A Community Based Mobility Model for Ad Hoc Network Research

Mirco Musolesi
Dept. of Computer Science
University College London
Gower Street, London
WC1E 6BT, United Kingdom
m.musolesi@cs.ucl.ac.uk

Cecilia Mascolo
Dept. of Computer Science
University College London
Gower Street, London
WC1E 6BT, United Kingdom
c.mascolo@cs.ucl.ac.uk

ABSTRACT

Validation of mobile ad hoc network protocols relies almost exclusively on simulation. The value of the validation is, therefore, highly dependent on how realistic the movement models used in the simulations are. Since there is a very limited number of available real traces in the public domain, synthetic models for movement pattern generation must be used. However, most widely used models are currently very simplistic, their focus being ease of implementation rather than soundness of foundation. As a consequence, simulation results of protocols are often based on randomly generated movement patterns and, therefore, may differ considerably from those that can be obtained by deploying the system in real scenarios.

Movement is strongly affected by the needs of humans to socialise or cooperate, in one form or another. Fortunately, humans are known to associate in particular ways that can be mathematically modelled and that have been studied in social sciences for years.

In this paper we propose a new mobility model founded on social network theory. The model allows collections of hosts to be grouped together in a way that is based on social relationships among the individuals. This grouping is then mapped to a topographical space, with movements influenced by the strength of social ties that may also change in time.

We have validated our model with real traces by showing that the synthetic mobility traces are a very good approximation of human movement patterns. We have also run simulations of AODV and DSR routing protocols on the mobility model and show how the message delivery ratio is affected by this type of mobility.

1. INTRODUCTION

The definition of realistic mobility models is one of the most critical and, at the same time, difficult aspects of the simulation of applications and systems designed for mobile environments. Currently, there are very few and very recent public data banks capturing node movement in real large-scale mobile ad hoc environments. For example, researchers at Intel Research Laboratory in Cambridge and the University of Cambridge distributed Bluetooth purpose-made devices to people, in order to collect data about human movements and to study the characteristics of the co-location patterns among people. These experiments were firstly conducted among students and researchers in Cambridge [6] and then among the participants of InfoCom 2005 [16]. Other similar projects are the Wireless Topology Discovery project at the University of California at San Diego [24] and the campus-wide WaveLan traffic measurement and analysis exercises that have been carried out at Dartmouth College [12]. At this institution, a project with the aim of creating a repository of publicly available traces for the mobile networking community has also been started [21].

Until now, real movement traces have been rarely used for evaluation and testing of protocols and systems for wireless networks, with the only exception of [39] and [15], in which the authors used, respectively, the movement traces collected from a campus scenario and direct empirical observations of the movements of pedestrians in downtown Osaka as a basis of the design of their models.

In general, synthetic models have been largely preferred [5]. The reasons of this choice are many. First of all, as mentioned, the available data are limited. Second, these traces are related to very specific scenarios and their validity is difficult to generalize. However, as we will discuss later in the paper, these data show surprising common statistical characteristics, such as the same distribution of the duration of the contacts and inter-contacts intervals. Third, the available traces do not allow for sensitivity analysis of the performance of the algorithm, since the values of the parameters that characterize the simulation scenarios, such as the distribution of the speed or the density of the hosts, cannot be varied. Finally, in some cases, it may be important to have a mathematical model that underlines the movement of the hosts in simulations, in order to study its impact on the design of protocols and systems.

Many mobility models for the generation of synthetic traces have been presented (a survey can be found in [5]). The most widely used of such models are based on random individual movement; the simplest, the Random Walk mobility model (equivalent to Brownian motion), is used to represent pure random movements of the entities of a system [7]. A slight enhancement of this is the Random Way-Point mobility model [18], in which pauses are introduced between changes in direction or speed. More recently, a large number of more sophisticated mobility models for ad hoc network research have been presented [4, 17, 22].

However, all synthetic movement models are suspect because it is quite difficult to assess to what extent they map reality. It is not hard to see, even only with empirical observations, that the random mobility models generate behaviour that is most unhuman-like. This analysis is confirmed by the examination of the available real traces. As we will discuss later in this paper, mobility models based on random mechanisms generate traces that show properties (such as the duration of the contacts between the mobile nodes and the inter-contacts time) very distant from those extracted from real scenarios.

Our work is based on a simple observation. In mobile ad hoc networks, mobile devices are usually carried by humans, so the movement of such devices is necessarily based on human decisions and socialization behaviour. For instance, it is important to model the behaviour of individuals moving in groups and between groups, as clustering is likely in the typical ad hoc networking deployment scenarios of disaster relief teams, platoons of soldiers, groups of vehicles, etc. In order to capture this type of behaviour, we define a model for group mobility that is heavily dependent on the structure of the relationships among the people carrying the devices. Existing group mobility models fail to capture this social dimension [5].

Fortunately, in recent years, social networks have been investigated in considerable detail, both in sociology and in other areas, most notably mathematics and physics. Mathematical models of such networks have been empirically shown to be useful in describing many types of relationships, including real social relationships [31, 28].

In this paper, we propose a new mobility model that is founded on social network theory. One of the inputs of the mobility model is the social network that links the individuals carrying the mobile devices. The model allows collections of hosts to be grouped together in a way that is based on social relationships among the individuals. This grouping is only then mapped to a topographical space, with topography biased by the strength of social ties. The movements of the hosts are also driven by the social relationships among them. The model also allows for the definition of different types of relationships during a certain period of time (i.e., a day or a week). For instance, it might be important to be able to describe that in the morning and in the afternoon of weekdays, relationships at the workplace are more important than friendships and family one, whereas the opposite is true during the evenings and weekends.

We evaluate our model using real mobility traces provided by Intel Research Laboratory in Cambridge and we show that the model provides a good approximation of real movements in terms of some fundamental parameters, such as the distribution of the contacts duration and inter-contacts time. In particular, the data show that an approximate power law holds over a large range of values for the inter-contacts time. Instead, contacts duration distribution follows a power law for a more limited range. These characteristics of distribution are also very similar to those observed by the researchers at the University of California at San Diego and Dartmouth College [6].

The proposed model is partially based on the work presented in [25]. With respect to that short paper, many aspects of the model have been revised to try to map reality with more accuracy. More specifically, in this work the formation of the groups is based on an algorithm for the detection of communities in social networks [27]. The placement of the groups and the dynamics of the hosts in the geographic space have also been completely re-designed. Furthermore, this paper presents a thorough evaluation of the model and a comparison with real traces, which is not presented in [25].

The paper has the following structure: Section 2 contains a definition of social network and illustrates some of the results offered by social network theory. Section 3 shows how these results can be used to design a social network founded mobility model. Section 4 illustrates the results of the evaluation of the model based on the comparison with real traces; some simulation results about the impact of the proposed mobility model on the performance of the AODV and DSR protocols are also discussed. In Section 5 we compare the proposed mobility model with the current state of the art and we outline our current research directions. Section 6 concludes the paper, summarizing the original contribution of our work.

2. SOCIAL NETWORKS

A social network describes a set of people (or groups of people) with some patterns of contact or interaction among each others [36]. Research studies in the area of social networks started in the 1920s [9]. However, the first significant quantitative results were presented by Rapoport [33] and his colleagues in the 1950s and 1960s in a series of papers in which they analyzed the statistics of epidemic diffusion in populations characterised by different social structures. Whilst this was pioneering exploratory work, it was not rigorous from a scientific point of view. However, in that period, a renewed interest in graph theory led to the definition of the so-called random graphs by Paul Erdős and Alfred Rényi [8]. This, then, was the beginning of the complex networks research area, investigating properties such as their topology, average diameter and degree of connectivity, as well as the presence of clusters in networks.

