STEPS - An Approach for Human Mobility Modeling

Anh Dung Nguyen[1,2], Patrick Sénac[1,2], Victor Ramiro[3], and Michel Diaz[2]

[1] ISAE/University of Toulouse, Toulouse, France
[2] LAAS/CNRS, Toulouse, France
[3] NIC Chile Research Labs, Santiago, Chile
{anh-dung.nguyen,patrick.senac}@isae.fr, vramiro@niclabs.cl,
michel.diaz@laas.fr

Abstract. In this paper we introduce Spatio-TEmporal Parametric Stepping (STEPS) - a simple parametric mobility model which can cover a large spectrum of human mobility patterns. STEPS makes abstraction of spatio-temporal preferences in human mobility by using a power law to rule the nodes movement. Nodes in STEPS have preferential attachment to favorite locations where they spend most of their time. Via simulations, we show that STEPS is able, not only to express the peer to peer properties such as inter-contact/contact time and to reflect accurately realistic routing performance, but also to express the structural properties of the underlying interaction graph such as small-world phenomenon. Moreover, STEPS is easy to implement, flexible to configure and also theoretically tractable.

1 Introduction

Human mobility is known to have a significant impact on performance of networks created by wireless portable devices e.g. MANET, DTN. Unfortunately, there is no model that is able to capture all the characteristics of human mobility due to its high complexity. In this paper we introduce Spatio-TEmporal Parametric Stepping (STEPS) - a new powerful formal model for human mobility or mobility inside social/interaction networks. The introduction of this new model is justified by the lack of modeling and expressive power, in the currently used models, for the spatio-temporal correlation usually observed in human mobility.

We show that preferential location attachment and location attractors are invariants properties, at the origin of the spatio-temporal correlation of mobility. Indeed, as observed in several real mobility traces, while few people have a highly nomadic mobility behavior the majority has a more sedentary one.

In this paper, we assess the expressive and modeling power of STEPS by showing that this model successes in expressing easily several fundamental human mobility properties observed in real traces of dynamic network:

1. The distribution of human traveled distance follows a truncated power law.
2. The distribution of pause time between travels follows a truncated power law.

3. The distribution of inter-contact/contact time follow a truncated power law.
4. The underlying dynamic graph can emerge a small-world structure.

The rest of this paper is structured as follows. After an overview of the state of the art in Section 2, we present the major idea behind the model in Section 3. Section 4 formally introduces STEPS as well as some implementation issues. Section 5 shows capacity of STEPS to capture salient features observed in real dynamic networks, going from inter-contact/contact time to epidemic routing performance and small-world phenomenon. Finally we conclude the paper in Section 6.

2   Related Works

Human mobility has attracted a lot of attention of not only computer scientists but also epidemiologists, physicists, etc because its deep understanding may lead to many other important discoveries in different fields. The lack of large scale real mobility traces made that research is initially based on simple abstract models e.g. Random Waypoint, Random Walk (see [3] for a survey). These models whose parameters are usually drawn from an uniform distribution, although are good for simulation, can not reflect the reality and even are considered harmful for research in some cases [14]. In these model, there is no notion of spatio-temporal preferences.

Recently, available real data allows researchers to understand deeper the nature of human mobility. The power law distribution of the traveled distance was initially reported in [6] and [1] in which the authors study the spatial distribution of human movement based on mobile phone and bank note traces. The power law distribution of the inter-contact time was initially studied by Chaintreau et al. in [4]. In [8], Karagiannis et al. confirm this and also suggest that the inter-contact time follows a power law up to a characteristic time (about 12 hours) and then cut-off by an exponential decay.

Based on these findings, some more sophisticated models have been proposed. In [5], the authors have proposed an universal model being able to capture many characteristics of human daily mobility by combining different sub-models. With a lot of parameters to configure, the complexity of this type of model make them hard to use.

