Small World in Motion (SWIM): Modeling Communities in Ad-Hoc Mobile Networking

Sokol Kosta, Alessandro Mei, and Julinda Stefa
Department of Computer Science
Sapienza University of Rome, Italy
Email: {kosta, mei, stefa}@di.uniroma1.it

Abstract—The complexity of social mobile networks (networks of devices carried by humans–e.g. sensors or PDAs–and communicating with short-range wireless technology) makes protocol evaluation hard. A simple and efficient mobility model such as SWIM reflects correctly kernel properties of human movement and, at the same time, allows to evaluate accurately protocols in this context. In this paper we investigate the properties of SWIM, in particular how SWIM is able to generate social behavior among the nodes and how SWIM is able to model networks with a power-law exponential decay dichotomy of inter contact time and with complex sub-structures (communities) as the ones observed in the real data traces. We simulate three real scenarios and compare the synthetic data with real world data in terms of inter-contact, contact duration, number of contacts, and presence and structure of communities among nodes and find out a very good matching. By comparing the performance of BUBBLE, a community-based forwarding protocol for social mobile networks, on both real and synthetic data traces, we show that SWIM not only is able to extrapolate key properties of human mobility but also is very accurate in predicting performance of protocols based on social human sub-structures.

Index Terms—Mobility model, small world, simulations, forwarding protocols in mobile networks.

I. INTRODUCTION

Mobile ad-hoc networking has presented many challenges to the research community, especially in designing suitable, efficient, and well performing protocols. The practical analysis and validation of such protocols often depends on synthetic data, generated by some mobility model. The model has the goal of simulating real life scenarios [1] that can be used to tune networking protocols and to evaluate their performance. Till a few years ago, the model of choice in academic research was the Random Way-Point model (RWP) [2], simple and very efficient to use in simulations.

Recently, with the aim of understanding human mobility [3], [4], [5], [6], [7], many researchers have performed real-life experiments by distributing wireless devices to people. From the data gathered during the experiments, they have observed the typical distribution of metrics such as inter-contact time (time interval between two successive contacts of the same people) and contact duration. Inter-contact time, which corresponds to how often people see each other, characterizes

the opportunities of packet forwarding between nodes. Contact duration, which limits the duration of each meeting between people in mobile networks, limits the amount of data that can be transferred. In [4], [5], the authors show that the distribution of inter-contact time is a power-law. Later, in [6], it has been observed that the distribution of inter-contact time is best described as a power law in a first interval on the time scale (12 hours or 1 day, in the experiments under analysis), then truncated by an exponential cut-off. Conversely, [8] proves that RWP yields exponential inter-contact time distribution. Therefore, it has been established clearly that models like RWP are not good to simulate human mobility, raising the need of new, more realistic mobility models for mobile ad-hoc networking.

In a recent work [9] the authors present small world in motion (SWIM), a simple mobility model that generates small worlds. The model is very simple to implement and very efficient in simulations. The mobility pattern of the nodes is based on a simple intuition on human mobility: People go more often to places not very far from their home and where they can meet a lot of other people. By implementing this simple rule, SWIM is able to raise social behavior among nodes, which we believe to be the base of human mobility in real life. The authors validate the model using real traces and compared the distribution of inter-contact time, contact duration and number of contact distributions between nodes, showing that synthetic data that SWIM generates match very well real data traces. They show experimentally that SWIM generates traces whose inter-contacts distribution's head follows a power law, and proved mathematically the exponential decaying of the tail of such distribution. Furthermore, they show that SWIM can predict well the performance of simple forwarding protocols. They compare the performance of two forwarding protocols—Epidemic Forwarding [10] and (a simplified version of) Delegation Forwarding [11]—on both real traces and synthetic traces generated with SWIM. The performance of the two protocols on the synthetic traces accurately approximates their performance on real traces, supporting the claim that SWIM is a good model for human mobility.

In this paper we investigate the way SWIM generates social behavior among nodes and uses it to yield a social mobile network: A network whose nodes indeed form social clusters (communities). We describe how to use SWIM to create different target scenarios—conference, university campus, city—

that differ from the number of nodes involved, the shape, the number and cardinality of the communities therein, the slope of the power-law head of the inter-contact distribution, and more. Moreover, we show how to obtain networks with a target power-law exponential decay dichotomy of inter contact time, and with a given membership of nodes to communities as the ones observed in the real data traces. We simulate three real scenarios and compare the synthetic data with the real ones in terms of inter-contact, contact duration, number of contacts, and presence of communities among nodes and find out a very good matching. With the goal to check how more sophisticated protocols—the ones that are based on community sub-structures present on the network—perform on SWIM, we use BUBBLE [12], the state of the art of the community-based forwarding protocols for social mobile networks. We compare BUBBLE's performance on both real and synthetic data traces and show that SWIM not only is able to extrapolate key properties of human mobility but also is very accurate in predicting performance of protocols based on complex social human sub-structures.

The rest of the paper is organized as follows: Section II briefly reports on current work in the field; in Section III we briefly review the way SWIM operates, whereas in Section IV we show how to use SWIM in order to generate a target scenario. In Section V we concretize the findings of Section IV by comparing three different real world traces with the respective SWIM traces, in terms of creation of complex sub-structures (communities), and distribution of inter-contacts, contact duration and number of contacts among the nodes. Later on, in Section VI we compare the performance of BUBBLE on the real traces with the SWIM-generated traces, and conclude with Section VII.