In the recent years, various types of networks (such as the Internet, the World Wide Web and biological networks) have been investigated by researchers especially in the statistical physics community. Theoretical models have been developed to reproduce the properties of these networks, such as the so-called small worlds model proposed by Watts and Strogats [41] or various scale-free models [29, 40]. Excellent reviews of the recent progress in complex and social networks analysis may be found in [1] and [29].

However, as discussed by Newman and Park in [31], social networks appear to be fundamentally different from other types of networked systems. In particular, even if social networks present typical small-worlds behaviour in terms of the average distance between pairs of individuals (the so-called *average path length*), they show a greater level of clustering. In particular, in [31] the authors observe that the level of clustering seen in many non-social systems is no greater than in those generated using pure random models. Instead in social networks, clustering appears to be far greater than in networks based on stochastic models. The authors suggest that this is strictly related to the fact that humans usually organize themselves into *communities*. Examples of social networks used for these studies are rather diverse and include, for instance, networks of

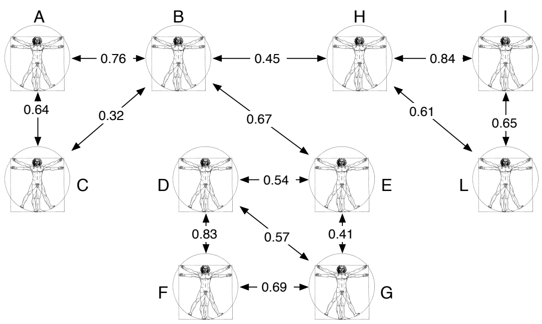

Figure 1: Example of social network.

coauthorships of scientists [28] and the actors in films with Kevin Bacon [41].

Many mathematical models have been proposed in the recent years to generate synthetic social networks [41] that show the same properties of real ones. We will use these results in order to generate realistic social networks structures that are one of the fundamental inputs of the proposed mobility model.

3. DESIGN OF THE MOBILITY MODEL

In this section we show how we designed a mobility model which is founded on the results of social network theories briefly introduced. The description of the mobility model, mirroring its conceptual steps, is organized as follows:

- Firstly, we describe how we model social relationships and, in particular, how we use *social networks as input* of the mobility model.
- Secondly, we present the *establishment of the model*: we discuss how we identify communities and groups in the network and how the communities are associated to a geographical space. Our observation here is that people with strong social links are likely to be geographically colocated often or from time to time.
- Thirdly, we describe the algorithm that is at the basis of the *dynamics* of the nodes, that, again, is based on the strength of social relationships. We argue that individuals with strong social relationships move towards (or within) the same geographical area.

3.1 Using Social Networks as Input of the Mobility Model

3.1.1 Modelling Social Relationships

One of the classic ways of representing social networks is *weighted graphs*. An example of social network is represented in Figure 1. Each node represents one person. The weights associated with each edge of the network is used to model the strength of the interactions between individuals [36]. It is our explicit assumption that these weights, which are expressed as a measure of the strength of social ties, can also be read as a measure of the likelihood of geographic colocation, though the relationship between these quantities is not necessarily a simple one, as will become apparent. We model the degree of social interaction between two people using a value in the range $[0, 1]$. 0 indicates no interaction; 1 indicates a strong social interaction. Different social networks can be valid for different parts of a day or of a week[1].

As a consequence, the network in Figure 1 can be represented by the 10×10 symmetric matrix $\mathbf{M}$ showed in Figure 2, where the names of nodes correspond to both rows and columns and are ordered alphabetically. We refer to the matrix representing the social relationships as *Interaction Matrix*.

$$\mathbf{M} = \begin{bmatrix} 1 & 0.76 & 0.64 & 0.11 & 0.05 & 0 & 0 & 0.12 & 0.15 & 0 \\ 0.76 & 1 & 0.32 & 0 & 0.67 & 0.13 & 0.23 & 0.45 & 0 & 0.05 \\ 0.64 & 0.32 & 1 & 0.13 & 0.24 & 0 & 0 & 0.15 & 0 & 0 \\ 0.11 & 0 & 0.13 & 1 & 0.54 & 0.83 & 0.57 & 0 & 0 & 0 \\ 0.05 & 0.67 & 0.24 & 0.54 & 1 & 0.2 & 0.41 & 0.2 & 0.23 & 0 \\ 0 & 0.13 & 0 & 0.83 & 0.2 & 1 & 0.69 & 0.15 & 0 & 0 \\ 0 & 0.23 & 0 & 0.57 & 0.41 & 0.69 & 1 & 0.18 & 0 & 0.12 \\ 0.12 & 0.45 & 0.15 & 0 & 0.2 & 0.15 & 0.18 & 1 & 0.84 & 0.61 \\ 0.15 & 0 & 0 & 0 & 0.23 & 0 & 0 & 0.84 & 1 & 0.65 \\ 0 & 0.05 & 0 & 0 & 0 & 0 & 0.12 & 0.61 & 0.65 & 1 \end{bmatrix}$$

Figure 2: Example of an Interaction Matrix representing a simple social network.

The generic element $m_{i,j}$ represents the interaction between two individuals i and j. We refer to the elements of the matrix as the *interaction indicators*. The diagonal elements represent the relationships that an individual has with himself and are set, conventionally, to 1. In Figure 1, we have represented only the links associated to a weight equal to or higher than 0.25.

The matrix is symmetric since, to a first approximation, interactions can be viewed as being symmetric. However, it is worth underlining that we are using a specific measure of the strength of the relationships. It is probable that by performing psychological tests, the importance of a relationship, such as a friendship, will be valued differently by the different individuals involved; in our modelization, this would lead to an asymmetric matrix. We plan to investigate this issue further in the future.

The Interaction Matrix is also used to generate a *Connectivity Matrix*. From matrix $\mathbf{M}$ we generate a binary matrix $\mathbf{C}$ where a 1 is placed as an entry c_{ij} if and only if $m_{i,j}$ is greater than a specific threshold t (i.e., 0.25). The Connectivity Matrix extracted by the Interaction Matrix in Figure 2 is showed in Figure 3. The idea behind this is that we have an "interaction" threshold above which we say that two people are interacting as they have a strong relationship. The Interaction Matrix (and, consequently, the Connectivity Matrix) can be derived by available data (for example, from a sociological investigation) or using mathematical models that are able to reproduce characteristics of real social networks. As we will discuss in Section 4.2.2, the default implementation of our model uses the so-called Caveman model [41] for the generation of synthetic social networks with realistic characteristics (i.e, high clustering and low average path length). However, this is a customizable aspect and, if there are insights on the type of scenario to be tested, a user-defined matrix can be inputed.

[1]Let us consider a family of three people, with one child. During the days, when the child is at school and the parents at their workplaces, their social relationship is weak (i.e., represented with low values in the matrix). During the evening, the social ties are stronger as the family members tend to be co-located (i.e., high values in the matrix). The relationship between two colleagues sharing the same office will be represented with a value higher than these family relationships during the working hours in week days.

$$\mathbf{C} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

Figure 3: Example of a Connectivity Matrix representing a simple social network.

3.1.2 *Detection of Community Structures*

The simulation scenario is established by mapping groups of hosts to certain areas in geographical space. After the definition of the social graph described above, groups, i.e., the highly connected set of nodes in the graph, need to be isolated. Fortunately, some algorithms can be exploited for this purpose.

We use the algorithm proposed by Newman and Girvan in [30] to detect the presence of community structures in social networks represented by matrices, like the Connectivity Matrix that we have defined in the previous section. This algorithm is based on the calculation of the so-called *betweenness* of edges. This provides a measure of the centrality of nodes. For example, considering two communities connected by few inter-community edges, all the paths through the nodes in one community to nodes in the other must traverse one of these edges, that, therefore, will be characterised by a high betweenness. Intuitively then, one of the possible estimation of the centrality of an edge is given by the number of shortest (geodesic) paths between all pairs of vertices that run along it. In other words, the average distance between the vertices of the network has the maximum increase when the nodes with the highest betweenness are removed.

