitates identifying different association patterns, such as: (1) total assortative behavior, when nodes only interact with other nodes from the same community, (2) total dissortative behavior, where nodes only interact with other nodes from different communities, or (3) an arbitrary association pattern.

6.4 Real Traces

Table 6.2 summarizes the GPS traces we use in our analysis.

Parameter	Quinta	Statefair
Avg. Speed ($\pm\sigma$)(m/s)	1.27($\pm$0.16)	0.58($\pm$0.12)
Avg. Pause ($\pm\sigma$)(sec)	4.61($\pm$1.60)	7.20($\pm$6.18)
Area (meters x meters)	840x840	1260x1260
Duration (sec)	900	8000
Sampling frequency (sec)	1	30
# nodes	98	19
# communities (size)	5(13,27,17,12,29)	2(10,9)

Table 6.2: Summary of the GPS traces we use to calibrate and validate SAGA.

Quinta, refers to the "Quinta da Boa Vista Park" trace, first presented in [6]. It is a GPS trace collected at a park in the city of Rio de Janeiro, Brazil. The park has many trees, lakes, caves, and trails. It houses the National Museum of Natural History and the city Zoo. The *Statefair* trace [56] corresponds to daily GPS track logs collected at the NC State Fair held

in North Carolina, USA.

The information in table 6.2 reflects the speed and pause time measured from the traces and averaged over all nodes, at every sample, during the duration of the trace. The last line in the table show the sizes of the communities identified in the traces, after applying our community identification method. This method is described in the following section.

6.5 Model Calibration

One of the contributions of the SAGA mobility modeling framework is its calibration from real mobility traces using formal, yet simple, statistical learning techniques. In this section we describe our model calibration approach in detail.

6.5.1 Community Identification

Communities of nodes, which are represented by the indicators $\xi_1, \xi_2, \ldots, \xi_N$, are identified using the *complete-linkage agglomerative hierarchical clustering* technique (see for example [19]). Each node is characterized through a vector of features that includes average time spent in each cell, average node speed, and pause time. We are then able to build a matrix $M_{N\times(I+2)}$ with the mobility features extracted from the trace. This matrix has N rows (one row per node) and $I+2$ columns (one column per cell, plus the last two columns for the average speed and pause time measured).

Since we measure dissimilarities among nodes using Euclidean distances, we work with the *logit* transformation of the average time spent on each cell. In addition, to deal with issues associated with zero counts for some cells, we add a single pseudo count to each cell before applying any transformation. The number of clusters is selected using the Bayesian Information Criteria (BIC), as discussed in [22]. The BIC associated with a vector of clustering indicators $\xi = (\xi_1, \xi_2, \ldots, \xi_N)$ is computed as

$$BIC(\xi) = loglik(\xi) - \frac{1}{2}K\log(N)$$

where $loglik(\xi)$ is the logarithm of the likelihood associated with a Gaussian finite mixture model where observations are clustered according to ξ, K represents the number of distinct communities in ξ, and N is the total number of nodes. BIC is computed for different models

and number of clusters, and the one with the largest BIC is selected. All computations were carried out using the package `mclust` in the statistical software `R` [1]; sample code for our model calibration approach is available at [2].

6.5.2 Extracting Model Parameters from Traces

Once communities are identified using the approach described in the previous section, a multivariate logit model [63] is used to estimate the geographical preference parameters $c_{i,k}$ and the affinity matrix $\Theta = [\theta_{k,l}]$. More specifically, let $y_{n,t} = i$ if node n is in cell i at time t. A multivariate logit model assumes that

$$\log\left\{\frac{\Pr(y_{n,t}=i)}{\Pr(y_{n,t}=1)}\right\} = c_{i,\xi_n} + \sum_{l=1}^{K} \theta_{\xi_n,l} \log(1 + a_{i,l,t-1})$$

or, equivalently,

$$\Pr(y_{n,t}=i) \propto \exp\left\{c_{i,\xi_n} + \sum_{l=1}^{K} \theta_{\xi_n,l} \log(1 + a_{i,l,t-1})\right\},$$

where, for identifiability, and without any loss of generality, we set $c_{1,k} = 0$. Hence, the logit model agrees with the structure of the intensity in (6.2).

Computations were carried out using the package `mlogit` in the statistical software `R`. This package uses a Fisher scoring algorithm, which is a variant of the Newton method, to find the maximum likelihood estimates for the parameters of the model along with their standard errors. As before, sample code is available at [2]. To reduce the computational burden associated with fitting these models, we exploit the conditional independence assumption underlying this model and subsample some of the traces. This subsampling has minimal effect on the results as discussed in more detail in Section 6.6.

6.5.3 Other Considerations

We constructed our model using the logarithm of the number of nodes on each cell to measure preferential attachment. This is at odds with the choices made by previous authors, who have preferred the use of linear terms. Our choice was motivated by a formal model comparison among the class of models based on the Bayesian information criteria (BIC). More

generally, it should be possible to estimate the functional form of the preferential attachment using nonparametric techniques. This avenue will be explored elsewhere.

Selection of cell sizes is another important issue for our model. In our examples we have selected the sizes in a mostly ad hoc manner. However, since our model implicitly assumes that the mobility model follows a non-homogeneous Poisson process, the problem of selecting the cell sizes is similar to the problem of selecting bandwidths in kernel density estimate problem, which can be used through the use of generalized cross-validation procedures.

6.6 Validation and Experimental Results

In this section we validate SAGA in two case studies corresponding to the traces described in the previous section, namely the Quinta and Statefair traces. We show how SAGA manages to capture the community structure and geography diversity of the traces. As baseline for our comparison, we also use two other mobility regimes, namely Random Waypoint (RWP) [15] and Natural [13]. We start by briefly describing the RWP and Natural mobility; then we present the setup we used to generate mobility traces under these regimes using the Scengen [62] simulation tool.

6.6.1 Baseline Mobility Regimes

6.6.1.1 RWP Mobility

Random Waypoint (RWP) mobility has been widely used in the study of multi-hop ad-hoc wireless networks (MANETs). Under RWP mobility, mobile nodes are initially placed in the area being simulated according to a given distribution (typically, a uniform distribution). Each node remains in its position for a given period of time, called *pause time* P uniformly chosen in the interval $[0, P_{max}]$, where P_{max} is a pre-specified parameter. After this period, the mobile node chooses a new destination uniformly distributed in the simulation area, and a velocity, also uniformly distributed in the interval $[v_{min}, v_{max}]$, where both v_{min} and v_{max} are pre-specified parameters. Once the destination is reached, the node pauses again and chooses another destination and velocity, as described above.