II. Related work

The problem of defining a mobility model for human mobility is felt as an important one in the community and in the literature. In the last few years there have been a considerable number of papers on this topic. The mobility model recently presented in [13] generates movement traces using a model which is similar to a random walk, except that the flight lengths and the pause times in destinations are generated based on Levy Walks—with power law distribution. In the past, Levy Walks have been shown to approximate well the movements of animals. The model produces inter-contact time distributions similar to real world traces. However, since every node moves independently, the model does not generate any social structure between the nodes. In [14], the authors present a mobility model based on social network theory which takes in input a social network and discuss the community patterns and groups distribution in geographical terms. They validate their synthetic data with real traces and show a good matching between them.

The work in [15] presents a new mobility model for clustered networks. Moreover, a closed-form expression for the stationary distribution of node position is given. The model captures the phenomenon of emerging clusters, observed in real partitioned networks, and correlation between the spatial speed distribution and the cluster formation.

In [16], the authors present a mobility model that simulates the every day life of people that go to their work-places in the morning, spend their day at work and go back to their homes at evenings. Each one of this scenarios is a simulation per se. The synthetic data they generate match well the distribution of inter-contact time and contact durations of real traces.

Recently, in [17] the authors proposed the SLAW mobility model, which is a modification of the Levy-walk based model, where the human waypoints are modeled as fractals. The model seems to match the inter-contact distribution of some of the real traces, and to predict quite accurately performance of simple forwarding protocols. Yet, no results are presented in terms of contact and contact number distributions and in the structure in communities of the resulting network, and it seems to be hard to be used in theoretical analysis.

In a very recent work, Barabasi et al. [18] study the trajectory of a very large (100,000) number of anonymized mobile phone users whose position is tracked for a six-months period. They observe that human trajectories show a high degree of temporal and spatial regularity, each individual being characterized by a time independent characteristic travel distance and a significant probability to return to a few highly frequented locations. They also show that the probability density function of individual travel distances are heavy tailed and also are different for different groups of users and similar inside each group. Furthermore, they plot also the frequency of visiting different locations and show that it is well approximated by a power law. All these observations are in contrast with the random trajectories predicted by Levy flight and random walk models, and support the intuition behind SWIM [9], a mobility model recently proposed. This model is very simple to simulate, matches well the statistical properties of real traces gained from experiments done with real devices, and in the same time, predicts well performance of simple forwarding protocols for ad-hoc wireless networks where devices are carried by mobile humans.

III. Small World in Motion: A brief overview

A good mobility model should be simple, and, simultaneously, predict well the performance of networking protocols on real mobile networks. Simplicity helps understand and distill fundamental ingredients of "human" mobility, eases implementation of the model and makes it a useful tool in supporting theoretical work. A good model should also generate traces with the same statistical properties that real traces have. Statistical distribution of inter-contact time and number of contacts, among others, are useful to characterize the behavior of a mobile network. At the same time, the model should be accurate in predicting protocol performance on real networks: If a protocol performs well (or bad) in the model, it should also perform well (or bad) in the real network, as accurately as possible.

SWIM [9] is the first model to meet all these requirements simultaneously. It is based on simple observations on our

everyday life: We usually prefer *popular* and *nearby* places to "not frequented" and distant ones. The best supermarket or the most popular restaurant that are also not far from where we live, for example. Moreover, usually there are just a few places where a person spends a long period of time (for example home and work office or school), whereas there are lots of places where she stays less, e.g. post office, bank, cafeteria, etc. Lastly, our movement speed depends on the distance: We walk to the buss stop, ride to the grocery store, and, fly to get to another continent. These are the basic intuitions SWIM is built upon: Trading-off proximity and popularity, adequate distribution of waiting time and speed proportional to the distance covered. More in detail, to each node is assigned a so called *home*, which is a randomly and uniformly chosen point over the network area. Then, the node itself assigns to each possible destination a *weight* that grows with the popularity of the place and decreases with the distance from home. The weight represents the probability for the node to chose that place as its next destination. The network area is divided into many small contiguous cells that represent possible destinations. Each cell has a squared area, and its size depends on the transmitting range of the nodes: The size of the cell is such that its diagonal is equal to the transmitting radius of the nodes—so that a node within a given cell is able to communicate with all other nodes within that same cell. Each node builds a *map* of the network area, also calculates a weight for each cell in the map, with the following formula:

$$w(C) = \alpha \cdot distance(h_A, C) + (1 - \alpha) \cdot seen(C). \quad (1)$$

where A is a network node, h_A is the home point of A, C is the given cell, $distance(h_A, C)$ is a function that decays as a power law as the distance between node A and cell C increases, α is a constant in $[0; 1]$, and $seen(C)$ is the fraction of nodes A saw in C the last time it was there. The value $seen(C)$ is initially zero, and it is updated each time A reaches C. The weight represents the probability for the node to chose that place as its next destination.

Once a node has chosen its next destination, it starts moving towards it following a straight line and with a constant speed that equals the movement distance. When reaching destination the node decides how long to remain there. One of the key observations is that in real life a person usually stays for a long time only in a few places, whereas there are many places where he spends a short period of time. Therefore, the distribution of the waiting time follows a bounded power law.

The particular choices of the distance and seen function affect the other parameters, especially α. The distance function we use is the one proposed in [9]:

$$distance(x, C) = \frac{1}{(1 + k||x - y||)^2}, \quad (2)$$

where x is the position of the home of the current node, and y is the position of the center of cell C. Moreover, we use as $seen(C)$ function the fraction of the nodes seen in cell C.