Therefore, in order to extract the communities from the network, nodes characterized by high values of centrality are progressively detected in subsequent rounds. At each round, one of the edges of the host with the highest centrality is removed. The final result is a network composed of (fairly distinguishable) groups of hosts (i.e., the communities).

The complexity of this algorithm is $O(mn^2)$, considering a graph with m edges and n vertices. The calculation of the shortest path between a particular pair of vertices can be perfomed using a breadth-first search in time $O(m)$ and there are $O(n^2)$ vertices. However, in [30], Newman and Girvan proposed a faster algorithm with a complexity equal to $O(mn)$. A coincise description of this algorithm for the calculation of the betweenness can be found in the appendix of this paper.

As we said, the algorithm can be run a number of times on the graph, severing more and more links and generating a number of distinguishable communities. However we also need a mechanism to stop the algorithm when further cuts would decrease the quality of the results: this would mean that we have reached a state when we have meaningful communities already. We adopted a solution based on the calculation of an indicator defined as *modularity* Q [30]. This quantity measures the proportion of the edges in the network that connect vertices within the same community minus the expected value of the same quantity in a network with the same community division but random connections between the vertices. If the number of edges within the same community is no better than random, the value of Q is equal to 0. The maximum value of Q is 1; such a value indicates very strong community structure. In real social networks, the value of Q is usually in the range $[0.3, 0.7]$ [30]. The analytical definition of the modularity of a network division is presented in Section B of the appendix. At each run the algorithm severs one edge and measures the value of Q. The algorithm terminates when the obtained value of Q is less than the one obtained in the previous edge removal round. This is motivated by the fact that Q presents one or, at maximum, but much more rarely, two local peaks: therefore, we can stop when the first local peak is reached. This is clearly an approximation since the value of the other possible local peak (if exists) may be higher, but it has been observed that the quality of the division that we obtain is, in the vast majority of the cases, very good [30]. Also, by adopting this technique, we considerably simplify the computational complexity of the algorithm.

In order to illustrate this process, let us now consider the social network in Figure 1. Three communities (that can be represented by sets of hosts) are detected by running the algorithm: $C_1 = \{A, B, C\}$, $C_2 = \{D, E, F, G\}$ and $C_3 = \{H, I, L\}$. Now that the communities are identified given the matrix, there is a need to associate them with a location.

3.2 Establishment of the Model: Placement of the Communities in the Simulation Space

After the communities are identified, each of them is randomly associated to a specific location (i.e., a square) on a grid[2]. We use the symbol $S_{p,q}$ to indicate a square in position p, q. The number of rows and columns are inputs of the mobility model.

Going back to the example, in Figure 4 we show how the communities we have identified can be placed on a 3x4 grid (the dimension of the grid is configurable by the user and influences the density of the nodes in each square). The three communities C_1, C_2, C_3 are placed respectively in the grid in the squares $S_{a,2}$, $S_{c,2}$ and $S_{b,4}$.

Once the nodes are placed on the grid, the model is established and the nodes move around according to social-based attraction laws as explained in the following.

3.3 Dynamics of the Mobile Hosts

As described in the previous section, a host is initially associated to a certain square in the grid. Then, in order to drive movement, a goal is assigned to the host. More formally, we say that a host i is associated to a square $S_{p,q}$ if its goal is inside $S_{p,q}$. Note that host i is not necessarily always positioned inside the square $S_{p,q}$, despite this association (see below).

The goal is simply a point on the grid which acts as *final destination* of movement like in the Random Way-Point model, with the exception that the selection of the goal is not as random.

3.3.1 *Selection of the first goal*

[2]A non random association to the particular areas of the simulation area can be devised, for example by deciding pre-defined *areas of interest* corresponding for instance to real geographical space. However, this aspect is orthogonal to the work discussed in this paper.

When the model is initially established, the goal of each host is randomly chosen inside the square associated to its community (i.e, the first goals of all the hosts of the community C_1 will be chosen inside the square $S_{a,2}$).

3.3.2 Selection of the subsequent goals

When a goal is reached, the new goal is chosen according to the following mechanism. A certain number of hosts (zero or more) is associated to each square $S_{p,q}$ at time t. Each square (i.e., place) exterts a certain *social attractivity* for a certain host. The social attractivity of a square is a measure of its importance in terms of the social relationships for the host taken into consideration. The social importance is calculated by evaluating the strength of the relationships with the hosts that are moving towards that particular square (i.e., with the hosts that have a current goal inside that particular square). More formally, given $C_{S_{p,q}}$ (i.e., the set of the hosts associated to square $S_{p,q}$), we define *social attractivity* of that square towards the host, i SA_{p,q_i}, as follows

$$SA_{p,q_i} = \frac{\sum_{\substack{j=1 \\ j \in C_{S_{p,q}}}}^{n} m_{i,j}}{w}$$

where w is the cardinality of $C_{S_{p,q}}$ (i.e., the number of hosts associated to the square $S_{p,q}$). In other words, the social attractivity of a square in position (p,q) towards a host i is defined as the sum of the interaction indicators that represent the relationships between i and the other hosts that belong to that particular square, normalized by the total number of hosts associated to that square. If $w = 0$ (i.e., the square is empty), the value of SA_{p,q_i} is set to 0.

The new goal is then randomly chosen inside the square characterised by the highest social attractivity; it may be again inside the same square or in a different one. New goals are chosen inside the same area when the input social network is composed by loosely connected communities (in this case, hosts associated with different communities have, in average, weak relationships between each others). On the other hand, a host may be attracted to a different square, when it has strong relationships with both communities. From a graph theory point of view, this means that the host is located between two (or more) clusters of nodes in the social network[3].

Let us suppose, for example, that host A has reached its first goal inside the square $S_{a,2}$. The new goal is chosen by calculating the social attractivities of all the squares that compose the simulation space and then by choosing the highest. If, say, square $S_{c,2}$ exterts the highest attractivity (for example, because a host with strong relationship with node A has joined that community), the new goal will then be selected inside that square.

3.3.3 Social Network Reconfigurations and their Effects on the Dynamics of Mobile Hosts

Like in everyone's life, the day movement are governed by different patterns of mobility which depend on the people we need to interact with. For example, most people spend a part of their day at work, interacting with colleagues, and another part at home with their families. In order to model this, we allow the association of different social networks to different periods of a simulation.

[3]This is usually the case of hosts characterised by a relatively high betweenness that, by definition, are located *between* two (or more) communities.

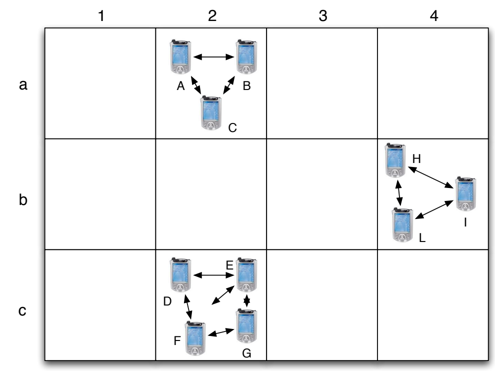

Figure 4: Example of initial simulation configuration.

Periodically, the social networks at the basis of the mobility model can be changed. The interval of time between changes is an input of the model. When the reconfiguration of the underlying social network happens, nodes are assigned to the new communities that are detected in the network using the algorithms described in Section 3.1.2. Communities are then randomly associated to squares in the simulation space. This assignement does not imply immediate relocation of the nodes, instead, it conditions the choice of the next goal. In fact, goals are chosen inside the square of the grid to which the community they belong to is assigned. So hosts will move towards their destination gradually. The nodes start moving towards the geographical region where other nodes that have strong interactions with them will converge. This mirrors the behaviour, for instance, of commuters who travel home every evening to join their families.