6.6.1.2 The Natural Mobility Regime

We also compare our results against mobility regimes that follow the *preferential attachment* principle. As representative of this family of mobility regimes, we use the Natural mobility model, or simply *Natural* [13].

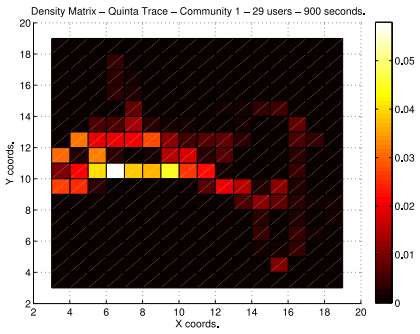
As discussed in Section 6.2, *Natural* is based on attraction points, where the attractiveness of each point is proportional to the *attractor*'s popularity given by the number of nodes at or going towards it and inversely proportional to the distance to it. Thus, the probability $\Pi(a_i)$ that a node z_k chooses an attractor a_i among all possible attractors is proportional to the portion of the total attractiveness it carries: $\Pi(a_i) = \frac{\mathcal{A}_{a_i,z_k}}{\sum_j \mathcal{A}_{a_j,z_k}}$. The attractiveness of an attractor is then defined as:

$$\mathcal{A}_{a_i,z_k} = \frac{(1 + \sum_{z_j \in \mathbb{Z}, z_j \neq z_k} B(a_i, z_k))}{\sqrt{(X_{a_i} - X_{z_k})^2 + (Y_{a_i} - Y_{z_k})^2}} \tag{6.4}$$

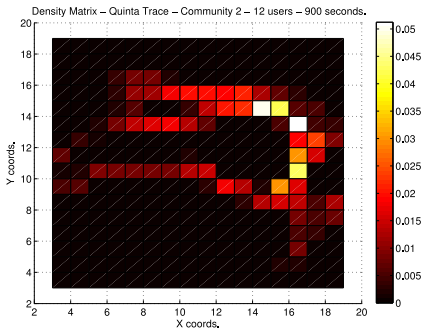
where $B(a_i, z_k)$ is an indicator variable, with $B = 1$ if the individual z_k is going toward or staying at attractor a_i and 0 otherwise, and X and Y are the coordinates of a node and an attractor. In our implementation, we divided the simulation area in equally sized squares, or cells, and consider each cell to be an attraction point. The coordinates (X_{a_i}, Y_{a_i}) mark the center of the i-th attraction point. Once the new destination is known, the node travels towards it with a velocity uniformly distributed in the interval $[v_{min}, v_{max}]$. A pause time is randomly selected once arriving at the destination before choosing another destination and beginning the process again.

6.6.2 Case Study: Quinta

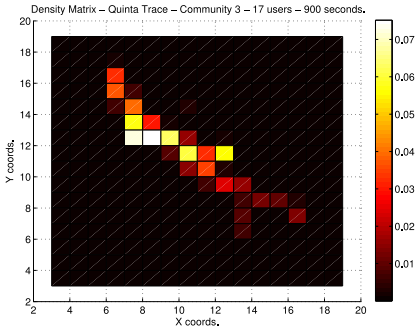
We capture the social structure of the subjects present in the Quinta trace by applying the community identification method described in Section 6.5.1. As a result, we identify 5 distinct communities of sizes 13, 27, 17, 12, and 29 nodes. Figure 6.6.2 shows the average intensity over time associated with each of these communities as well as the average intensity associated with the whole population. In addition, Table 6.3 shows the average speed and average pause time for each community, along with 95% confidence intervals. Note that, for the most part, communities are well separated by the predilection of their members for different geographical regions. The exception are communities 3 and 4, whose members seem to prefer very similar locations. However, it is very clear that members of these two communities have



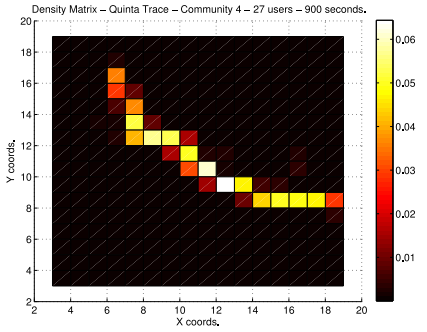
(a) Community 1 (29 nodes)



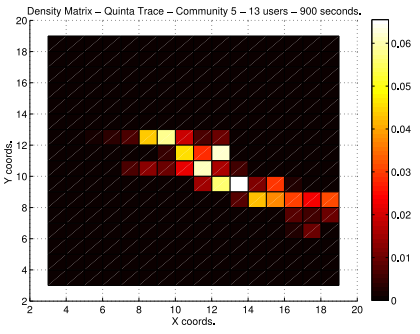
(b) Community 2 (12 nodes)



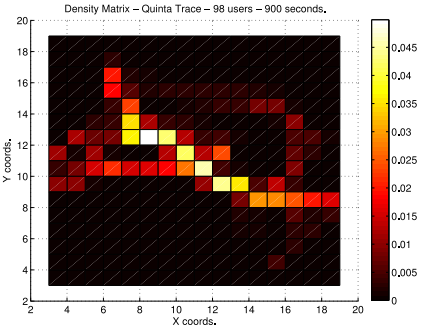
(c) Community 3 (17 nodes)



(d) Community 4 (27 nodes)



(e) Community 5 (13 nodes)



(f) All Nodes (98 nodes)

Figure 6.1: Average intensity over time for the user communities in the Quinta trace (from 6.1(a) to 6.1(e)) and overall intensity for all nodes from all communities (6.1(f)).

different speeds and pause times.

Community	Speed (m/s)	Pause Time (sec)
1	1.124 (1.117, 1.132)	3.316 (2.303, 4.330)
2	1.453 (1.444, 1.462)	2.365 (1.356, 3.375)
3	1.028 (1.017, 1.038)	5.951 (1.974, 9.928)
4	1.139 (1.132, 1.146)	2.036 (1.432, 2.640)
5	1.103 (1.092, 1.114)	4.320 (3.092, 5.548)

Table 6.3: Speed and pause times for user communities in the Quinta trace.