In [9] the parameter k is set to $k = .05$, which accounts for the small size of the experiment area of the scenarios

Experimental data set	Cambridge	Infocom 05	Infocom 06
Device	iMote	iMote	iMote
Network type	Bluetooth	Bluetooth	Bluetooth
Duration (days)	11	3	3
Granularity (sec)	600	120	120
Devices number	36	41	78
Internal contacts number	10,873	22,459	191,336
Average Contacts/pair/day	0.345	4.6	6.7

TABLE I
THE THREE EXPERIMENTAL DATA SETS.

simulated. We have though noticed that in SWIM there are two different ways of scaling: Either by changing the parameter k or by changing the communication range of the nodes. Indeed, the latter determines the size of network area. Moreover, the parameter k in the distance function affects the probability of choosing nearby cells. As can be noticed in Equation 2, bigger values for k yield bigger values for the $distance(x, C)$ function, by giving more importance to the distance component of the formula in Equation 1. In this work we aim to generate simulated scenarios where node movements (contacts) are highly related to their home-points on the network area. Thus, we use a bigger value for the k parameter (we set $k = 5$). Then, differently from [9], we get the scaling effect by decreasing the communication range: We set it to be 0.04.

Since the weight that a node assigns to a place represents the probability that the node chooses it as its next destination, the value of α has a strong effect on the node's decisions—the larger is α, the more the node will tend to go to places near its home. The smaller is α, the more the node will tend to go to "popular" places.

Finally, SWIM takes in input the following parameters:

- n: the number of nodes in the network;
- r: the transmitting radius of the nodes;
- the simulation time in seconds;
- coefficient α that appears in Equation 1;
- distribution of the home-points;
- the distribution of the waiting time at destination (slope of the power-law distribution and upper bound).

The output of the simulator is a text file containing records on each main event occurrence (meet, depart, start, and, finish).

IV. SWIM MADE SIMPLE TO USE

In a number of research works ([4], [12], and others) it is shown that the social relationships among the nodes in an ad-hoc mobile network can efficiently and correctly be detected by using the k-clique algorithm [19]. Moreover, [20], [21], [22], [23], [18] show that the distribution of social community members in cities follows a power law distribution. In this section we show that SWIM generates synthetic networks that have all these properties.

The α parameter of Equation 1 determines whether nodes prefer popular sites ($\alpha \longrightarrow 0$) or nearby ones ($\alpha \longrightarrow 1$). Thus, if the distribution of the home-points is uniform, small values of α generates a preferential visiting of sites (sites that are popular are visited by more and more nodes). Conversely, big values of α give more preference to sites that are nearby
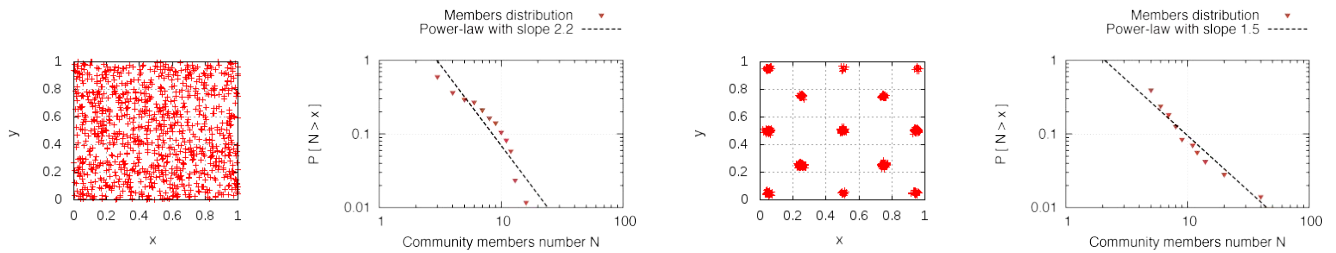
(a) Home-points uniformly distributed.

(b) Community membership distribution, $\alpha = 0$, home-points uniform.

(c) Home-points distributed with preferential attachment.

(d) Community membership distribution, $\alpha = 1$, home-points with preferential attachment.

Fig. 1. Distributions of home-points and respective community membership number after simulation. Waiting time slope 1.45, waiting time bound 1 day, network of 1000 nodes, radius 0.01.

nodes' home-points. Thus, node agglomerations (communities) with a power law distribution of members should be obtained through SWIM in the two following different ways: Either small values of α, and a uniform distribution of home-points over the network area, or big values of α and a distribution of home-points that follows a power-law. However, the properties of the structure in communities are different. We empirically prove these claims by conducting the following experiments. In the first one, we set the α value to 0, and uniformly distribute 1000 home-points that correspond to 1000 nodes over the network area. Then we let the simulation run, and, at the end, we detect the communities generated using the k-clique algorithm [19], and plot the cumulative distribution of the membership number in a log-log scale. The result is exactly what we expected: The community membership distribution is very close to a power law (see Figures 1(a)) and 1(b)). For the second method we operate in the following way: We first distribute 13 nodes' home-points in 13 different sites (we want to generate a network with 13 communities), trying to keep the mutual distance as big as possible. Then, we let the other nodes choose their home-point site with a preferential attachment (that is, in the process the probability of choosing a cell as home-point depends linearly with the number of home-points in the cell itself in that moment). The resulting initial distribution of home-points is shown in Figure 1(c). Then we run the simulation using $\alpha = 1$, detect the communities and plot the cumulative distribution of members number. Even in this case the result is what we expected: The distribution of the number of members of the communities detected by the k-clique algorithm [19] is very close to a power law (see Figure 1(d)).