4. IMPLEMENTATION AND EVALUATION

In order to evaluate our model we have performed a number of tests, in particular, we have taken real mobility traces collected by Intel Research Laboratory in Cambridge. We have then tested our model using realistic social networks and compared the mobility patterns with the Intel traces. We have also compared the performance of AODV [32] and DSR [19] using the Random Way-Point and our Community based mobility models. In this section, we will present and discuss the results of our simulations comparing them with these data from real scenarios.

4.1 Implementation of the model

We implemented a movement patterns generator that produce primarily traces for the ns-2 simulator [23], one of the most popular in the ad hoc network research community. However, the generator is also able to produce traces in a XML meta-format that can be parsed and trasformed into other formats (for example, by using XSLT) such as the one used by GlomoSim/Qualnet [42]. The model is available for downloading at the following URL: *[omitted for blind review]*.

4.2 Validation of the Model using Real Movement Traces

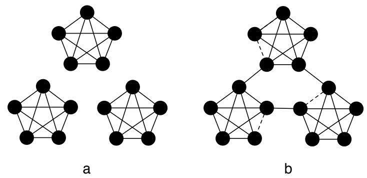

Figure 5: Generation of the social network in input using the Caveman model: (a) initial configuration with 3 disconnected 'caves'. (b) generated social network after the rewiring process.

In this section, we present a comparison of the properties of the movement patterns generated by our mobility model with those of the real traces provided by Intel Research Laboratory in Cambridge. The description of these measurement exercise is presented in [6]. In that paper, the authors also compare their results with other publicly available data sets provided by McNett and Voelker from University of California at San Diego [24] and by Henderson et alii from Dartmouth College [12] showing evident similarities between the patterns movements collected by the three different groups. For this reason, we decided to compare the traces obtained by using our mobility model only with the data provided by the researchers in Cambridge.

4.2.1 Description of the Data Sets

The traces were collected by Intel researchers using iMotes (a modified version of the Berkeley Motes) [2] equipped with Bluetooth. The iMotes were then given to members of the staff of Intel Research Laboratory and University of Cambridge. The iMotes were packed in keyfobs in order to make sure that people carried them around. Each iMotes logged contacts data in a flash memory using the standard Bluetooth Baseband layer inquiry procedure. Every contact was stored as a tuple composed of three fields, the MAC address of the other device, the start and the end of the interval of time of the contact. Every iMotes collected information, not only on the other samplers, but also on the other Bluetooth devices in reach. The iMotes were programmed to perform an inquiry for 5 seconds every $2+\Delta$ minutes with Δ randomly chosen in the range $[-12, 12]$ seconds. This correction was introduced to avoid undesired synchronization effects, i.e., to avoid that the iMotes performed inquiries at the same time. In fact, iMotes are not able to perform and reply to inquiries at the same time.

4.2.2 Description of the Simulation

We tested our mobility model using several runs generating different mobile scenarios and we compared the results with the real movement patterns provided by Intel and synthetic traces generated using a Random Way-Point model.

We tested our model considering a scenario composed of 100 hosts in a simulation area of 5 $km \times 5$ km, divided into a grid composed of 625 squares of 200 m (i.e., the numbers of rows and columns of the grid were set to 25). We chose a relatively large simulation scenario, with a low population density, in order to better see the differences in the results obtained with a Random Way-Point model. In fact, in small simulation areas, the limited possible movements and the higher probability of having two nodes in the same transmission range may affect the simulation results introducing side-effects that are not entirely due to the mobility model.

We also assumed that each device is equipped with an omnidirectional antenna with a transmission range of 250 m, modeled using a free space propagation model. The speeds of the nodes were randomly generated according to a uniform distribution in the range $[1-6]$ m/s. The duration of the simulation is one day and the reconfiguration interval is equal to 8 hours. These values have not been chosen to reproduce the movements described by the traces provided by Intel, rather, we were more interested in observing if similar patterns could be detected in synthetic and real traces. In other words, our goal has mainly been to verify whether the movement patterns observed in Intel traces were reproduced by our mobility model.

A key aspect of the initialization of our model is the selection of the social network in input. We implemented a generator of synthetic social networks using the so-called Caveman Model proposed by Watts [41]. The social network is built starting from K fully connected graphs (representing communities living in isolation, like primitive men in caves). According to this model, every edge of the initial network in input is re-wired to point to a node of another cave with a certain probability p. The re-wiring process is used to represent random interconnections between the communities. Figure 5.a shows an initial network configuration composed by 3 disconnected communities (*caves*) composed by 5 individuals; a possible social network after random rewiring is represented in Figure 5.b.

Individuals of one cave are closely connected, whereas populations belonging to different caves are sparsely connected. Therefore, the social networks generated using this model are characterized by a high clustering coefficient and low average path length. It has been proved that this model is able to reproduce social structures very close to real ones [41]. We generated social networks with different rewiring probabilities, also considering the case of disconnected communities (i.e., $p = 0$).

We also implemented a movement patterns generator based on the Random Way-Point model. We generated traces with the same simulation scenarios in terms of size of the area and characteristics of the mobile devices, with hosts that move with a speed uniformly distributed in the range $[1-6]$ m/s and stop time equal to $[1-10]$ m/s.

We repeated the experiments using a number of runs sufficient to achieve a 10% confidence interval.

4.2.3 Simulation Results

The emergent structure of the network derived by analyzing the Intel traces is typically exponential [1]; in fact, the *degree of connectivity* shows a local peak near the average. Our mobility model (indicated with CM) produces a similar type of distribution as shown in Figure 6. The peak shifts to the right as the density of the squares increases. We analyzed two further properties of the movement patterns, the contact duration and the inter-contacts time. We adopt the same definitions used by the authors of [6] in order to be able compare the results. We define *contact duration* as the time interval in which two devices are in radio range. We define *inter-contacts time* as the time interval between two contacts. These indicators are particularly important in ad hoc networking and, in particular, in *opportunistic mobile networks*, such as delay tolerant mobile ad hoc networks [26, 20]: inter-contacts times define the

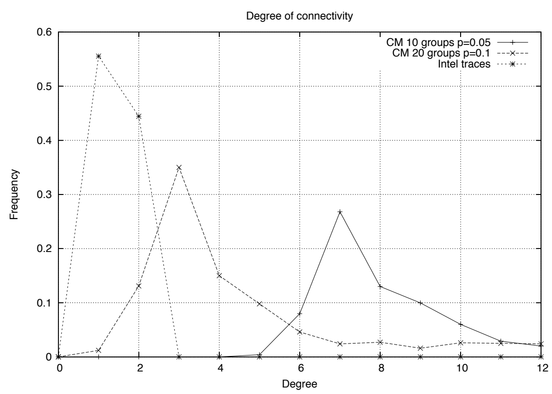

Figure 6: Distribution of the degree of connectivity.

frequency and the probability of being in contact with the recipient of a packet or a potential carrier in a given time period.

Figure 7 shows the comparison between the inter-contacts time and the contact duration cumulative distributions[4] using log-log coordinates. These distributions are extracted from the real and synthetic traces generated by the Random Way-Point (indicated with RWP) and our Community based mobility model with different rewiring probabilities p.

With respect to the inter-contacts time, our traces (excluding the case with $p = 0$ that we will discuss separately) shows an approximate power law behaviour for a large range of values like those extracted from Intel data. A similar pattern can be observed in UCSD and Dartmouth traces [6]. The cumulative distribution related to Random Way-Point, instead, shows a typical exponential distribution. The same behaviour can be observed for the traces generated using our Community based mobility model with a probability of rewiring equal to 0. In fact, in this case, the only movements of the hosts outside the assigned square happen when a reconfiguration takes place (i.e., a new generation of the social networks takes place and a consequent new assignment to different squares in the grid are performed). However, the case of disconnected and isolated communities is not so realistic. As far as the contacts time distribution is concerned, we observe a power law behaviour for a much more limited range of values and, in general, with a lower angular coefficient of the interpolating line. The traces from Dartmouth College and UCSD also show a power law distribution with different angular coefficients [6]. It seems that data related to different scenarios are characterized by different types of power law distribution.