Due to the high computational burden associated with fitting multivariate logit models when both the number of cells and the number of observations is large, we computed our estimates based on a subsample of the original 900 seconds available for inference. This is well justified in this case because of the Markovian structure of the model. To assess the robustness of the procedure we consider various subsample rates (every other second, every five seconds, and every ten seconds). The point estimators were very similar in every case; all the results reported below are based on a subsample of 450 data points generated by retaining every other second of the original trace.

Table 6.4 presents the estimates of the affinity parameters generated by the multivariate logit model, along with 95% confidence intervals. Note that there is evidence of both assortative and disassortative communities. Also, the confidence intervals suggests that a few affinity parameters are indistinguishable from 0 (e.g., $\theta_{2,3}$), but the majority are clearly statistically significant.

As previously pointed out, we use the *spatial density* of mobile nodes to characterize and describe their mobility. In other words, rather than focusing on individual nodes, we are interested in characterizing how node mobility affects the behavior of the network as a whole. In the case of spatial node density, it directly affects fundamental network properties, such as connectivity and capacity, and thus has considerable impact on core network functions, e.g., medium access and routing. Spatial node density can be defined as the number of nodes located in a given unit area. In order to study node's spatial density behavior, we define the node density distribution as the percentage of cells containing more than n nodes.

Figure 6.2 shows the spatial node density distribution for the Quinta trace along with density distributions associated with SAGA, RWP, and Natural mobility regimes. The curve

Θ	1	2	3	4	5
1	-0.138 (-0.258,-0.012)	0.299 (0.239,0.363)	-0.135 (-0.21,-0.071)	0.369 (0.307,0.495)	-0.072 (-0.123,-0.008)
2		-0.411 (-0.479,-0.341)	-0.004 (-0.061,0.039)	0.425 (0.357,0.497)	0.151 (0.110,0.205)
3			0.183 (0.113,0.269)	0.494 (0.417,0.561)	0.339 (0.284,0.374)
4				0.512 (0.405,0.657)	0.452 (0.393,0.518)
5					-0.077 (-0.133,-0.022)

Table 6.4: Affinity parameters associated with the five communities identified by our clustering algorithm. Since the matrix is symmetric, only its upper triangular part is presented.

labeled "*Initial*" shows the trace's density distribution at the beginning of the trace collection interval, while the curve labeled "*Quinta*" shows the distribution at the end of the trace collection interval. The same initial distribution collected from the trace and shown in this figure was used in all the simulations to feed the initial node's placement, leveling the initial conditions in the simulations for all the studied mobility regimes. Data points for the RWP, Natural, and SAGA mobility regimes were averaged over 10 runs.

Note that SAGA's density distribution follows closely the distribution of the real trace. In the case of RWP, the majority of cells present similar number of nodes, and no cells contain significantly higher concentration of nodes. This is also the behavior observed for Natural which can be explained as follows: in the beginning of the simulation the initial distribution weighs on the computation of the attractiveness of the cells, but as the simulation proceeds more and more nodes may choose, with small, but greater than zero probability, cells where there are no other nodes present. Over time, this behavior spreads the nodes around different cells, leveling the attractiveness of cells in general, approaching uniformity and RWP's density distribution. SAGA, on the other hand, is able to maintain the original spatial node density.

Next we consider the average intensities over time, which are presented in Figure 6.3. These plots denote the average percentage of time, nodes choose to be in each cell. More specifically, Figure 6.3(a) shows the average intensity for Quinta and demonstrates that nodes have a strong preference for some areas of the park, which is partially driven by its geographic characteristics (e.g., the presence of a lake in the top right corner of the park). Note that SAGA

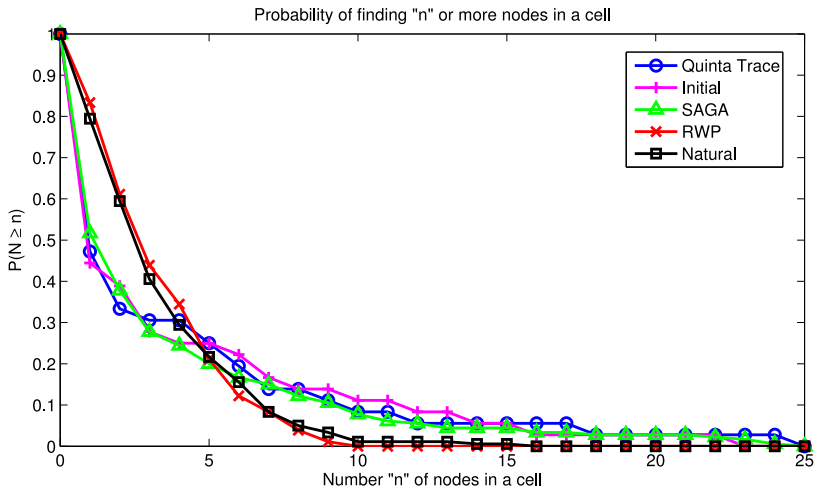


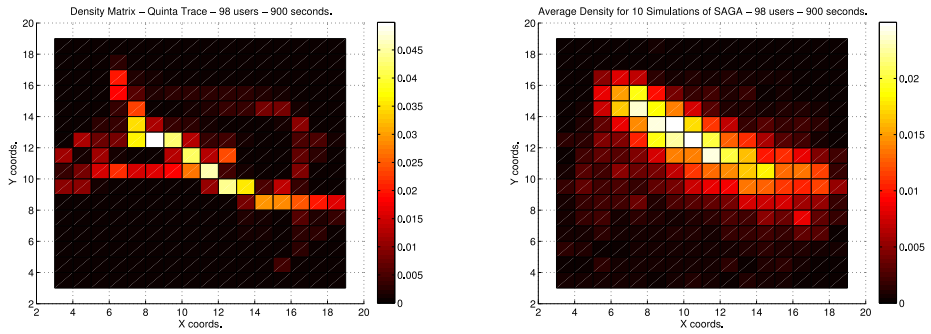
Figure 6.2: Initial and final spatial node density distribution for the Quinta trace compared with the node density distribution of RWP, Natural, and SAGA.

(Figure 6.3(b) is capable of preserving this observed spatial preferences, while both RWP (Figure 6.3(c)) and Natural (Figure 6.3(d)) tend to converge towards a homogeneous distribution over the whole region, which is clearly suboptimal in this context.