Both methods ($\alpha = 0$ uniform distribution of home-points, and, $\alpha = 1$ preferential attachment distribution of home-points) raise community agglomeration among the nodes and, moreover, the community membership number follows a power law distribution. In the first case there is a higher mixing of nodes in the network. Indeed, when $\alpha = 0$ nodes prefer popular sites, no matter where in the network these sites are. In the second case, the nodes chose only in base of the distance between home-points and sites. This generates a "neighborhood effect": nodes that are neighbors chose to go to similar sites, and thus, tend to meet more frequently,

and thus, to belong to the same community. Indeed, when we calculate the correlation between the distance among home points and the community membership, we have the following results: For $\alpha = 0$ and uniform home-point distribution the correlation is 0, whereas, for $\alpha = 1$ and preferential attachment distribution of home-points the correlation is -0.53. This gives an important hint on how to generate traces with a given community structure among nodes: Use large values of α and distribute the home-points in such a way that nodes that should be in the same community are neighbors. We will use this hint in the next section, where we will show how SWIM simulates real-scenario community structures.

As we already mentioned, the distribution of the inter-contact times—intervals between two consecutive meetings of the same couple of nodes—is important, not only because it characterizes statistically the network, but also because it impacts the design of opportunistic forwarding algorithms [7], [5]. As shown in [5], the interval within which inter-contacts follow a power law distribution, and the slope of this distribution strongly affects the performance of forwarding protocols. Thus, a model that is able to yield, by simple tuning of a few parameters, a mobile trace with a required inter-contact distribution in terms of slope and power-law like head interval, becomes more than valuable. This is SWIM's case: From our experimental testings we observe that the inter-contact distribution can be controlled by the setting of the waiting time slope and upper bound. The upper bound determines the interval within which inter-contacts follow a power law distribution, whereas the slope determines the slope of this distribution. For the sake of space here we show experimental results related to the upper bound of the waiting time. As you can see from Figure 2 the waiting-time upper bound determines the exact interval within which inter-contacts follow a power-law distribution. For values beyond the bounds, in all cases the distribution decays exponentially. This also can be used to simulate a mobile scenario with a given power-law exponential dichotomy of inter-contacts among network members. In the next section we will see how SWIM can be tuned to simulate real scenarios in terms of communities structuring, community membership, and distribution of statistical properties.
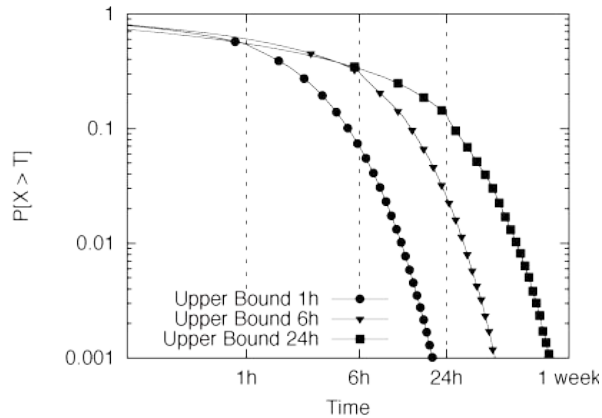
Fig. 2. Distribution of inter-contacts for different waiting-time upper bounds. Waiting time slope 1.45, network of 1000 nodes, radius 0.01.

Simualted Scenario	Cambridge	Infocom 05	Infocom 06
Communication range	0.04	0.04	0.04
Duration (days)	11	3	3
Devices number	36 mobile	41	78 mobile
Value of α	.95	.75	.75
Waiting time slope	1.45	1.45	1.45
Waiting time bound	24h	12h	12h

TABLE II
THE THREE SIMULATED DATASETS.

V. SWIM COMMUNITIES VS REAL COMMUNITIES

Here we show how to get a simulation of a mobile network whose communities are the approximately the same of a given real scenario. For this, we simulate 3 real scenarios, whose traces were gathered during experiments done with real devices carried by people. We will refer to these traces as *Cambridge*, *Infocom 05* and *Infocom 06*. Characteristics of these data sets such as inter-contact and contact distribution have been observed in several previous works [4], [24], [5]. In Table I we show the statistical details of each trace, whereas here below we describe each experiment scenario and provide details on the communities present therein.

- In *Cambridge* [25] the authors used Intel iMotes to collect the data. The iMotes were distributed to two groups of students (Year1 and Year2) of the University of Cambridge and were programmed to log contacts of all visible mobile devices. Also, a number of stationary nodes were deployed in various locations around the city of Cambridge UK. The data of the stationary iMotes will not be used in this paper. The number of mobile devices used is 36 (plus 18 stationary devices). This data set covers 11 days.
- In *Infocom 05* [26] the same devices as in *Cambridge* were distributed to students attending the Infocom 2005 student workshop. Participants belong to different social communities (depending on their country of origin, research topic, etc.)The number of devices is 41. This experiment covers approximately 3 days.
- In *Infocom 06* the scenario was similar to Infocom 05

except that the scale is larger, with 78 participants. Participants were selected so that 34 out of 78 form 4 subgroups by academic affiliation: ParisA with 10 participants, ParisB with 4 participants, Lausanne 5 participants, and, Barcelona 15 participants. In addition, 20 long range iMotes were deployed at several places in the conference site to act as access points. However, the data from these fixed nodes is not used in this paper.

The authors in [27] use the *k*-clique algorithm [19] to uncover communities from the traces of the real experiments. In the Cambridge scenario, they detect two main communities of 11 members each, that correspond respectively to the students of the first and the second year. In the Infocom 06 scenario they observe that mostly of the participants with the same academic affiliation (Paris A, Paris B, Lausanne and Barcelona) do belong to the same communities detected by the *k*-clique algorithm. As for the Infocom 05 scenario, the authors in [27] do not give details on the community detection. However, they give some information on the participants: There are four groups based on academic affiliation, of respectively 10, 6, 4, 4 members each.