In [9], the authors propose SLAW - a Random Direction liked model, except that the traveled distance and the pause time distributions are ruled by a power law. An algorithm for trajectory planning was added to mimic the human behavior of always choosing the optimal path. Although being able to capture statistical characteristics like inter-contact, contact time distribution, the notion of spatio-temporal preferential attachment is not expressed. Moreover, the routing protocol performance results have not been compared with real traces.

Another modeling stream is to integrating social behaviors in the model. In [10], a community based model was proposed in which the movement depends on the relationship between nodes. The network area is divided in zones and the social attractivity of a zone is based on the number of friends in the

same zone. The comparisons of this model with real traces show a difference for the contact time distribution. Moreover, the routing performance has not been shown.

Time Varying Community [7] is another interesting model in which the authors try to model the spatio-temporal preferences of human mobility by creating community zones. Nodes have different probabilities to jump in different communities to capture the spatial preferences. Time structure was build on the basis of night/day and days in a week to capture temporal preferences.

A recent research shows that some mobility model (including [7]), despite of the capacity of capturing the spatio-temporal characteristics, deviate significantly the routing performances compared to ones obtained with real traces [12]. This aspect that has not always been considered in existing models is indeed really important because it shows how a model can confront the real dynamic networks.

3   Characterizing Human Mobility

STEPS is inspired by observable characteristics of the human mobility behaviour, specifically the spatio-temporal correlation. Indeed, people share their daily time between some specific locations at some specific time (e.g. home/office, night/day). This spatio-temporal pattern repeats at different scales and has been recently observed on real traces [7].

On a short time basis (i.e. a day, a week), we can assume that one have a finite space of locations. We define two mobility principles :

- Preferential attachment: the probability for a node to move in a location is inverse proportional to the distance from his preferential location.
- Attractor: when a node is outside of his preferential location, he has a higher probability to move closer to this location than moving farther.

From this point of view, the human mobility can be modeled as a finite state space Markov chain in which the transition probability distribution express a movement pattern. In the next section, we answer the question what exactly this probability distribution. Figure 1(a) illustrates a Markov chain of 4 states which corresponds to 4 locations: (A) House, (B) Office, (C) Shop and (D) Other places.

4   Model Description

In STEPS, a location is modeled as a zone in which a node can move freely according to a random mobility model such as Random Waypoint. The displacement between zones and the staying duration in a zone are both drawn from a power law distribution whose the exponent value expresses the more or less localized mobility. By simply tuning the power law exponent, we can cover a large spectrum of mobility patterns from purely random ones to highly localized ones. Moreover, complex heterogeneity can be described by combining nodes with
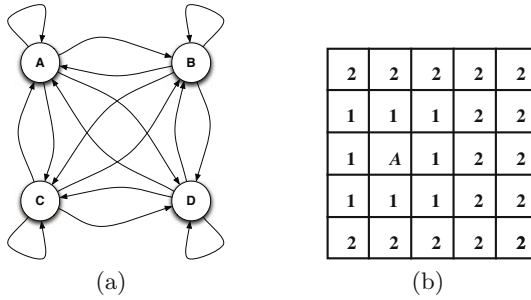
Fig. 1. (a) Human mobility modeling under a markovian view: States represent different localities e.g. House, Office, Shop and Other places, and transitions represent the mobility pattern. (b) 5×5 torus representing the distances from a location A to the other locations.

different mobility patterns as defined by their preferential zones and the related attraction power. Group mobility is also supported on our implementation.

4.1   The Model

Assume that the network area is a square torus divided in $N \times N$ square zones. The distance between theses zones is defined according to a metric (here we use Chebyshev distance). Figure 1(b) illustrates an example of a 5×5 torus with the distances from the zone A. One can imagine a zone as a geographic location (e.g. building, school, supermarket) or a logical location (i.e. a topic of interest such as football, music, philosophy, etc). Therefore, we can use the model to study human geographic mobility or human social behaviours. In this paper, we deal only with the first case.