Another way to validate SAGA is through a *collocation matrix*, which represents the average amount of time that any pair of nodes spend in the same cell. Figure 6.4 presents the collocation matrices for the Quinta trace, as well as RWP, Natural, and SAGA. Note that Quinta's collocation matrix exhibits a pattern that matches the community structure we identified in Section 6.5.1 quite well. More specifically, note that members of communities 3 and 4 tend to have high collocation. However, members of communities 1, 2, and 5 have fairly low collocation with other members of their same community. In this case, these patterns agree with the structure of the affinity matrix, but this is not necessarily the case in general (for example, see our analysis of the Statefair trace in the following subsection). Our model replicates this structure fairly well, while RWP and Natural (which treat all nodes identically) exhibit uniform collocation matrices.

6.6.3 Case study: Statefair

Similarly to the Quinta trace, we use the method described in Section 6.5.1 to capture the social structure of the subjects present in the Statefair trace. In this trace, we identify 2 distinct communities of sizes 9 and 10 nodes. Figure 6.6.3 shows the average intensity over



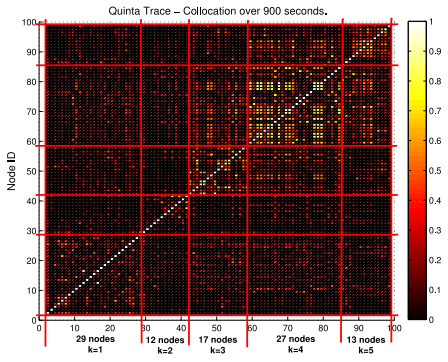
(a) Quinta Trace

(b) SAGA

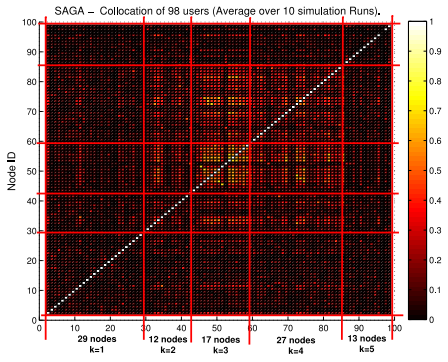
(c) RWP

(d) Natural

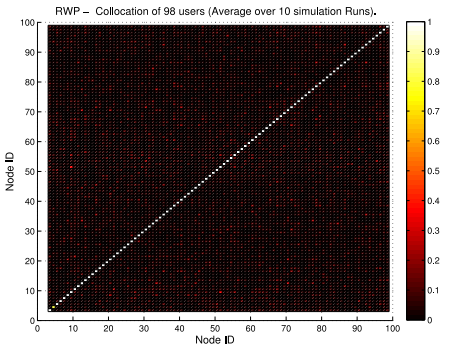
Figure 6.3: Density over time for the Quinta scenario.



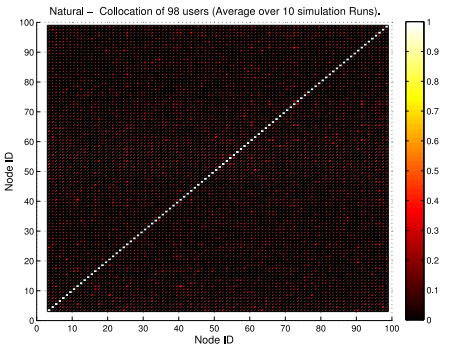
(a) Quinta Trace Collocation Matrix



(b) SAGA Collocation Matrix

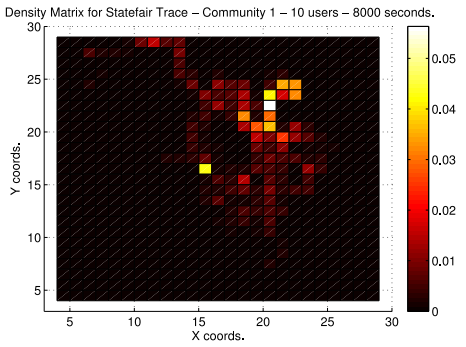


(c) RWP Collocation Matrix

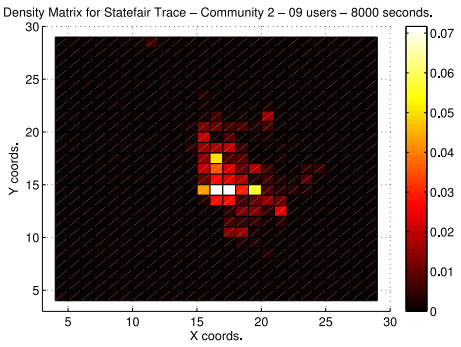


(d) Natural Collocation Matrix

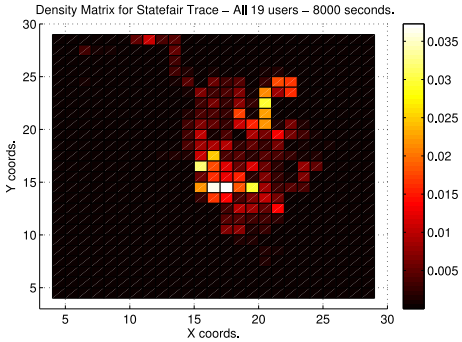
Figure 6.4: Collocation matrices for the Quinta scenario.



(a) Community 1 (9 nodes)



(b) Community 2 (10 nodes)



(c) All Nodes (19 nodes)

Figure 6.5: Average intensity over time for the user communities in the Statefair trace (6.5(a) and 6.5(b)) and overall intensity for all nodes from all communities (6.5(c)).

time associated with each of these communities as well as the average intensity associated with the whole user population. In addition, Table 6.5 shows the average speed and average pause time for each community, along with 95% confidence intervals. Like in the Quinta trace, we manage to clearly differentiate one community from the other and show how members of each community favor different geographical regions. In this case study, not only their preferred location are very distinct for each community, but also their speeds and pause times.

Community	Speed (m/s)	Pause Time (sec)
1	0.606 (0.596, 0.616)	4.663 (1.259, 8.066)
2	0.435 (0.426, 0.444)	8.574 (0.077, 17.071)

Table 6.5: Speed and pause times for user communities in the statefair trace.

The Statefair trace is longer but was sampled at a lower frequency than Quinta. Hence, in this scenario, we were able to use the totality of the measurements (267 samples) in the trace. and there was no need for subsampling due to the lower sampling frequency in this trace (30 seconds).