The next step of our study is to generate with SWIM each of the real scenarios, and to see how the desired social-grouping of the nodes can be gained. Let us start with Cambridge 05. There are 36 students involved, grouped by academic year in two groups: Year1 and Year2. As we mentioned, in the real trace only 11 students per group form a community. Hence, we "associate" 22 nodes to each of the community-forming students in the real experiment. To each group we assign a "center point" in the network area: $p_1 = (.05; .05)$ and $p_2 = (.95; .95)$ (for respectively groups Year1 and Year2). The members of each group is given a home point obtained by perturbing the respective center point with a Gaussian distribution of standard deviation of 0.01. The remaining 14 nodes is assigned a home point obtained with a uniform distribution over the network area.

Unlike the Cambridge scenario, in the Infocom 05 scenario we have no exact information on the communities detected, but only on the initial affiliation of some of the members. Hence, we base our experiments on this information: 4 groups of respectively 10, 6, 4 and 4 members each. Therefore, to simulate this scenario we divide our nodes in 4 groups of as much members as in the real case. For each group we assign a central point as follows: $p_1 = (.95; .95)$ for group 1, $p_2 = (95; .05)$ for group 2, $p_3 = (.05; .95)$ for group 3, $p_4 = (.05; .05)$ for group 4. The members of each group is given a home point obtained by perturbing the respective center point with a Gaussian distribution of standard deviation of 0.01. The remaining nodes are assigned home-points chosen uniformly and randomly over the network area.

Infocom 06 is similar to Infocom 05 but with exact community member information: four communities (ParisA, ParisB, Lausanne and Barcelona) of respectively 10, 4, 5, and, 15 members each. Therefore, to simulate this scenario we divide 34 nodes in 4 groups of as much members as in the real case. For each group we assign a central point

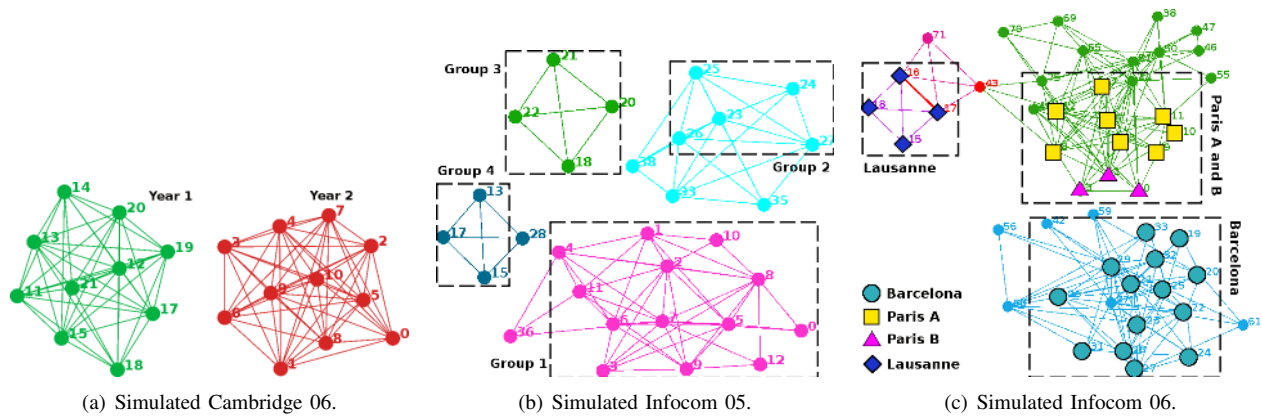(a) Simulated Cambridge 06.   (b) Simulated Infocom 05.   (c) Simulated Infocom 06.

Fig. 3.   Communities detected in the synthetic traces.

as follows: $p_1 = (.01; .01)$ for Paris A, $p_2 = (.013; .013)$ for Paris B, $p_3 = (.95; .01)$ for Lausanne, and, $p_4 = (.5; .95)$ for Barcelona. Note that the members of the two Paris groups are initially placed close, in order to simulate social connection among them. The members of each group is given a home point obtained by perturbing the respective center point with a Gaussian distribution of standard deviation of 0.01. The remaining nodes are assigned home-points chosen uniformly and randomly over the network area.

The rest of the simulation input is showed in Table II. In particular, the choice of the α value is done based on the grade of relationship people typically have in conferences vs university. We follow the suggestions of the authors in [9] and use the same values of α for each scenario: .95 and .75 for respectively the Cambridge and the Infocom's scenarios. Also, the choice of the waiting time bound is done based on the real traces inter-contact time distribution's head. In the Cambridge case it follows a power law for up to 24 hours, whereas in both Infocom scenarios it decays as power law for up to 12 hours.

In Figure 3 are shown the communities detected from the synthetic traces. As can be seen, in each simulated scenario the community-detection reflects very well the real scenario: Nodes whose affiliation was emulated by assigning adjacent home-points result being members of the same community detected after the simulation. This means that SWIM preserves initial "social relationships" among nodes in the same way as a real social mobile network does.

In [9] the authors advocate SWIM's merits by plotting the cumulative distributions of inter-contacts, contact duration and contact number among nodes. The matching between the simulated and the real distributions is very good. In their work they use a uniform distribution for the nodes' home-points over the network area. Here we changed this distribution in order to be able to obtain a desired target social scenario. Therefore, we still need to compare inter- contacts, contact-number and contact-duration of the so modified SWIM traces with the real ones. We present the relative results in Figures 4, 5, and 6.

As can be seen from the figures, even by changing the way with which the home-points are assigned to nodes, SWIM still yields synthetic traces with similar statistical properties to the

real ones. Moreover, as it can be seen from Figures 4(a), 5(a), and 6(a), the head's length of the inter-contacts distributions is determined by the waiting-time upper bound used in each of the simulations: 24h in the Cambridge case, and 12h in both Infocom's cases. In all of the three experiments, SWIM proves to be an excellent way to generate synthetic traces that approximate real traces in both community raising and statistical properties. It is particularly interesting that the same choice of parameters gets goods results for all the metrics under consideration at the same time.