In the so structured space, each node is associated to a preferential zone Z_0. For the sake of simplicity we assume is this paper that each node is attached to one zone, however this model can be extended by associating several preferential zones to each node. The movement between zones is driven by a power law satisfying two mobility principles described above. The pdf of this power law is given by

$$P[D = d] = \frac{\beta}{(1 + d)^{\alpha}}, \tag{1}$$

where d is the distance from Z_0, α is the power law exponent that represents the attractor power and β is a normalizing constant.

From (1) we can see that : the farther a zone is from the preferential zone, the less probability the node to move in (i.e. principle of preferential attachment). On the other hand, when a node is outside of its preferred zone he has a higher probability to move closer to this one than moving farther (i.e. principle of attraction).

Algorithm 1: STEPS algorithm

Input: Initial zone $\leftarrow Z_0$
repeat
- Select randomly a distance d from the probability distribution (1);
- Select randomly a zone Z_i among all zones that are d distance units away from Z_0;
- Select randomly a point in Z_i;
- Go linearly to this point with a speed randomly chosen from $[v_{\min}, v_{\max}], 0 < v_{\min} \leq v_{\max} < +\infty$;
- Select randomly a staying time t from the probability distribution (2);
while t *has not elapsed* **do**
  | Perform Random Waypoint movement in Z_i ;
end
until *End of simulation*;

The staying time in a zone is also driven by a power law

$$P\left[T = t\right] = \frac{\omega}{t^\tau}, \tag{2}$$

where τ is the temporal preference degree of node and ω is a normalizing constant.

From this small set of modeling parameters the model can cover a full spectrum of mobility behaviours. Indeed, according to the value of the α exponent a node has a more or less nomadic behavior. For instance,

- when $\alpha < 0$, nodes have a higher probability to choose a long distance than a short one and so the preferential zone plays the repulsion role instead of a attraction one,
- when $\alpha > 0$, nodes are more localized,
- when $\alpha = 0$, nodes move randomly towards any zone with a uniform probability.

We summarize the description of STEPS in Algorithm 1.

4.2   Markov Chain Modeling

In this section, we introduce the Markovian model behind STEPS. This analytical analysis makes it possible to derive routing performance bounds when using the model. From a formal point of view STEPS can be modeled by a discrete-time Markov chain of N states where each state corresponds to one zone. Note that the torus structure gives nodes the same spatial distribution wherever their preferential zones (i.e. they have the same number of zones with an equal distances to their preferential zones). More specifically, for a distance d, we have $8 \times d$ zones with equal distances from Z_0. Consequently, the probability to choose one among these zones is

$$P\left[Z_i|d_{Z_i Z_0} = l\right] = \frac{1}{8l}P\left[D = l\right] = \frac{\beta}{8l(1 + l)^\alpha}. \tag{3}$$

Because the probabilities for a node to jump to any zone do not depend on the residing zone but only on the distance from Z_0, the transition probabilities of the Markov chain is defined by a stochastic matrix with similar lines. Therefore the resulting stochastic matrix is a idempotent matrix (i.e. the product of the matrix by itself gives the same matrix)

$$P = \begin{bmatrix} p(Z_0) \ p(Z_1) \ \dots \ p(Z_{n-1}) \\ p(Z_0) \ p(Z_1) \ \dots \ p(Z_{n-1}) \\ \vdots \quad \vdots \quad \ddots \quad \vdots \\ p(Z_0) \ p(Z_1) \ \dots \ p(Z_{n-1}) \end{bmatrix} .$$

Hence it is straight-forward to deduce the stationary state of the Markov chain

$$\Pi = \big(p(Z_0) \ p(Z_1) \ \dots \ p(Z_{n-1}) \big) . \tag{4}$$

From this result, it is interesting to characterize the inter-contact time (i.e. the delay between two consecutive contacts of the same node pair) of STEPS because this characteristic is well known to have a great impact on routing in dynamic networks. To simplify the problem, we assume that a contact occurs if and only if two nodes are in the same zone. Let two nodes A and B move according to the underlying STEPS Markov chain and initially start from the same zone. Let assume that the movement of A and B are independent. For each instant, the probability that the two nodes are in the same zone is

$$p_{contact} = P_A(Z_0)P_B(Z_0) + \dots + P_A(Z_{n-1})P_B(Z_{n-1})$$
$$= \sum_{i=0}^{n-1} P(Z_i)^2 = \sum_{i=0}^{dMax} \frac{1}{8i} \left[\frac{\beta}{(i+1)^\alpha} \right]^2 \quad ,$$

where $dMax = \lfloor \frac{\sqrt{N}}{2} \rfloor$ is the maximum distance a node can attain in the torus.