By plotting the same distributions using semi-log coordinates (see Figure 8), the differences between the curves corresponding to real traces and those generated using the Random Way-Point mobility model are even more evident. The exponential nature of the cumulative distribution of the inter-contacts time[5] extracted by the latter is clearly reflected by the approximated straight line that is shown in the figure.

Figure 9.a and 9.b show the influence of the speed respectively on the cumulative distributions of the inter-contacts time and contacts duration. We simulated scenarios with host speed uniformly distributed in the range $[1-6]$, $[1-10]$ and $[1-20]m/s$. The cumulative distributions related to all these scenario can be approximated with a power law function for a wide range of values.

In many of our experiments, the coefficient of the power law of the distribution of the Intel traces is different from those related to synthetic traces generated using our model. Different coefficients can be observed in the available sets of real traces. In a sense, it seems that the values of these coefficients characterize the various mobile settings. It is worth noting that currently there are not available theoretical models that justify the emergence of these distributions.

The impact of the density of the population in the simulation scenario is presented in Figure 10. We simulated scenarios composed of 100, 200, 300 nodes with a starting number of groups for the Caveman model, respectively equal to 10, 20, 30, and a rewiring probability of 0.2. Also in these scenarios, the inter-contacts time and contacts duration distributions follow a similar pattern. As discussed previously, our aim was not to exactly reproduce the traces provided by Intel. However, quite interestingly, we observe that the inter-contacts time distribution lie in between the curves representing the scenario composed of 100 and 200 nodes. The number of nodes recorded in the Intel experiments was in fact 140. Instead, the contacts duration distribution is bounded by the curves extracted by these two synthetic traces for a smaller range of values. Finally, in Figure 11 we consider a scenario composed of 100 hosts connected by a social network generated using different initial numbers of groups (i.e., caves) as input for the Caveman model (with a re-wiring probability equal to 0.1). By varying the number of groups, the density of the squares of the grid changes. The power law patterns can be observed in all the scenario, also with a large number of small initial groups.

4.3 Influence of the Choice of the Mobility Model on Routing Protocols Performance

4.3.1 Simulation Description

In order to be able to compare routing protocol performance with exisiting results, we tested the community model in case of dense networks. Using ns-2, we simulated a scenario composed of 50 hosts and we compared the performance in terms of delivery ratio of the AODV [32] and DSR [19] protocols. We used a $1000m \times 1000m$ simulation area with a maximum node transmission range equal to $250m$. We chose the two-ray pathloss model as propagation model and at the MAC layer, the IEEE 802.11 DCF protocol was used with a bandwidth equal to 2 Mbps. We started 10 sessions between randomly chosen hosts[6] using CBR traffic with data packet size and sending rate respec-

[4]Cumulative distributions are generally used instead of frequency distributions to avoid the issues related to the choice of the bins of the plot. It is possible to prove that if a set of data shows a power law behaviour using a frequency histogram, its cumulative distribution also follows the same pattern.

[5]This behaviour has been theoretically studied and predicted by Sharma and Mazumdar in [37].

[6]This kind of traffic can be considered as a worst case sce-

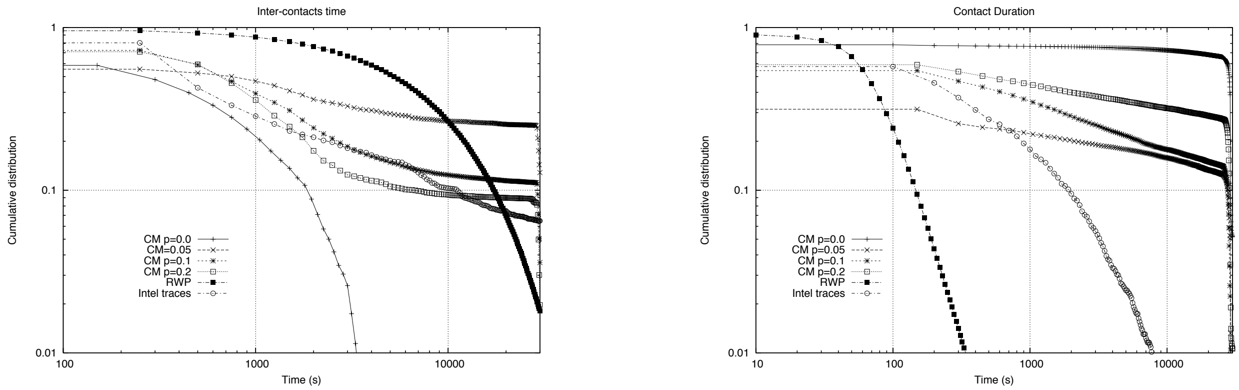

Figure 7: Comparison between synthetic and real traces (log-log coordinates) : (a) cumulative distribution of inter-contacts time in seconds; (b) cumulative distribution of contacts duration in seconds.

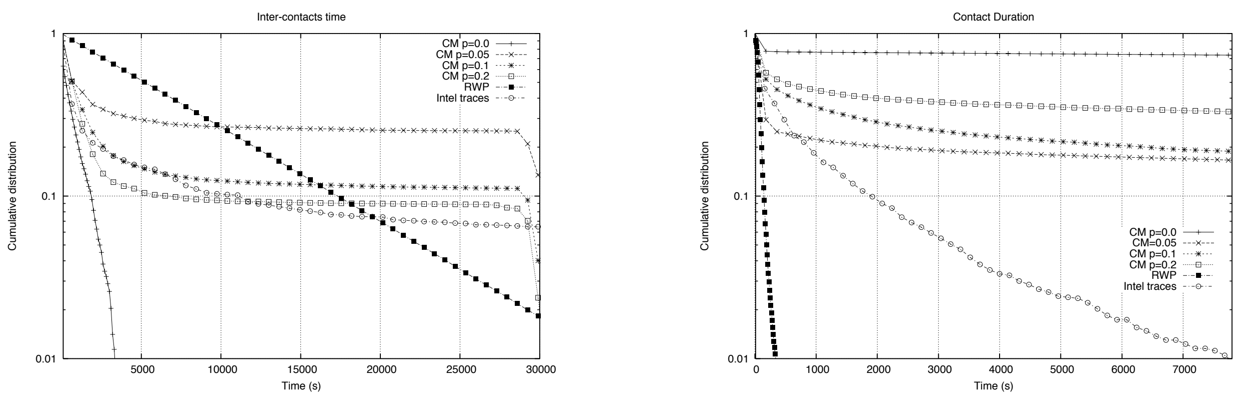

Figure 8: Comparison between synthetic and real traces (semi-log coordinates): (a) cumulative distribution of inter-contacts time in seconds; (b) cumulative distribution of contacts duration in seconds.

tively equal to 512 bytes and 4 packets/second. The simulation time was equal to 2 hours.

We studied the influence of the speed on the performance comparing the results obtained by using the Random Way-Point model and the Community based mobility model presented in this paper. Every node in the simulation is moving at the same speed. With respect to the Random Way-Point model, the stopping times are chosen randomly in the interval $[1-10]m/s$. As far as our mobility model is concerned, the reconfiguration interval was set to 1 hour. The social network in input was generated with the Caveman model with 5 groups of 10 individuals and a re-wiring probability equal to 0.1. The simulation scenario was divided into a $5x5$ grid. We performed a number of runs sufficient to achieve a 10% confidence interval.