Table 6.6 presents the estimates of the affinity parameters generated by the multivariate logit model, along with 95% confidence intervals. Different from our first case study, in this scenario there is only evidence of assortative behavior between the two communities.

Θ	1	2
1	0.386	0.716
	(0.223, 0.551)	(0.634, 0.799)
2		0.907
		(0.811, 1.004)

Table 6.6: Affinity parameters associated with the two communities identified by our clustering algorithm.

Figure 6.6 shows the spatial node density distribution metric, as defined above, for the Statefair trace. This figure also shows the density distributions associated with SAGA, RWP, and Natural mobility. Once again, we show that the clustering behavior observed in real live traces is preserved by SAGA. This is in contrast with the preferential attachment based mobility regime, which, by accounting only for the number of nodes in a given location and disregarding social interactions and geographical preferences, leads to undesirable homogeneous node spatial density in steady state. Such steady-state behavior approaches the behavior observed in random mobility regimes as exemplified by the RWP.

We now consider the average intensities over time, which are presented in Figure 6.7. As also observed in the Quinta scenario, the average intensity measured for the Statefair scenario, as shown in Figure 6.7(a), presents some concentration of nodes in certain areas. Figure 6.7(b) shows that SAGA is capable of preserving the spatial preferences observed in the real trace, while both RWP (Figure 6.7(c)) and Natural (Figure 6.7(d)) tend to degrade towards a homogeneous distribution.

Figure 6.8 presents the collocation matrices for the Statefair trace, as well as SAGA. In the case of RWP and Natural, the collocation matrices for the Statefair scenario look very similar to those presented in our first case study, where it is not possible to see any collocation

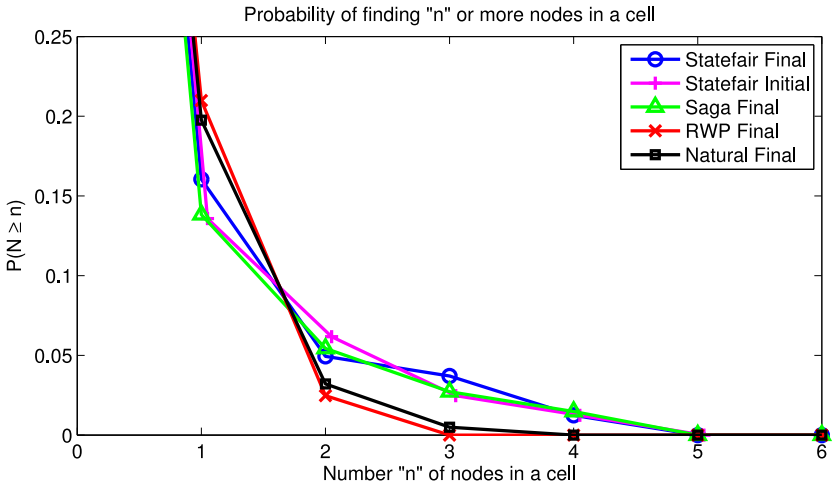
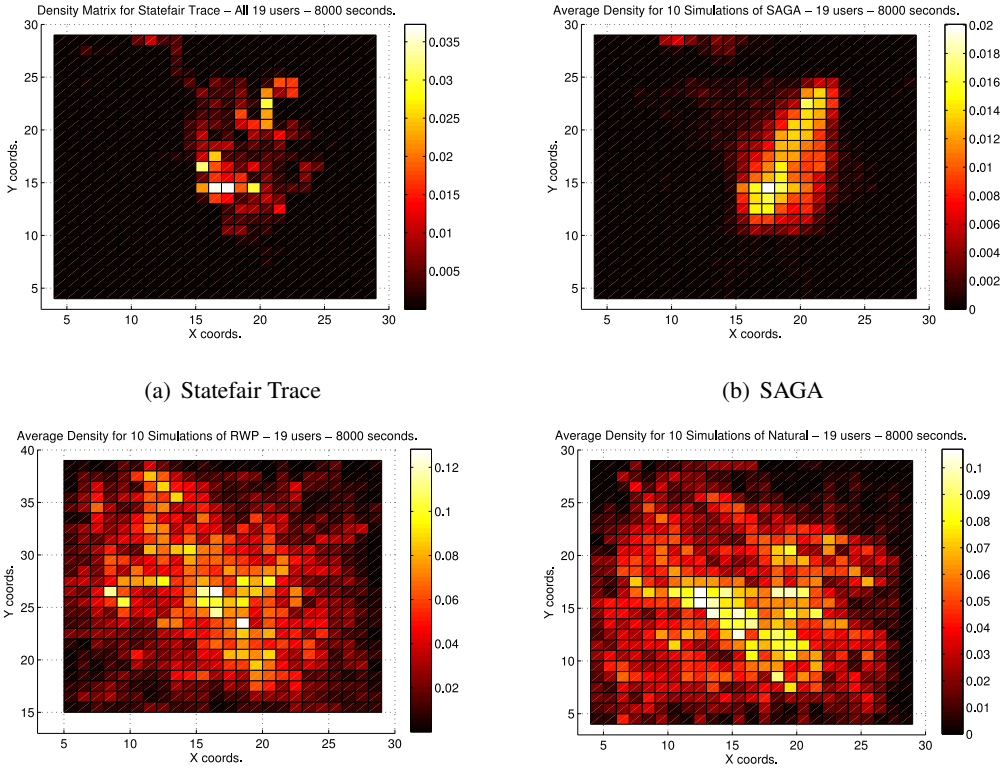


Figure 6.6: Initial and final spatial node density distribution for the Statefair trace compared with the node density distribution of RWP, Natural, and SAGA.



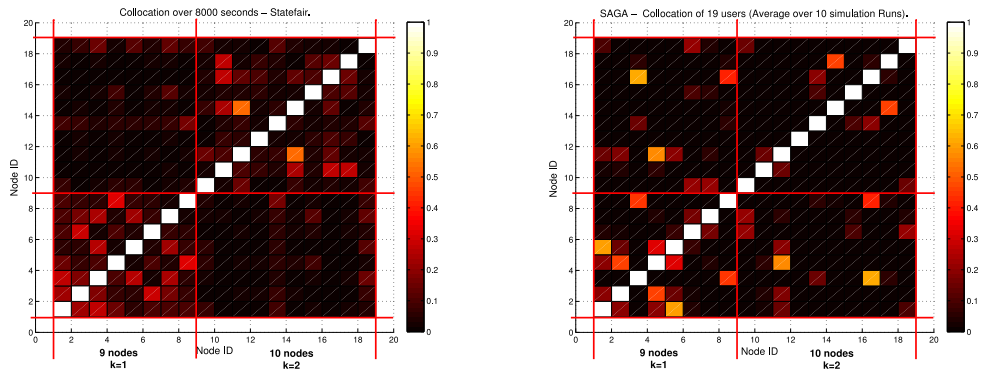
(a) Statefair Trace

(b) SAGA

(c) RWP

(d) Natural

Figure 6.7: Density over time for the Statefair scenario.