Let ICT be the discrete random variable which represents the number of instants elapsed before A and B are in contact again. One can consider that ICT follows a geometric distribution with the parameter $p_{contact}$, i.e. the number of trials before the first success of an event with probability of success $p_{contact}$. Hence, the pdf of ICT is given by

$$P\left[ICT = t\right] = (1 - p_{contact})^{t-1} p_{contact}. \tag{5}$$

It is well known that the continuous analog of a geometric distribution is an exponential distribution. Therefore, the inter-contact time distribution for i.i.d. nodes can be approximated by an exponential distribution. This is true when the attractor power α is equals to 0 (i.e. nodes move uniformly) because there is no spatio-temporal correlation between nodes. But when $\alpha \neq 0$, there is a higher correlation in their movement and in consequence the exponential distribution is not a good approximation. Indeed, [2] reports this feature for the Correlated Random Walk model where the correlation of nodes induces the emergence of a power law in the inter-contact time distribution. A generalized closed formula
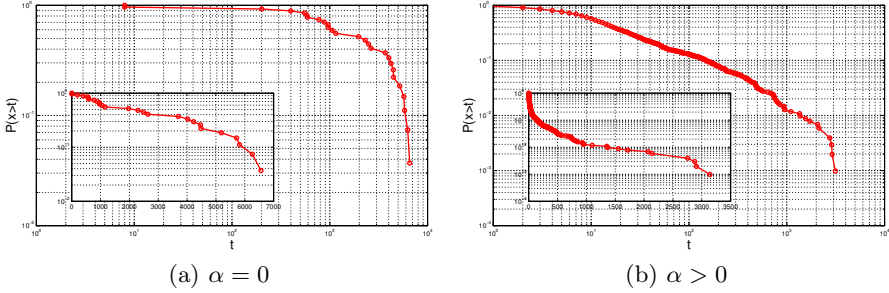
(a) $\alpha = 0$ (b) $\alpha > 0$

Fig. 2. Theoretical Inter-Contact Time Distribution of STEPS

for STEPS inter-contact time is an on-going work. We provide here simulation results related to this feature.

Figure 2 gives the linear-log plot of the complementary cumulative distribution function (CCDF) of the inter-contact time that results from a simulation of the Markov chain described above when $\alpha = 0$ and $\alpha > 0$. In the first case, the inter-contact time distribution fits an exponential distribution (i.e. is represented by a linear function the linear-log plot) while in the second case it fits a power law distribution (i.e. is represented by a linear function in the log-log plot) with an exponential decay tail. This result confirms the relationship between the spatio-temporal correlation of nodes and the emergence of a power law in inter-contact time distribution.

5   Model Properties

It is worth mentioning that a mobility model should express the fundamental properties observed in real dynamic networks. In this section, we show that STEPS can really capture the seminal characteristics of human mobility and that when used for testing routing performance STEPS deliver the same performances as the ones observed on top of real traces (this features is too often neglected when introducing a new mobility model).