4.3.2 Simulation Results

Using the Random Way-Point mobility model, as expected and confirming the results obtained by the authors of these protocols [32, 19], the delivery ratio decreases as the speed increases (Figure 12). Instead, using our model, the decreasing trend of the delivery ratio is less evident, since the emerging structure is composed of groups of hosts moving in limited areas (i.e., the square of the grids) that are ‘bridged’ by hosts roaming among them. In other words, the movement of most of the hosts is constrained in geographical terms so topology changes are less frequent than in the case of a pure random model. As it is possible to observe in Figure 12, the difference in terms of performance using the two mobility models is more evident for the DSR protocol. In case of fixed hosts (i.e., with a speed equal to 0), the delivery ratio that we obtained using our mobility model is lower than in the scenarios with a speed greater than 0, since in the former case, there may be disconnected communities, whereas in the latter, hosts move between communities, providing a link between them.

5. DISCUSSION

nario; in reality, it is probable that sessions will be between hosts of the same community. We plan to investigate this aspect in the future, developing a social networks founded traffic generator.

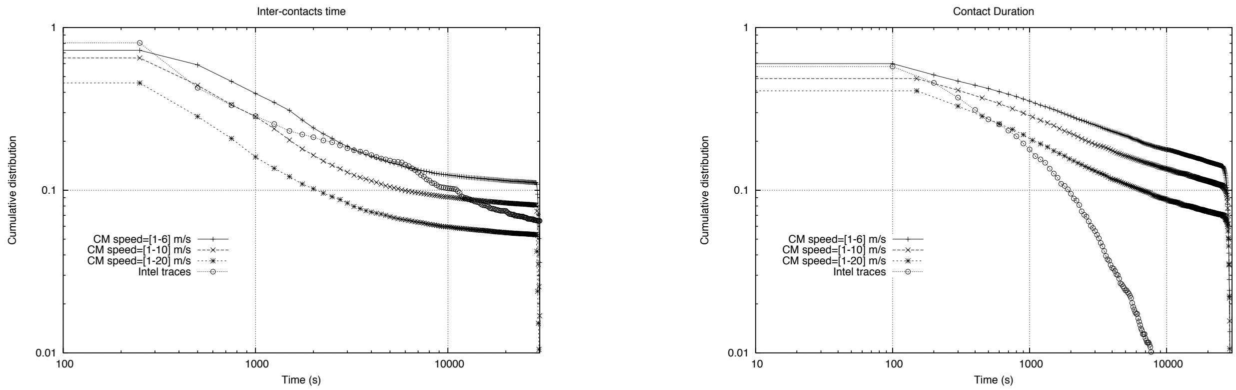

Figure 9: Influence of the hosts speed: (a) cumulative distribution of inter-contacts time in seconds; (b) cumulative distribution of contacts duration in seconds.

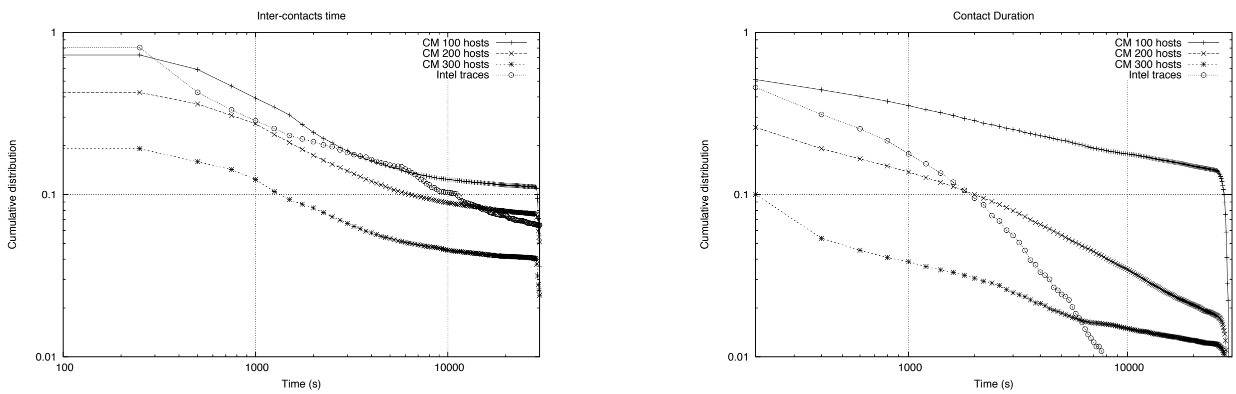

Figure 10: Influence of the density of population: (a) cumulative distribution of inter-contacts time in seconds; (b) cumulative distribution of contacts duration in seconds.

5.1 Related Work

Many mobility models have been presented with the aim of allowing scalability testing of protocols and algorithms for mobile ad hoc networking. A comprehensive review of the most popular mobility models used by the mobile ad hoc research community can be found in [5]. However, it is interesting and, at the same time, surprising to note that even the best solutions and approaches have only been tested using completely random models such as the Random Way-Point model, without grouping mechanisms. A more refined approach used a simple groups mobility model which still had a large random component in the way groups were created and moved [14]. The almost pervasive adoption of such models has generated a considerable amount of work that builds on the reasonableness of random mobility models.

The work most directly related to ours can be found in [13]. This model is predicated upon similar assumptions, but is considerably more limited in scope. In that model hosts are statically assigned to a particular group during the initial configuration process, whereas our model accounts for movement between groups. Moreover, the authors claim that mobile ad hoc networks are scale-free, but the typical properties of scale-free networks are not exploited in the design of the model presented by the authors. The scale-free distribution of mobile ad hoc networks is still not proven in general, since practical measurements are not currently available. Scale-free properties are strictly dependent on the movements of hosts and therefore are dependent on the actual simulated scenarios/applications [11]. With respect to this work, we allow the setting of the initial social network, which conditions the movement patterns, this enables different kinds of networks to emerge, including small world and scale free.

In recent years, many researchers have tried to refine existing models in order to make them more realistic. In [17], a technique for the creation of a mobility models that include the presence of obstacles is presented. The specification of obstacles is based on the use of Voronoi graphs in order to derive the possible pathways in the simulation space. This approach is orthogonal to ours; this would be an interesting extension of the model as discussed in the

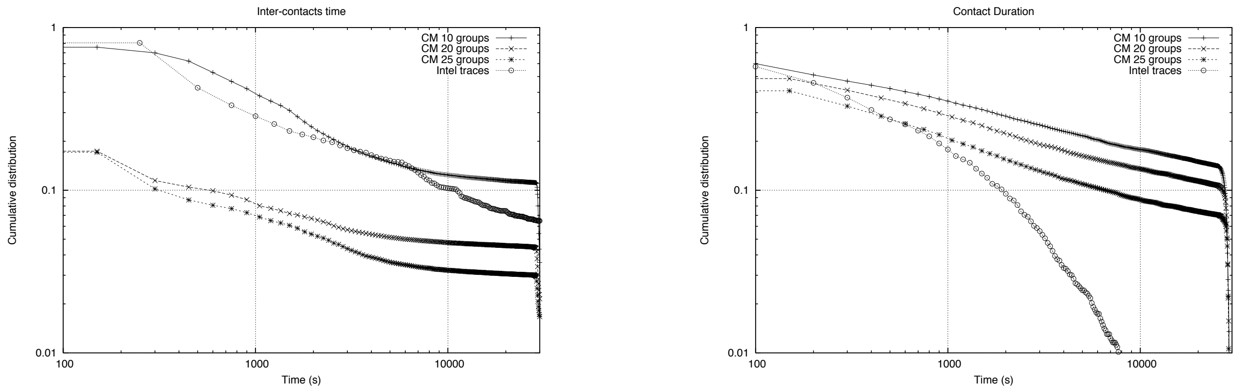

Figure 11: Influence of the number of initial number of groups in input to the network generator based on the Caveman model: (a) cumulative distribution of inter-contacts time in seconds; (b) cumulative distribution of contacts duration in seconds.