VI. PERFORMANCE OF COMPLEX PROTOCOLS ON SWIM AND REAL TRACES

Generating synthetic traces that look real from a statistical point of view and that have a community structure similar to the one we can experience in the real experiments is definitely important, but not enough. Our model has the ultimate goal of being used to validate networking protocols. One important parameter is how nodes in the same community interact and, even more importantly, how communities overlap and interplay. To evaluate the quality of SWIM to this respect, we believe that the best is to see how it performs when doing the job this model was created for—to give reliable information on the performance of non-trivial, social based protocols. SWIM has already been shown to be adequate in predicting performance of Epidemic [10] and a simplified version of Delegation Forwarding [11]. Both these protocols are based on flooding techniques, and do not make use of any social component of the network. Since we want to test the adequacy of SWIM in predicting the performance of more complex forwarding protocols, we use BUBBLE [12], whose forwarding technique totally relies upon relationship among network nodes, their belonging to communities, and upon the overlapping nature of communities. BUBBLE works as follows: Each node is assigned a *global rank*, that determines the centrality of the node in the network, and a *local rank* for each community it belongs to. When a message is generated for a given destination, say D, then the message is first forwarded towards nodes with higher and higher *global rank*, till it reaches a relay that belongs to $D's$ community. Then, it
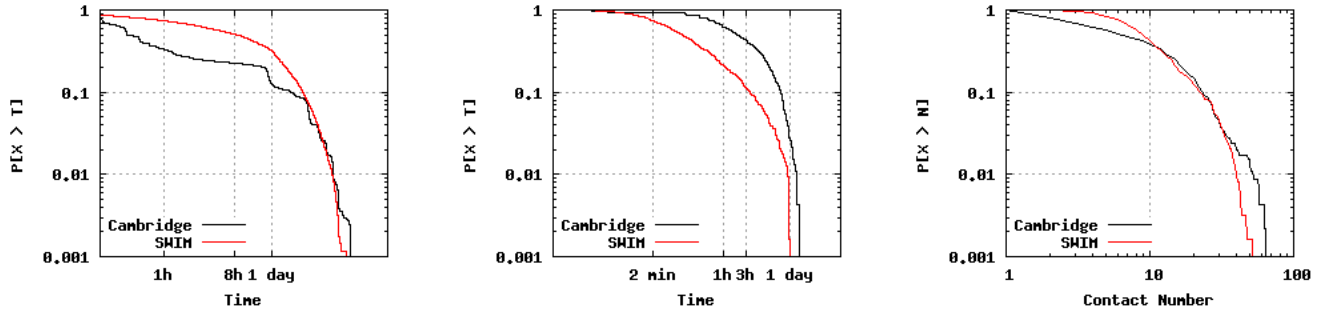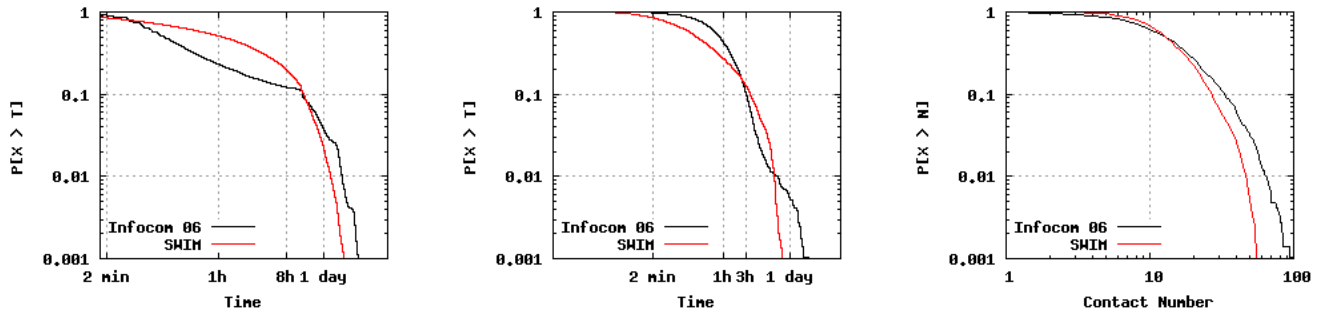
(a) Distribution of the inter-contact time in Cambridge and in SWIM.

(b) Distribution of the contact duration for each pair of nodes in Cambridge and in SWIM.

(c) Distribution of the number of contacts for each pair of nodes in Cambridge and in SWIM.

Fig. 4. SWIM and Cambridge.
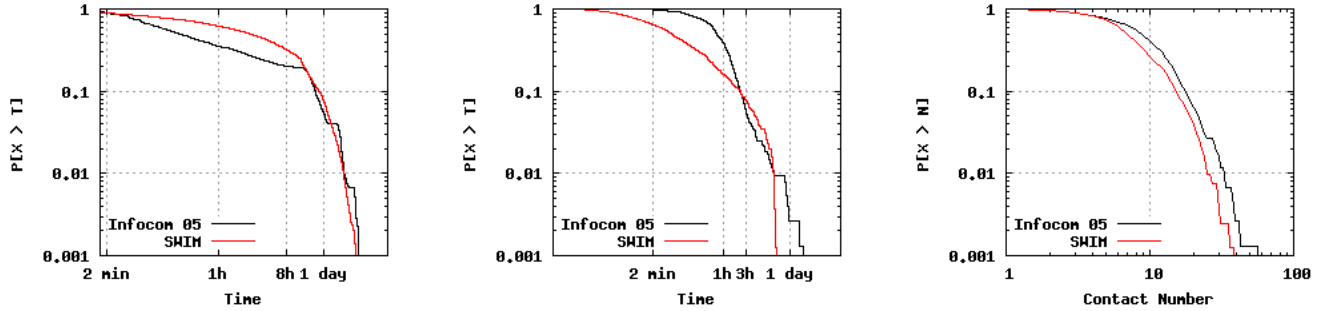
(a) Distribution of the inter-contact time in Infocom 05 and in SWIM.