5.1   Inter-Contact Time vs. Contact Time Distributions

Inter-Contact Time Distribution. The inter-contact time is defined as the delay between two encounters of a pair of nodes. Real trace analysis suggest that the distribution of inter-contact time can be approximated by a power law up to a characteristic time (i.e. about 12 hours) followed by an exponential decay [8]. In the following we will use the set of traces presented by Chaintreau et al. in [4] as base of comparison with STEPS mobility simulations. Figure 3(a) shows the aggregate CCDF of the inter-contact time (i.e. the CCDF of inter-contact time samples over all distinct pairs of nodes) for different traces. In order to demonstrate the capacity of STEPS to reproduce this feature, we configured STEPS to exhibit the results observed in the Infocom 2006 conference trace. Table 1 summarizes the characteristics of this trace.
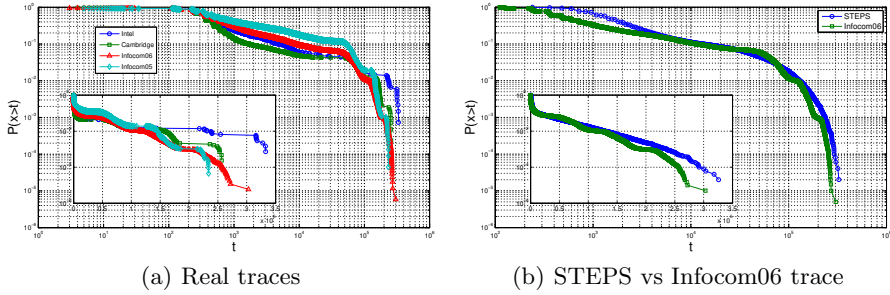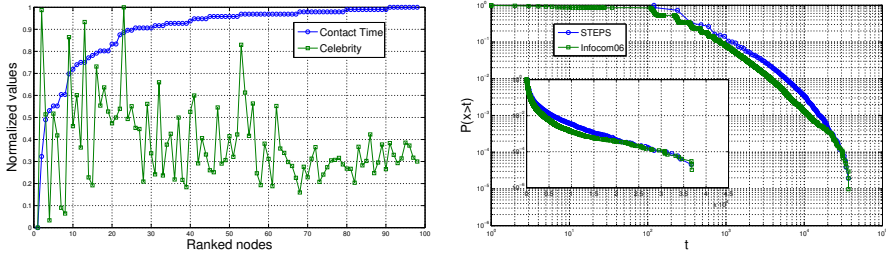
(a) Real traces   (b) STEPS vs Infocom06 trace

Fig. 3. CCDF of Inter-Contact Time

Table 1. Infocom 2006 trace

Number of nodes	98
Duration	4 days
Technology	Bluetooth
Average inter-contact time	1.9 hours
Average contact duration	6 minutes

To simulate the conference environment, we create a 10×10 torus of size $120 \times 120m^2$ that mimics rooms in the conference. The radio range is set to $10m$ which corresponds to Bluetooth technology. Figure 3(b) shows the CCDF of inter-contact time in log-log and lin-log plots. We observe that the resulting inter- contact time distribution as given by the STEPS simulations fits with the one given by the real trace.

Contact Time Distribution. Because of the potential diversity of nodes behavior it is more complicated to reproduce the contact duration given by real traces. Indeed, the average time spent for each contact depends on the person (e.g. some people spend a lot of time to talk while the others just check hands). To the best of our knowledge, the abstract modeling of social behavior has not been studied precisely yet. We measured the average contact duration and the celebrity (i.e. the global number of neighbor nodes) of the Infocom06 nodes and ranked them according to their average contact duration. The result is plotted in Figure 4(a). According to this classification, it appears that the more/less popular the person is, the less/more time he spends for each contact. Because the contact duration of STEPS depends principally on the pause time of the movement inside zone (i.e. the pause time of RWP model), to mimic this behavior we divided nodes in four groups. Each group corresponds to a category of mobility behavior: highly mobile nodes, mobile nodes, slightly mobile and rarely mobile. The pause time for each groups is summarized in Table 2.

With this configuration, we aim to mimic the behavior observed in Infocom06 trace where a large percentage of nodes have short contacts and a few nodes have long to very long contacts. The CCDFs of contact time of STEPS

(a) Social Characteristic Observed in In-   (b) CCDF of Contact Time of STEPS vs
focom06   Infocom06 Trace

Fig. 4. Contact Time Behavior of STEPS vs Real Trace

Table 2. Group categories

Dynamicity categories	RWP pause time range (s)	Number of nodes
Very high	[0, 60]	65
High	[60, 900]	15
Low	[900, 3600]	10
Very low	[3600, 43200]	8

and Infocom06 trace as shown in Figure 4(b) show that STEPS can also capture
with high accuracy this mobility behavior.