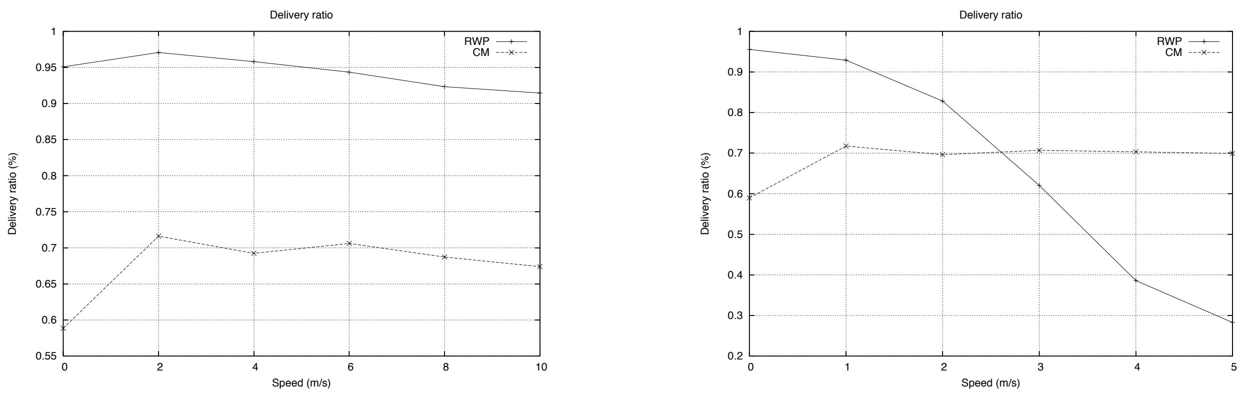

Figure 12: Influence of the mobility model on the protocol performance (delivery ratio vs speed): (a) AODV; (b) DSR.

next section.

Tuduce and Gross in [39] present a mobility model based on real data from the campus wireless LAN at ETH in Zurich. They use a simulation area divided into squares and derive the probability of transitions between adjacent squares from the data of the access points. Also in this case, the session duration data follow a power law distribution. This approach can be a refined version of the Weighted Way-Point Mobility Model [15], based on the probability of moving between different areas of a campus using a Markov model. Moreover, Tuduce and Gross' model represents the movements of the devices in an infrastructure-based network and not ad hoc settings. In [22], the authors try to reproduce the movements of pedestrians in downtown Osaka by analysing the characteristics of the crowd in subsequent instants of time and maps of the city using an empirical methodology. In general, the main goal of these works is to try to reproduce the specific scenarios with a high degree of accuracy. We focus, instead, on the cause of these movements, trying to capture the social dimensions that lead to general emergent human movement patterns.

Some interesting studies have been recently carried out on the connectivity of ad hoc networks with respect to complex networks theory. Glauche et alii in [11] discuss some network properties using percolation theory [38], that is, an application of complex networks theory derived by the investigation of physical phenomena such as phase transitions in molecular lattices. In [35], the authors present mathematical results about the possible emergence of scale-free structures in ad hoc networks. However, the authors consider only *fixed* ad hoc networks (such as peer-to-peer networks), without analysing the influence of movement in the definition of their model.

5.2 Possible Improvements of the Model and Current Research Directions

A number of possible features can be added to the mobility model presented in this paper in order to increase its realism. Many researchers in the ad hoc community have been focussing on different aspects of this problem. Some of the techniques are orthogonal to this work and can be integrated in our model. We plan to explore these possible refinements in the future. More specifically, some of the most important improvements can be summarized as follows:

- **Non random assignment of the nodes to the geographical locations** A possible improvement of the model may be the assignment of the communities based on real mapping between groups of people and geographical locations (such as students moving around lecture rooms and halls in a campus, etc.). An example of this kind of mobility models is [15]. In the current implementation, the generation of the social networks is based on the mathematical model described above and the placement of the communities is random. This allows for multiple runs with different automatically generated social networks and mobile scenarios. However, the current implementation can be easily modified and replaced by a custom initialization of the simulation settings.

- **Movement determined by pre-defined trails** Many existing mobility models are based on the definition of *trails* or *paths* that are used to define the movements of the mobile nodes in the simulation scenarios. Examples are the Manhattan model [5] or other models used to simulate protocols and systems for vehicular ad hoc networks [34]. In these models hosts move between different locations following precise paths that represent roads or motorways. We plan to study the effects of the introduction of pre-defined trails in our model, in particular to characterize the movements between different communities (i.e., squares of the grid) with better accuracy.

- **Presence of obstacles** As discussed in the previous subsection, in [17] Jardosh et alii propose a modified version of the Random Mobility Model that allows for the insertion of obstacles in the simulation space. The definition of obstacles can also be easily integrated in our mobility model.

Finally, we plan to study the connectivity of the generated mobile networks also in relation to the social networks given in input using results from graph theory studies [3].

6. CONCLUSIONS

We have presented a new mobility model based on social network theory and predicated on the assumption that mobility patterns are driven by the fact that devices are carried by humans and that the movements are strongly affected by the relationships between them.

The paper has described the generation of the mobility model, its implementation and an evaluation based on the comparison between our approach, existing random mobility models and real movement traces. We have shown that our mobility model generates traces that present characteristics similar to real ones, in terms of inter-contacts time and contacts duration. We have also compared the performance in terms of delivery ratio of the AODV and DSR protocols using the Random Way-Point model and our Community based model.

Acknowledgments

We acknowledge the support of Intel Research Laboratory in Cambridge. They provided us with the traces we used for the evaluation of our model. We are also grateful to Paolo Costa, Ilias Leontiadis and George Roussos for their comments on an early draft of this paper.

7. REFERENCES

[1] R. Albert and A.-L. Barabasi. Statistical mechanics of complex networks. *Review of Modern Physics*, 74:47–97, 2002.

[2] M. Balazinska and P. Castro. Characterizing Mobility and Network Usage in a Corporate Wireless Local-Area Network. In *Proceedings of ACM MobiSys'03*, pages 303–316, 2003.

[3] B. Bollobas. *Random Graphs.* Cambridge University Press, Second edition, 2001.

[4] J.-Y. L. Boudec and M. Vojnovic. Perfect simulation and stationarity of a class of mobility models. In *Proceedings of IEEE INFOCOM'05*, pages 72–79, March 2005.

[5] T. Camp, J. Boleng, and V. Davies. A survey of mobility models for ad hoc network research. *Wireless Communication and Mobile Computing Special Issue on Mobile Ad Hoc Networking: Research, Trends and Applications*, 2(5):483–502, 2002.

[6] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Pocket Switched Networks: Real-world mobility and its consquence for opportunistic forwarding. Technical Report UCAM-CL-TR-617, University of Cambridge, Computer Laboratory, February 2005.

[7] A. Einstein. *Investigations on the Theory of the Brownian Movement.* Dover Publications, 1956.

[8] P. Erdos and A. Renyi. On random graphs. *Publicationes Mathematicae*, 6:290–297, 1959.

[9] L. C. Freeman. Some antecedents of social network analysis. *Connections*, 19:39–42, 1996.

[10] L. C. Freeman. A Set of Measures of Centrality based on Betweenness. *Sociometry*, pages 35–41, 40.

[11] I. Glauche, W. Krause, R. Sollacher, and M. Greiner. Continuum percolation of wireless ad hoc communication networks. *Physica A*, 325:577–600, 2003.

[12] T. Henderson, D. Kotz, and I. Abyzov. The changing usage of a mature campus-wide wireless network. In *Proceedings of ACM MobiCom'04*, pages 187–201, 2004.