(b) Distribution of the contact duration for each pair of nodes in Infocom 05 and in SWIM.

(c) Distribution of the number of contacts for each pair of nodes in Infocom 05 and in SWIM.

Fig. 5. SWIM and Infocom 05.

(a) Distribution of the inter-contact time in Infocom 06 and in SWIM.

(b) Distribution of the contact duration for each pair of nodes in Infocom 06 and in SWIM.

(c) Distribution of the number of contacts for each pair of nodes in Infocom 06 and in SWIM.

Fig. 6. SWIM and Infocom 06.

is forwarded towards nodes with higher and higher *local rank* within $D's$ community, till it reaches D.

We computed the local/global ranks of nodes exactly as the authors in [12]. For the global rank we generated a uniform traffic set of 1000 messages, forwarded them with flooding technique, and, counted for each node, the number of times it belonged to the path of a delivered message. Similarly, for the local ranks, the messages were generated to have source/destination within a given community. The experiment was repeated 10 times and the resulting global/local ranking values were normalized and averaged.

We then run BUBBLE on both real and synthetic traces generated with SWIM, and measure its performance in terms of *cost* (total of replicas for delivered messages), and *success ratio*, for different values of time to live (TTL) of messages. The results are shown respectively in Figures 7 and 8. As you can see, the results are more than satisfying. Most importantly, this is not due to a customized tuning that has been optimized for these forwarding protocols, it is just the same tuning of the previous section that has been used to compute the statistical properties of SWIM. This can be important methodologically: To tune SWIM on a particular scenario with a given number of communities, you can concentrate on a few well known and important statistical properties like inter-contact time, number
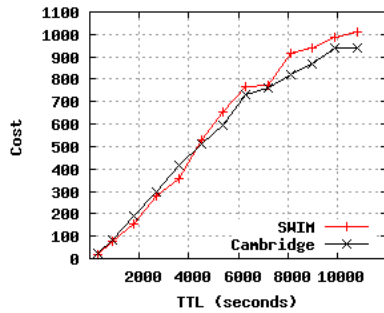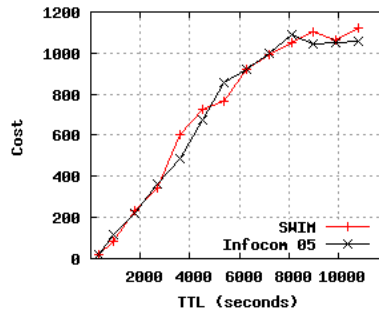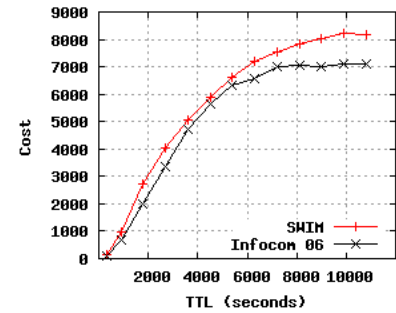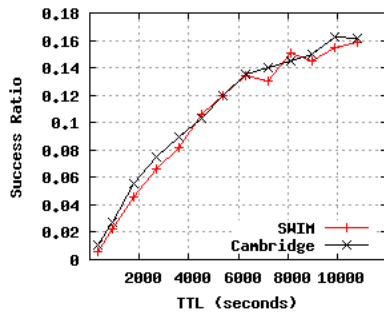
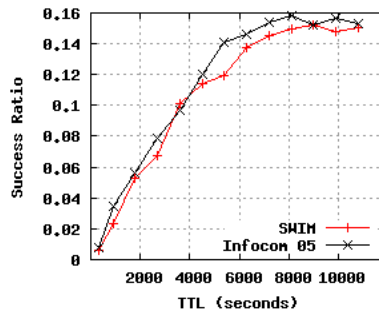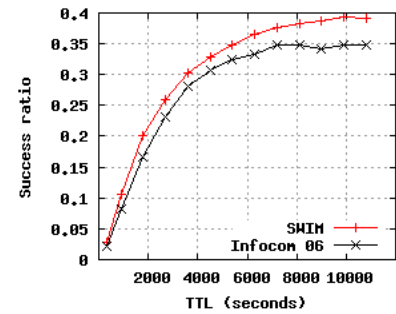| (a) Cambridge and SWIM. | (b) Infocom 05 and SWIM. | (c) Infocom 06 and SWIM. |

Fig. 7. Cost of BUBBLE on real traces and SWIM.

| (a) Cambridge and SWIM. | (b) Infocom 05 and SWIM. | (c) Infocom 06 and SWIM. |

Fig. 8. Success ratio of BUBBLE on real traces and SWIM.

of contacts, and duration of contacts. Then, you can have a good confidence that the model is properly tuned and usable to get meaningful estimation of the performance of complex protocols for social mobile networks.

VII. CONCLUSIONS

In this paper we show how SWIM, a mobility model for ad hoc networking that has been recently introduced, can generate networks with a community structure. SWIM is simple, proves to generate traces that look real, and, moreover, provides an accurate estimation of the performance of sophisticated forwarding protocols based on the community structure of the network. SWIM can be used to improve our understanding of human mobility, it can support theoretical work and it can be very useful to evaluate the performance of networking protocols in scenarios where the structure in communities of the network is important and where the number of nodes can scale up to very large mobile systems, for which we don't have real traces.

