Gauging Human Mobility Characteristics and its Impact on Mobile Routing Performance

Gautam S. Thakur*

Department of Computer and Information Science and Engineering,
University of Florida,
Gainesville FL
E-mail: gsthakur@cise.ufl.edu
* Corresponding author

Udayan Kumar

Department of Computer and Information Science and Engineering,
University of Florida,
Gainesville FL
E-mail: ukumar@cise.ufl.edu

Wei-Jen Hsu

Cisco Systems Inc.,
San Jose, CA
E-mail: wehsu@cisco.com

Ahmed Helmy

Department of Computer and Information Science and Engineering,
University of Florida,
Gainesville FL
E-mail: helmy@cise.ufl.edu

Abstract: A new generation of "behavior-aware" services and networks are emerging in what may define future mobile social networks. It is of great importance for current set of mobility models to understand and realistically model mobile users behavioral characteristics and also realistically reproduce their effects on the performance of networking protocols. Recent work on mobility modeling focused on replicating metrics of encounter statistics and spatio-temporal preferences to improve the realism of mobility models. No studies have been conducted, however, to show whether matching these metrics is sufficient to accurately reproduce structural dynamics and mobile networking protocol performance. In this study, we address these specific problems, and attempt to show the sufficiency (or lack thereof) of existing encounter and mobility metrics in reproducing realistic effects of mobility on networking protocols. We first analyze the characteristics of two well-established mobility models; the random direction and the time-variant community (TVC) models, and study whether they capture encounter statistics and preference patterns observed in realworld traces. Second, we introduce mobile user similarity, its definition, analysis and modeling. To define similarity, we measures the difference of the major spatio-temporal behavioral trends using their association matrix. Such measure is then used to cluster users into similarity communities and compare them in the traces and the mobility models. Finally, we contrast the performance of epidemic routing on the mobility models, to that based on extensive mobility traces. We provide three main findings: (i) Careful parameterization of the models can indeed replicate the metrics in question (e.g., inter-encounter time distribution). (ii) Our results show a rich set of similar communities in real mobile societies with distinct behavioral clusters of users. This is true for all the traces studied, with the trend being consistent over time. (iii) Even carefully crafted mobility models surprisingly result in structural dynamics and protocol performance that is dramatically different from the trace-driven performance. These findings strongly suggest that similarity should be explicitly captured in future mobility models, which motivates the need to re-visit mobility modeling to incorporate accurate behavioral models in the future.

Keywords: Delay Tolerant Network; Encounter Statistics; Mobility Models; Protocol testing; Similarity; Mobile Societies.

Reference to this paper should be made as follows: Thakur, G., Kumar, U., Hsu, -W., and Helmy, A. (2011) 'Delay Tolerant Network; Encounter Statistics; Mobility Models; Similarity; Mobile Societies', *Int. J. Sensor Networks*, Vol. 1, Nos. 3/4, pp.197–212.

Biographical notes: Gautam S. Thakur is doing his PhD in The Department of Computer and Information Science and Engineering at University of Florida, Gainesville. His research interest includes Mobility Modeling, traffic Measurement and Characterization (Internet & Vehicular), protocol testing and performance in Wireless Networks.

Udayan Kumar received his B.Tech degree from DA-IICT, Gandhinagar, India, MS from University of Florida and is currently pursuing PhD from University of Florida. His research interests lie in understanding users' social behavior from network traces and utilizing it to develop solutions for the challenges faced by adhoc and mobile social networks.

Wei-jen Hsu was born in Taipei, Taiwan, in March 1977. He received the B.S. degree in Electrical Engineering and the M.S. degree in Communication Engineering from National Taiwan University in June 1999 and June 2001, respectively. He received the Engineer degree in Electrical Engineering from University of Southern California in August 2006, and the Ph.D. degree in Computer Science from University of Florida in August 2008. He joined Cisco Systems, Inc. in 2008. His main research interest involves the utilization of realistic measurement data in various tasks in computer networks, including user modeling and behavior-aware protocol design.