[13] K. Hermann. Modeling the sociological aspect of mobility in ad hoc networks. In *Proceedings of MSWiM'03*, pages 128–129, San Diego, California, USA, September 2003.

[14] X. Hong, M. Gerla, G. Pei, and C.-C. Chiang. A group mobility model for ad hoc networks. In *Proceedings of MSWiM'99*, pages 53–60, 1999.

[15] W. Hsu, K. Merchant, H. Shu, C. Hsu, and A. Helmy. Weighted Waypoint Mobility Model and its Impact on Ad Hoc Networks. *ACM Mobile Computer Communications Review (MC2R)*, pages 59–63, January 2005.

[16] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot. Pockets Switched Networks and Human Mobility in Conference Environments. In *Proceedings of ACM SIGCOMM'05 Workshops*, pages 244–251, August 2005.

[17] A. Jardosh, E. M. Belding-Royer, K. C. Almeroth, and S. Suri. Real world Environment Models for Mobile Ad hoc Networks. *IEEE Journal on Special Areas in Communications - Special Issue on Wireless Ad hoc Networks*, 23(3), March 2005.

[18] D. Johnson and D. Maltz. Dynamic source routing in ad hoc wireless networks. In T. Imelinsky and H. Korth, editors, *Mobile Computing*, volume 353, pages 153–181. Kluwer Academic Publishers, 1996.

[19] D. B. Johnson and D. A. Maltz. Dynamic source routing in ad hoc wireless network. In T. Imielinski and H. Korth, editors, *Mobile Computing*, chapter 5, pages 153–181. Kluwer Academics Publisher, 1996.

[20] K. A. Khaled Harras and E. Belding-Royer. Delay Tolerant Mobile Networks (DTMNs): Controlled Flooding Schemes in Sparse Mobile Networks. In *IFIP Networking 2005*, pages 1180–1192, May 2005.
[21] D. Kotz and T. Henderson. CRAWDAD: A Community Resource for Archiving Wireless Data at Dartmouth. *IEEE Pervasive Computing*, 4(4):12–14, October-December 2005.
[22] K. Maeda, K. Sato, K. Konishi, A. Yamasaki, A. Uchiyama, H. Yamaguchi, K. Yasumotoy, and T. Higashino. Getting urban pedestrian flow from simple observation: Realistic mobility generation in wireless network simulation. In *Proceedings of MSWiM'05*, pages 151–158, September 2005.
[23] S. McCanne and S. Floyd. ns-2 network simulator. http://www.isi.edu/nsnam/ns/.
[24] M. McNett and G. M. Voelker. Access and mobility of wireless pda user. *Mobile Computing Communications Review*, 9(2):40–55, April 2005.
[25] M. Musolesi, S. Hailes, and C. Mascolo. An Ad Hoc Mobility Model Founded on Social Network Theory. In *Proceedings of MSWiM'04*, pages 20–24. ACM Press, October 2004.
[26] M. Musolesi, S. Hailes, and C. Mascolo. Adaptive routing for intermittently connected mobile ad hoc networks. In *Proceedings of WoWMoM 2005. Taormina, Italy.* IEEE press, June 2005.
[27] M. E. J. Newman. Scientific Collaboration Networks: II. Shortest Paths, Weighted Networks and Centrality. *Physical Review E*, 64, 2001.
[28] M. E. J. Newman. The Structure of Scientific Collaboration Networks. In *Proceedings of the National Academy of Science*, volume 98, pages 404–409, 2001.
[29] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 19(1):1–42, 2003.
[30] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69, February 2004.
[31] M. E. J. Newman and J. Park. Why Social Networks are Different from Other Types of Networks. *Physical Review E*, 68, 2003.
[32] C. E. Perkins and E. M. Royer. Ad hoc on-demand distance vector routing. In *Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications*, pages 90–100, February 1999.
[33] A. Rapoport. Contribution to the theory of random and biased networks. *Bulletin of Mathematical Biophysics*, (19):257–277, 1957.
[34] A. K. Saha and D. B. Johnson. Modeling Mobility for Vehicular Ad Hoc Networks. In *Proceedings of VANET'04*, pages 91–92, October 2004.
[35] N. Sarshar and R. Chowdhury. Scale-free and stable structures in complex ad hoc networks. *Physical Review E 026101*, 2003.
[36] J. Scott. *Social Networks Analysis: A Handbook*. Sage Publications, London, United Kingdom, second edition, 2000.
[37] G. Sharma and R. R. Mazumdar. Scaling laws for capacity and delay in wireless ad hoc networks with random mobility. In *IEEE International Conference on Communications (ICC'04)*, pages 3869– 3873, June 2004.
[38] D. Stauffer and A. Aharony. *Introduction to Percolation Theory*. Taylor and Francis, 1992.
[39] C. Tuduce and T. Gross. A Mobility Model Based on WLAN Traces and its Validation. In *Proceedings of INFOCOM'05*, pages 19–24, March 2005.
[40] A. Vasquez, R. Pastor-Satorras, and A. Vespignani. Large-scale topological and dynamical properties of the internet. *Physical Review E*, 67, 2003.
[41] D. J. Watts. *Small Worlds The Dynamics of Networks between Order and Randomness*. Princeton Studies on Complexity. Princeton University Press, 1999.
[42] X. Zeng, R. Bagrodia, and M. Gerla. Glomosim: A library for parallel simulation of large-scale wireless networks. In *Workshop on Parallel and Distributed Simulation*, pages 154–161, 1998.

APPENDIX

A. BETWEENNESS OF A NODE

The betweenness of a node in a graph is defined as the total number of shortest paths between pairs of nodes that pass through it [10]. The steps of the algorithm proposed by Newman are the following:

- For each vertex j, the shortest path reaching it from every vertices in the network is calculated by using the following algorithm:
  1. Assign to vertex j a distance d equal to 0;
  2. For each vertex i with assigned distance i, follow each attached edge to the vertex k and, if k has not already been assigned a distance, assign it distance $d+1$. i is declared a predecessor of k;
  3. Increment the value of the distance d by 1;
  4. Repeat from Step 2 until there are no unassigned vertices left.
- A variable b_i is assigned to every vertex of the network with initial value 1.
- Considering the vertices i sorted from the farthest to the nearest, the value of b_i is added to the corresponding variable on the predecessor vertex of i. If i has more than one predecessor, then b_i is equally subdivided among them.
- The resulting values of b_i represent the number of geodesic paths to vertex j that run through each vertex of the lattice. To calculate the betweenness for all paths, the b_i are added to a variable b'_i that is maintained for each variable. The entire calculation is repeated for all the vertices. The final value of b'_i represents the betweenness of vertex i.

The algorithm is presented and discussed in more details in [30].

B. MODULARITY

Let us consider a division of a network into k communities. We define a $k \times k$ symmetric matrix $\mathbf{e}$ the elements e_{ij} of which are the proportion of all edges in the network that link vertices in community i to vertices in community j. The trace of this matrix $Tr\ \mathbf{e} = \sum_i e_{ii}$ gives the proportion of edges in the network that connect vertices in the same community. A good division into communities should have a high value of this trace, meaning that a good portion of the edges of the network is of edges "inside" a community. However, this is not sufficient to judge the quality of the division. In fact, the case of one single community corresponds to the case of $Tr\ \mathbf{e} = 1$.

Therefore, we define the rows sums $a_i = \sum_j e_{ij}$ which represents the proportion of edges that connect to vertices in community i. Thus, it is possible to define the *modularity* Q of a network division as follows:

$$Q = \sum_i (e_{ii} - a_i^2) = Tr\ \mathbf{e} - ||\mathbf{e}^2||$$

where $||\mathbf{e}^2||$ is the sum of all the elements of the matrix $\mathbf{e}^2$. More details can be found in [30].