(a) Statefair Collocation Matrix

(b) SAGA Collocation Matrix

Figure 6.8: Collocation matrices for the Statefair scenario.

of pairs of nodes for more than 10% of the time. For that reason, and due to space constrains, we do not show these plots. The collocation matrices for the trace and SAGA exhibit a pattern that matches the community structure we identified in Section 6.5.1 quite well. More specifically, note that members of communities 1 and 2 tend to have high colocation, among nodes in their own community, but we also observe nodes from different communities with a higher degree of collocation. This is consistent with the fact that these are highly assortative communities, as described by their affinity matrices (Table 6.6). structure fairly well, while RWP and Natural (which treat all nodes identically) exhibit uniform collocation matrices.

6.7 Conclusion

In this work, we introduced a novel mobility model that accounts for both the social structure exhibited by the user population as well as the geographic diversity of the region of interest. We call our model SAGA for Socially– And Geography-Aware mobility model. SAGA captures social features through the use of *communities* which group users with similar features such as average time in a cell, average speed, and pause time. Acknowledging the importance of preserving– and being able to control node density over time, SAGA also accounts for geographic diversity by considering that different communities exhibit different interests for different locales. Besides introducing SAGA, the contributions of this work include: a model calibration approach based on formal statistical procedures to extract social structures

and geographical diversity from real traces and set SAGA's parameters. We validate our work by comparing SAGA with two real mobility traces and two other mobility regimes, namely Random Waypoint and Natural, the latter representing preferential-attachment based mobility models. We show how to extract social structures and the geographical diversity exhibited by real traces. Moreover, our experimental results demonstrate that, when compared to existing mobility regimes, SAGA is able to preserve the desired non-uniform node spatial density present in real user mobility, creating and maintaining clusters and accounting for differential node popularity and transitivity.

Chapter 7

Conclusion

In this work we have explored the spatial behavior of human mobility through through the analysis of a variety of mobility traces collected in different network environments. As a result, we identify characteristics exhibited by human mobility that, to the best of our knowledge, have not yet been revealed in previous studies. We used a number of real traces that record user mobility in wireless LANs as well as GPS traces. We show that, from a macroscopic level, human mobility is *symmetric*. In other words, the number of users that move from point A to point B approximates the number of users that go in the opposite direction. Additionally, we study the direction of movement which also exhibit symmetric behavior in both real– as well as synthetic mobility.

Motivated by the preliminary findings on the symmetry of human mobility we started to investigate the spatial density distribution properties of real mobility using mobility traces collected in a diverse set of scenarios. More specifically, we chose three scenarios where human mobility was registered using GPS tracking devices carried by users as they move around a city park, a University campus, and a State Fair event. By defining a metric called *spatial density distribution*, we showed that, for all three scenarios, the original non-homogeneous node density distribution does not change over time. We also show that mobility models that rely on computing the probability of choosing a node's new destination dependent on the position of the other nodes can exhibit long transient phases and a steady state similar to completely random behavior as observed when applying the RWP model. To the best of our knowledge, this is the first time that these observations have been reported.

Acknowledging the importance of preserving and being able to control node density over time, we proposed a mobility model called Intensity Waypoint (IWP) that takes a density map as a parameter and where the mobility description is as simple as in the RWP model. By feeding the IWP model with a density map extracted from mobility traces, we were able to match closely the density distribution generated by the IWP to the one generated by the real traces, also preserving the original node density distribution. Moreover, when simulating the operation of a network moving according to the proposed model, we found that performance results expressed by a number of network metrics also match closely results obtained under mobility governed by the real traces. We also compared our results to models whose steady-state tend to random mobility and show that network performance under such regimes deviates from performance under real trace mobility.

In order to study the spatial behavior in steady-state, we proposed a framework of first order ordinary differential equations and solved it numerically to find the steady-state spatial node density of generic waypoint-like mobility regimes. We validate our approach by comparing its results against GPS traces, and present steady-state spatial distribution for two synthetic mobility regimes in three different scenarios. We also use the proposed model to show the inability of mobility regimes that apply the preferential attachment principle to maintain node clusters and preserve the desired node density distribution. In the case of these type of mobility models, we show that in steady-state, their density distribution approaches the distribution of a totally random mobility regime, such as the RWP mobility model.

Finally, we looked into the social aspects of mobility. With our model IWP, we successfully managed to maintain node clustering. However, IWP still fails to implement a scenario where two given specific nodes want to communicate with each other and decide to meet or spend more time together. We then introduce a novel mobility model that accounts for both the social structure exhibited by the user population as well as the geographic diversity of the region of interest. We call our model SAGA for Socially– And Geography-Aware mobility model. SAGA captures social features through the use of *communities* which group users with similar features such as average time in a cell, average speed, and pause time. Besides introducing SAGA, the contributions of this work include: a model calibration approach based on formal statistical procedures to extract social structures and geographical diversity from real traces and set SAGA's parameters. We validate our work by comparing SAGA with two real

mobility traces and two other mobility regimes, namely Random Waypoint and Natural. We show how to extract social structures and the geographical diversity exhibited by real traces, preserving the desired non-uniform node spatial density present in real user mobility, creating and maintaining clusters and accounting for differential node popularity and transitivity.

Bibliography

[1] R: A language and environment for statistical computing - http://www.R-project.org.

[2] Abel Rodriguez and Bruno Nunes and Katia Obraczka. http://inrg.cse.ucsc.edu/inrgwiki/projects, 2012.

[3] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74(1):47–97, 2002.

[4] Mark Allman and Vern Paxson. On estimating end-to-end network path properties. In *SIGCOMM '99: Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication*, pages 263–274, New York, NY, USA, 1999. ACM.