5.2   Epidemic Routing Performance

A mobility model has not only to capture the salient features observed in real
traces but must also reproduce the performances given by routing protocols on
top of real traces. In order to assess the capacity of STEPS to offer this important
property we ran Epidemic routing on STEPS and Infocom06 trace and compared
the respective routing delays. For each trace, the average delay to spread a mes-
sage to all the nodes is measured in function of the number of nodes who received
the message. Figure 5 shows that STEPS is able to reflect at the simulation level
the performance of Epidemic routing when applied on real traces.

5.3   Human Mobility Structure Studied with STEPS

A mobility model should allow not only to express faithfully peer-to-peer inter-
actions properties such as inter-contact time and contact duration, but should
be able also to reproduce the fundamental structure of the underlying interac-
tion graph as modeled by a temporal graph, i.e. graphs with time varying edges.
The structural properties of static interaction graphs have been studied lead-
ing to the observation of numerous instances of real interaction graph with a
high clustering and a low shortest path length [13]. Such a structure of graph is
called small-world. With respect to routing, the small world structure induces
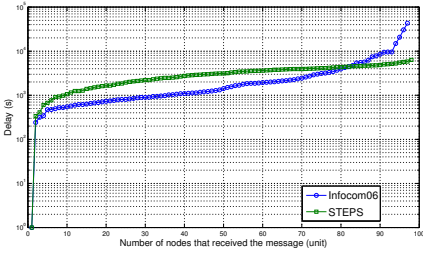fast message spreading in the underlying network.

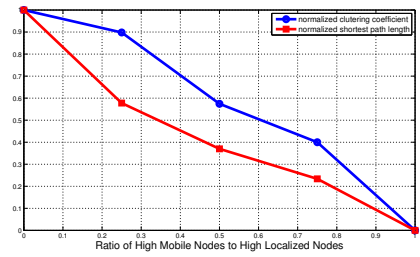**Fig. 5.** Epidemic Routing Delay of STEPS vs Infocom06 Trace

Fig. 6. Small-World Structure in STEPS

We extended the notions of clustering coefficient and shortest path length introduced in [13] for dynamic graph. In this paper, we present only the most important definitions due to the lack of space.

Let $\mathcal{G}(t) = (\mathcal{V}(t), \mathcal{E}(t))$ be a temporal graph with a time varying set $\mathcal{V}(t)$ of vertexes and with a time varying set $\mathcal{E}(t)$ of edges.

1. Temporal Clustering Coefficient : Let $\mathcal{N}(w, i)$ be the set of neighbors of node i in a time window w. The temporal clustering coefficient is defined as the ratio of the actual number of connections between neighbors of i to the theoretical number of connections between them during a time window w. Intuitively, it represents the cliquishness of a time varying friendship circle. Formally, that is

$$C_i = \frac{2 \sum_j |\mathcal{N}(w, j) \cap \mathcal{N}(w, i)|}{|\mathcal{N}(w, i)| \, (|\mathcal{N}(w, i)| - 1)}, \tag{6}$$

   where j is a neighbor of i and $|X|$ denotes the cardinal of X.

2. Temporal Shortest Path Length : Let $\mathcal{R}_{ij}(t) \in \{0, 1\}$ denotes a direct or indirect (i.e. via multiple connections at different times) connection between node i and node j at time t. The shortest path length between i and j is defined as the earliest instant when there is a connection between them. That is

$$L_{ij} = \inf\{t | \mathcal{R}_{ij}(t) = 1\}. \tag{7}$$

To visualize the phenomenon in STEPS, we create scenarios where there are two categories of nodes with different attractor power. In the first category, nodes have a high mobile behaviour (i.e. α is small) while in the second one, they are more localized (i.e. α is large). The idea is to "rewiring" a high clustered dynamic graph (i.e. with the population is in the second category) by introducing high mobile nodes into the population. By tuning the ratio of number of nodes between these two categories, we measure the metrics defined in (6) and (7). Figure 6 shows that between two extrema, with certain value of the ratio, the network have a structure where the clustering coefficient is high while the shortest path length is low. This result suggests the existence of the small-world phenomenon in dynamic graph.