REFERENCES

[1] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," *Wireless Communications and Mobile Computing Special issue on Mobile Ad Hoc Networking: Research, Trends and Applications*, vol. 2, no. 5, pp. 483–502, 2002.

[2] D. B. Johnson and D. A. Maltz, "Dynamic source routing in ad hoc wireless networks," in *Mobile Computing*, Imielinski and Korth, Eds. Kluwer Academic Publishers, 1996, vol. 353.

[3] J. Su, A. Chin, A. Popivanova, A. Goel, and E. de Lara, "User mobility for opportunistic ad-hoc networking," in *WMCSA '04: Proceedings of the Sixth IEEE Workshop on Mobile Computing Systems and Applications*. Washington, DC, USA, 2004.

[4] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot, "Pocket switched networks and human mobility in conference environments," in *WDTN '05: Proceeding of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking*. ACM Press, 2005.

[5] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on the design of opportunistic forwarding algorithms," in *INFOCOM '06. Proceedings of the 25th IEEE International Conference on Computer Communications*, 2006.

[6] T. Karagiannis, J.-Y. L. Boudec, and M. Vojnović, "Power law and exponential decay of inter contact times between mobile devices," in *MobiCom '07: Proceedings of the 13th annual ACM international conference on Mobile computing and networking*. ACM, 2007.

[7] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Pocket switched networks: Real-world mobility and its consequences for opportunistic forwarding," Computer Laboratory, University of Cambridge, Tech. Rep., 2006.

[8] H. Cai and D. Y. Eun, "Crossing over the bounded domain: from exponential to power-law inter-meeting time in manet," in *MobiCom '07: Proceedings of the 13th annual ACM international conference on Mobile computing and networking*. ACM, 2007.

[9] A. Mei and J. Stefa, "SWIM: A Simple Model to Generate Small Mobile Worlds," in *INFOCOM '09: Proceedings of The 28th IEEE Conference on Computer Communications*, April 2009.

[10] A. Vahdat and D. Becker, "Epidemic routing for partially connected ad hoc networks," Duke University, Tech. Rep. CS-200006, 2000.

[11] V. Erramilli, M. Crovella, A. Chaintreau, and C. Diot, "Delegation forwarding," in *MobiHoc '08: Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing*. ACM, 2008.

[12] P. Hui, J. Crowcroft, and E. Yoneki, "Bubble rap: social-based forwarding in delay tolerant networks," in *MobiHoc '08: Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing*. ACM, 2008.

[13] I. Rhee, M. Shin, S. Hong, K. Lee, and S. Chong, "On the levy-walk nature of human mobility," in *INFOCOM 2008. IEEE International Conference on Computer Communications. Proceedings*, 2008.

[14] M. Musolesi and C. Mascolo, "Designing mobility models based on social network theory," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 11, no. 3, pp. 59–70, 2007.

[15] M. Piorkowski, N. Sarafijanovic-Djukic, and M. Grossglauser, "On Clustering Phenomenon in Mobile Partitioned Networks," in *The First ACM SIGMOBILE International Workshop on Mobility Models for Networking Research*. ACM, 2008.

[16] F. Ekman, A. Keränen, J. Karvo, and J. Ott, "Working day movement model," in *The First ACM SIGMOBILE International Workshop on Mobility Models for Networking Research*. ACM, 2008.

[17] K. Lee, S. Hong, S. J. Kim, I. Rhee, and S. Chong, "SLAW: A Mobility Model for Human Walks," in *INFOCOM '09: Proceedings of The 28th IEEE Conference on Computer Communications*, April 2009, pp. 2106–2113.

[18] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, pp. 779–782, june 2008.

[19] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, no. 7043, pp. 814–818, June 2005.

[20] P. Bak, *How Nature Works: The Science of Self-Organized Criticality*. Springer-Verlag Telos, October 1996.

[21] R. Gunther, L. Levitin, B. Shapiro, and P. Wagner, "Zipf's law and the effect of ranking on probability distributions," *International Journal of Theoretical Physics*, vol. 35, no. 2, pp. 395–417, 1996.

[22] J. Laherrere and D. Sornette, "Stretched exponential distributions in Nature and Economy: "Fat tails" with characteristic scales," *European Physical Journals*, pp. 525–539, 1998.

[23] J. Laherrere, ""Parabolic fractal" distributions in Nature," *Sciences de la Terre et des Plantés*, April 1996.

[24] J. Leguay, A. Lindgren, J. Scott, T. Friedman, and J. Crowcroft, "Opportunistic content distribution in an urban setting," in *CHANTS '06: Proceedings of the 2006 SIGCOMM workshop on Challenged networks*. ACM, 2006.

[25] J. Leguay, A. Lindgren, J. Scott, T. Riedman, J. Crowcroft, and P. Hui, "CRAWDAD trace upmc/content/imote/cambridge (v. 2006–11–17)," Downloaded from http://crawdad.cs.dartmouth.edu/upmc/content/imote/cambridge, nov 2006.

[26] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, and A. Chaintreau, "CRAWDAD trace cambridge/haggle/imote/infocom (v. 2006–01–31)," Downloaded from http://crawdad.cs.dartmouth.edu/cambridge/haggle/imote/infocom, January 2006.

[27] P. Hui, E. Yoneki, S. Chan, and J. Crowcroft, "Distributed community detection in delay tolerant networks," in *MobiArch '07: Proceedings of 2nd ACM/IEEE international workshop on Mobility in the evolving internet architecture*. New York, NY, USA, 2007.