[5] Nils Aschenbruck, Aarti Munjal, and Tracy Camp. Trace-based mobility modeling for multi-hop wireless networks. *Comput. Commun.*, 34, May 2011.

[6] Tiago Azevedo, Rafael Bezerra, Carlos Campos, and Luís de Moraes. An analysis of human mobility using real traces. In *IEEE WCNC 2009.*

[7] Anand Balachandran, Geoffrey M. Voelker, Paramvir Bahl, and P. Venkat Rangan. Characterizing user behavior and network performance in a public wireless lan. In *ACM SIGMETRICS 2002.*

[8] Magdalena Balazinska and Paul Castro. Characterizing mobility and network usage in a corporate wireless local-area network. In *MobiSys 2003*, pages 303–316.

[9] 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*, 2(3):257 – 269, 2003.

[10] Christian Bettstetter. Smooth is better than sharp: A random mobility model for simulation of wireless networks. In *MSWIM '01*, 2001.

[11] Christian Bettstetter, Michael Gyarmati, and Udo Schilcher. An inhomogeneous spatial

node distribution and its stochastic properties. In *Proceedings of the 10th ACM Symposium on Modeling, analysis, and simulation of wireless and mobile systems*, MSWiM '07, 2007.

[12] V. Borrel, F. Legendre, M. Dias de Amorim, and S. Fdida. Simps: Using sociology for personal mobility. *Networking, IEEE/ACM Transactions on*, 17(3):831 –842, june 2009.

[13] Vincent Borrel, Marcelo Dias de Amorim, and Serge Fdida. On natural mobility models. In *WAC*, 2005.

[14] Olivier Bousquet and Manfred K. Warmuth. Tracking a small set of experts by mixing past posteriors. *J. Mach. Learn. Res.*, 3:363–396, 2003.

[15] Tracy Camp, Jeff Boleng, and Vanessa Davies. A survey of mobility models for ad hoc network research. *Wireless Communications and Mobile Computing*, 2:483–502, 2002.

[16] COPPE and UFRJ. http://www.ravel.ufrj.br/orla/ipanema/cobertura3.php, 2010.

[17] Jonathan Corbet, Alessandro Rubini, and Greg Kroah-Hartman. *Linux Device Drivers, 3rd Edition*. O'Reilly Media, Inc., 2005.

[18] CRAWDAD. http://crawdad.cs.dartmouth.edu/.

[19] B. S. Everitt, S. Landau, M. Leese, and D. Stahl. Cluster analysis. *Wiley. Fifth Edition*, 2011.

[20] Alaeddine El Fawal, Jean yves Le Boudec, Kave Salamatian, A. Self, and Limiting Epidemic Service. Self-limiting epidemic forwarding. Technical report, In The First IEEE WoWMoM Workshop on Autonomic and Opportunistic Communications - AOC2007, 2006.

[21] D. Fischer, K. Herrmann, and K. Rothermel. Gesomo: A general social mobility model for delay tolerant networks. In *Mobile Adhoc and Sensor Systems (MASS), 2010 IEEE 7th International Conference on*, pages 99 –108, nov. 2010.

[22] C. Fraley and A. E. Raftery (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, 2002.

[23] Michele Garetto and Emilio Leonardi. Analysis of random mobility models with partial differential equations. *IEEE Transactions on Mobile Computing*, 6(11):1204–1217, November 2007.

[24] Robert B. Gramacy, Manfred K. Warmuth, Scott A. Brandt, and Ismail Ari. Adaptive caching by refetching. In *In Advances in Neural Information Processing Systems 15*, pages 1465–1472. MIT Press, 2002.

[25] M. Gyarmati, U. Schilcher, G. Brandner, C. Bettstetter, Yun Won Chung, and Young Han Kim. Impact of random mobility on the inhomogeneity of spatial distributions. In *IEEE GLOBECOM 2008.*

[26] M. Haeri and A.H.M. Rad. Tcp retransmission timer adjustment mechanism using model-based rtt predictor. In *Control Conference, 2004. 5th Asian*, volume 1, pages 686 – 693 Vol.1, july 2004.

[27] L. Harfouche, S. Boumerdassi, and E. Renault. Towards a social mobility model. In *Personal, Indoor and Mobile Radio Communications, 2009 IEEE 20th International Symposium on*, pages 2876 –2880, sept. 2009.

[28] L. Harfouche, S. Boumerdassi, and E. Renault. Weighted social manhattan: Modeling and performance analysis of a mobility model. In *Personal Indoor and Mobile Radio Communications (PIMRC), 2010 IEEE 21st International Symposium on*, 2010.

[29] Simon Heimlicher and Kavé Salamatian. Globs in the primordial soup: the emergence of connected crowds in mobile wireless networks. In *Proceedings of the eleventh ACM international symposium on Mobile ad hoc networking and computing*, MobiHoc '10, pages 91–100, 2010.

[30] David P. Helmbold, Darrell D. E. Long, and Bruce Sherrod. A dynamic disk spin-down technique for mobile computing. In *MobiCom '96: Proceedings of the 2nd annual international conference on Mobile computing and networking*, pages 130–142, New York, NY, USA, 1996. ACM.

[31] Tristan Henderson, David Kotz, and Ilya Abyzov. The changing usage of a mature campus-wide wireless network. In *ACM MOBICOM 2004*.

[32] Mark Herbster and Manfred K. Warmuth. Tracking the best expert. *Mach. Learn.*, 32(2):151–178, 1998.

[33] Wei-jen Hsu, Kashyap Merchant, Haw-wei Shu, Chih-hsin Hsu, and Ahmed Helmy. Weighted waypoint mobility model and its impact on ad hoc networks. *SIGMOBILE Mob. Comput. Commun. Rev.*, 9:59–63, Jan 2005.

[34] IEEE Computer Society Standards Committee. Working group of the Microprocessor Standards Subcommittee and American National Standards Institute. *IEEE standard for binary floating-point arithmetic*. ANSI/IEEE Std 754-1985. 1985.

[35] V. Jacobson. Congestion avoidance and control. *SIGCOMM Comput. Commun. Rev.*, 25(1):157–187, 1995.

[36] A.P. Jardosh, E.M. Belding-Royer, K.C. Almeroth, and S. Suri. Real-world environment models for mobile network evaluation. *Selected Areas in Communications, IEEE Journal on*, 23(3):622 – 632, 2005.