6 Conclusion

In this paper, we introduce STEPS, a generic and simple mobility model which abstracts the spatio-temporal correlation of human mobility. Based on the principles of preferential attachment and location attractor, this model can cover a large spectrum of human mobility patterns by tuning a small set of parameters. Via simulations, the model is shown to be able to capture different characteristics observed on top of real mobility traces. On the other hand, the model can also reflect accurately realistic routing performances, one of important aspect often neglected in proposed mobility models. Moreover, STEPS can reflect the structural features of the underlying dynamic graph as well as the peer-to-peer ones. Finally, the underlying Markovian basis make it possible to derive analytical results from this model.

References

1. Brockmann, D., Hufnagel, L., Geisel, T.: The scaling laws of human travel. Nature 439(7075), 462–465 (2006)
2. Cai, H., Eun, D.: Toward stochastic anatomy of inter-meeting time distribution under general mobility models. In: Proceedings of the 9th ACM International Symposium on Mobile Ad hoc Networking and Computing, MobiHoc 2008, pp. 273–282. ACM, New York (2008)
3. Camp, T., Boleng, J., Davies, V.: A survey of mobility models for ad hoc network research. Wireless Communications and Mobile Computing 2(5), 483–502 (2002)
4. Chaintreau, A., Hui, P., Crowcroft, J., Diot, C., Gass, R., Scott, J.: Impact of human mobility on opportunistic forwarding algorithms. IEEE Transactions on Mobile Computing 6(6), 606–620 (2007)
5. Ekman, F., Keränen, A., Karvo, J., Ott, J.: Working day movement model. In: Proceeding of the 1st ACM SIGMOBILE Workshop on Mobility Models, pp. 33–40. ACM, New York (2008)
6. González, M., Hidalgo, C., Barabási, A.: Understanding individual human mobility patterns. Nature 453(7196), 779–782 (2008)
7. Hsu, W., Spyropoulos, T., Psounis, K., Helmy, A.: Modeling spatial and temporal dependencies of user mobility in wireless mobile networks. IEEE/ACM Transactions on Networking 17(5), 1564–1577 (2009)
8. Karagiannis, T., Le Boudec, J., Vojnovic, M.: Power law and exponential decay of inter contact times between mobile devices. IEEE Transactions on Mobile Computing, 183–194 (2010)
9. Lee, K., Hong, S., Kim, S., Rhee, I., Chong, S.: Slaw: A mobility model for human walks. dspace.kaist.ac.kr (2009)
10. Musolesi, M., Mascolo, C.: A community based mobility model for ad hoc network research. In: Proceedings of the 2nd International Workshop on Multi-hop Ad hoc Networks: from Theory to Reality, pp. 31–38. ACM, New York (2006)
11. Tang, J., Musolesi, M., Mascolo, C., Latora, V.: Characterising temporal distance and reachability in mobile and online social networks. ACM SIGCOMM Computer Communication Review 40(1), 118–124 (2010)

12. Thakur, G.S., Kumar, U., Helmy, A., Hsu, W.-J.: Analysis of Spatio-Temporal Preferences and Encounter Statistics for DTN Performance (July 2010)
13. Watts, D., Strogatz, S.: Collective dynamics of small-world networks. Nature 393(6684), 440–442 (1998)
14. Yoon, J., Liu, M., Noble, B.: Random waypoint considered harmful. In: IEEE Societies INFOCOM 2003, Twenty-Second Annual Joint Conference of the IEEE Computer and Communications, vol. 2 (2003)