Ahmed Helmy received Ph.D. in Computer Science '99 from the University of Southern California (USC), M.S. in Electrical Engineering (EE) '95 from USC, M.S. Eng. Math '94 & B.S. in EE '92 from Cairo University. He was a key researcher in the NS- 2 and PIM projects at USC/ISI from '95-99. Starting '99 has been on the EE Dept faculty at USC and the director of the wireless networking Lab. Since Fall '06, he became an associate professor and the director of the mobile networking laboratory in the CISE Dept at the University of Florida. He received the NSF CAREER Award in '02, in '00 Zumberge Research Award, and the best paper award at IEEE/IFIP MMNS '02. In '04 & '05, he got the best merit ranking in the EE-USC faculty. In '08 & '07, he was a finalist and winner in the ACM MobiCom SRC. In '10, he was a winner in the ACM MobiCom WiNTech demo competition. He is an area editor of the Ad Hoc Networks Journal Elsevier and IEEE Computer, and the workshop coordination chair for ACM SIGMOBILE. His research interests include design, analysis & measurement of wireless adhoc, sensor & mobile social networks, mobility modeling, multicast protocols, IP mobility & network simulation.

1 Introduction

The proliferation of highly capable mobile devices (e.g., laptops, smart phones, tablets) with multi-sensing capabilities greatly facilitates the capture of mobility traces(Helmy, et al., 2010; Kotz, et al., 2005) and the direct exchange of information through encounters. Mobility traces can then be used as guidelines for modeling purposes. More realistic mobility models have been created, by mimicking encounter statistics (Chaintreau, et al., 2007; Karagiannis, et al., 2007) or mobile user location visitation preferences (Hsu, et al., 2009). Much of the recent modeling work focused on encounter metrics; such as inter-encounter and hitting time distribution (Chaintreau, et al., 2007), meeting duration (Chaintreau, et al., 2007; Hsu, et al., 2009), or spatio- temporal profiles (Hsu, et al., 2009). These metrics are generally considered important to the operation of mobile networks in general, including DTNs, adhoc and sensor networks. DTNs are characterized by intermittent connectivity, limited end-to-end connectivity and node resources. Future social networks are expected to have classes of applications that are aware of mobile users' behavioral profiles and preferences and are likely to support peer-to-peer mobile networking including DTNs. A new generation of protocols is emerging, including behavior-aware communication paradigms (such as profile-cast (Hsu, et al., 2008)) and service architectures (such as participatory sensing (Nazir, et al., 2010; Shilton, et al., 2008)). Such behavior-aware communication paradigm leverages user behavior and preferences to achieve efficient operation in DTNs (e.g., interest-based target message forwarding; encounter-based routing, mobile resource discovery). Accurate models of mobile user behavioral profiles are essential for the analysis, performance evaluation, and simulation of such networking protocols. Hence, there is a compelling need to understand and realistically model mobile users behavioral profiles, similarity and clustering of user groups.

Earlier work on mobility modeling presented advances in random mobility models (e.g., RWP, RD (Camp, et al., 2002)), synthetic models that attempt to capture spatial correlation between nodes (e.g., group models (Bai et al., 2008)) or temporal correlation and geographic restrictions (e.g., Freeway, Manhattan, Pathway Models (Bai et al., 2008)). More recent models tend to be trace-driven and some account for location preferences and temporal repetition (Hsu, et al., 2009). However, similarity characteristics between clusters of nodes, which lie in the heart of behavior-aware networking, have not been modeled explicitly by these mobility models. Hence, it is unclear whether (and to which degree) similarity between mobile nodes is captured, and more importantly, how closely can such models be fine-tuned to replicate "social structures", such as groups with distinct behavior observed from real traces.

Also, no studies have been conducted to show whether matching metrics is sufficient to accurately reproduce DTN protocol performance. In this study, we thoroughly examine this specific problem, and attempt to show, for the first time, the sufficiency (or lack thereof) of existing encounter and mobility metrics in reproducing realistic effects of mobility on structural dynamics and the performance of networking protocols.

In this paper, first we analyze spatio-temporal properties and encounter statistics of two realistic wireless measurement traces. We then evaluate the same characteristic on the synthetic traces produced by two different mobility models; the random direction model (Royer, et al., 2001) and the time-variant community (TVC) model (Hsu, et al., 2009). Specifically, we analyze two commonly used encounter statistics; inter-meeting time and meeting duration, in addition to two spatio-temporal metrics; periodic re-appearance and location visitation preference. Our results show that models mimic such statistics if carefully tuned.