[37] Minkyong Kim and David Kotz. Extracting a mobility model from real user traces. In *IEEE INFOCOM*, 2006.

[38] David Kotz, Tristan Henderson, Ilya Abyzov, and Jihwang Yeo. CRAWDAD http://crawdad.cs.dartmouth.edu/ dartmouth/campus/syslog/01_04 v.2004-12-18, December 2004.

[39] K.K. Leung, T.E. Klein, C.F. Mooney, and M. Haner. Methods to improve tcp throughput in wireless networks with high delay variability [3g network example]. In *Vehicular Technology Conference, 2004. VTC2004-Fall. 2004 IEEE 60th*, volume 4, pages 3015 – 3019 Vol. 4, sept. 2004.

[40] The GNU C Library. `http://www.gnu.org/software/libc/manual`, 2009.

[41] Sunho Lim, Chansu Yu, and C.R. Das. Clustered mobility model for scale-free wireless networks. In *Local Computer Networks, Proceedings 2006 31st IEEE Conference on*, 2006.

[42] WenYue Lou and ChingYao Huang. Adaptive timer-based tcp control algorithm for wireless system. In *Wireless Networks, Communications and Mobile Computing, 2005 International Conference on*, volume 2, pages 935 – 939 vol.2, june 2005.

[43] R. Ludwig and K. Sklower. The eifel retransmission timer. *SIGCOMM Comput. Commun. Rev.*, 30(3):17–27, 2000.

[44] Liangping Ma, G.R. Arce, and K.E. Barner. Tcp retransmission timeout algorithm using weighted medians. *Signal Processing Letters, IEEE*, 11(6):569 – 572, june 2004.

[45] MATLAB. *version 7.10.0 (R2010a)*. The MathWorks Inc., 2010.

[46] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444, 2001.

[47] A. Mei and J. Stefa. Swim: A simple model to generate small mobile worlds. In *IEEE INFOCOM 2009.*

[48] Mirco Musolesi and Cecilia Mascolo. Designing mobility models based on social network theory. *ACM SIGMOBILE Mobile Computing and Communication Review*, 11:59–70, 2007.

[49] Mirco Musolesi and Cecilia Mascolo. Mobility models for systems evaluation. In *Middleware for Network Eccentric and Mobile Applications*, pages 43–62. Springer, 2009.

[50] William Navidi and Tracy Camp. Stationary distributions for the random waypoint mobility model. *IEEE Transactions on Mobile Computing*, 3:99–108, 2003.

[51] D.W. Ngwenya and G.P. Hancke. Estimation of srtt using techniques from the practice of spc and change detection algorithms. In *AFRICON, 2004. 7th AFRICON Conference in Africa*, volume 1, pages 397 –402 Vol.1, sept. 2004.

[52] Bruno A.A. Nunes and Katia Obraczka. On the invariance of spatial node density for realistic mobility modeling. In *Mobile Adhoc and Sensor Systems (MASS), 2011 IEEE 8th International Conference on*, pages 322 –331, oct. 2011.

[53] Amos R. Omondi. *Computer Arithmetic Systems: Algorithms, Architecture, and Implementation*. Prentice Hall International (UK) Limited, 1994.

[54] Z.J. Pan, Kang L.S., and Y.P. Chen. *Evolutionary Computation*. 2000.

[55] Michal Piórkowski, Natasa Sarafijanovic-Djukic, and Matthias Grossglauser. On clustering phenomenon in mobile partitioned networks. In *Proceeding of the 1st ACM SIGMOBILE workshop on Mobility models*, MobilityModels '08, pages 1–8, 2008.

[56] Injong Rhee, Minsu Shin, Seongik Hong, Kyunghan Lee, and Seongjoon Kim. CRAWDAD http://crawdad.cs.dartmouth.edu/ ncsu/mobilitymodels/GPS/KAIST v.2009-07-23, July 2009.

[57] I. Rubin and C.W. Choi. Impact of the location area structure on the performance of signalling channels in wireless cellular networks. In *IEEE Commun. Mag.*, volume 35, pages 108–115, 1997.

[58] Scalable Network Technologies. Qualnet 4.0.

[59] Libo Song, David Kotz, Ravi Jain, and Xiaoning He. Evaluating location predictors with extensive wi-fi mobility data. In *IEEE INFOCOM 2004.*

[60] Diane Tang and Mary Baker. Analysis of a local-area wireless network. In *ACM MOBICOM 2000*.

[61] Diane Tang and Mary Baker. CRAWDAD http://crawdad.cs.dartmouth.edu/ stanford/gates v.2003-10-16, October 2003.

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

[63] K. Train. Discrete choice modelling, with simulations. *Cambridge University Press*, 2004.

[64] C. Tuduce and T. Gross. A mobility model based on wlan traces and its validation. In *IEEE INFOCOM 2005.*

[65] Jingxin Wang, Jian Yuan, Xiuming Shan, Zhenming Feng, Junwei Geng, and Ilsun You. Samob: A social attributes based mobility model for ad hoc networks. In *Innovative*

Mobile and Internet Services in Ubiquitous Computing (IMIS), 2011 Fifth International Conference on, pages 444 –449, 30 2011-july 2 2011.

[66] Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications*. Number 8. Cambridge University Press, 1994.

[67] H. White, S. Boorman, and R. Breiger. Social structure from multiple networks: I. blockmodels of roles and positions. *American Journal of Sociology*, 81(4):730–80, 1976.

[68] Shusen Yang, Xinyu Yang, Chao Zhang, and E. Spyrou. Using social network theory for modeling human mobility. *Network, IEEE*, 24(5):6 –13, september-october 2010.

[69] J. Yoon, M. Liu, and B. Noble. Random waypoint considered harmful. In *INFOCOM 2003.*, 2003.

[70] Jungkeun Yoon, Brian D. Noble, Mingyan Liu, and Minkyong Kim. Building realistic mobility models from coarse-grained traces. In *MobiSys 2006*.

[71] Xiaolan Zhang, Giovanni Neglia, Jim Kurose, and Don Towsley. Performance modeling of epidemic routing. *Comput. Netw.*, 51:2867–2891, July 2007.

[72] Chen Zhao and M.L. Sichitiu. N-body: Social based mobility model for wireless ad hoc network research. In *Sensor Mesh and Ad Hoc Communications and Networks (SECON), 2010 7th Annual IEEE Communications Society Conference on*, pages 1 –9, june 2010.