Second, we address issues related to mobile user similarity, its definition, analysis and modeling. Similarity, in this study, is defined by mobility preferences, and is meant to reflect the users interests to the extent that can be captured by wireless measurements of on-line usage. To define similarity, we adopt a behavioral-profile based on users mobility and location preferences using an on-line association matrix representation, and then use the cosine product of their weighted Eigen-behaviors to capture similarity between users. This quantitatively compares the major spatio-temporal behavioral trends between mobile network users, and can be used for clustering users into similarity groups or communities. Note that this may not reflect social ties between users or relationships per se, but does reflect mobility- related behavior that will affect connectivity and network topology dynamics in a DTN setting.

We analyze similarity distributions of mobile user populations in two settings. The first analysis aims to establish deep understanding of realistic similarity distributions in such mobile societies. It is based on real measurements of over 8860 users for a month in four major university campuses, USC (Helmy, et al., 2010) IBM Watson , Dartmouth (Kotz, et al., 2005) and UF. It may be reasonable to expect some clustering of users that belong to similar affiliations, but quantification of such clustering and its stability over time is necessary for developing accurate similarity models. Furthermore, on-line behavior that reflects distribution of active wireless devices may not necessarily reflect work or study affiliations or social clustering. For DTNs, on-line activity and mobility preferences translate into encounters that are used for opportunistic message forwarding, and this is the focus of our study rather than social relations per se. The second similarity analysis we conduct aims to investigate whether existing mobility models provide a reasonable approximation of realistic similarity distributions found in the campus traces.

Our results show that among mobile users, we can discover distinct clusters of users that are similar to each other, while dissimilar to other clusters. This is true for all campuses, with the trend being consistent and stable over time. We find an average modularity of 0.64, clustering coefficient of 0.86 and path length of 0.24 among discovered clusters. Surprisingly, however, we find that the existing mobility models do not explicitly capture similarity and result in homogeneous users that are all similar to each other (in one big cluster). This finding generalizes to all other mobility models that produce homogeneous users, not only the mobility models studied in this paper. Thus the richness and diversity of user behavioral patterns is not captured in any degree in the existing models. Our findings strongly suggest that unless similarity is explicitly captured in mobility models, the resulting behavioral patterns are likely to deviate dramatically from reality, sometimes totally missing the richness in the similarity distribution found in the traces. Furthermore, this indicates our current inability to accurately simulate and evaluate similarity-based protocols, services and architectures using mobility models.

Finally, we perform epidemic routing (Vahdat, et al., 2000) on the synthetic (model generated) traces and real network traces and compares their network performance. Surprisingly, through systematic analysis, we find that even when mobility models reflect equivalent spatio-temporal and encounter statistics, they exhibit large DTN routing performance discrepancy with the real scenarios. Furthermore, they clearly show the insufficiency of existing encounter and preference metrics as a measure of mobility model goodness. Systematically establishing a new set of meaningful mobility metrics should certainly be addressed in future works. This also motivates the need to re-visit mobility modeling to incorporate accurate behavioral models in the future.

The paper is structured as: In section 3, we describe various types of wireless traces used in this study, in section 4, we discuss mobility models used to compare against real-world traces. We study human mobility characteristics in the section 5 and compare the routing performance in section 6. Finally, we discuss the results in section 7 and concludes this paper in section 8.

2 Related Work

Delay Tolerant Networks (DTNs) are essentially opportunistic networks. These types of networks do not demand permanent connectivity between source and destination; instead attempt to make best use of any scheme available that can get the message across. Mobility of the nodes is often realized for transferring the messages. Design of any communication protocols for DTNs is heavily dependent on how well the underlying mobility is understood (Bai et al., 2003; Spyropoulos, et al., 2006a). There are two elemental ways to design and test the protocols for DTNs, namely Trace-Based and Mobility Model based (Bai et al., 2008). In case of trace based design and evaluation, a mobility trace can be downloaded from a limited number of trace repositories (Helmy, et al., 2010; Kotz, et al., 2005). These trace are from the real world and capture real mobility patterns of the users belonging to the traces. For the trace collection environment, testing the protocol on the traces would produce most realistic results. But there are quite a few drawbacks of using real traces such as limited number of traces, not capturing all scenarios and inability to generalize the results based on a few traces. Due to these drawbacks, researchers have proposed models that capture key characteristics of human mobility and produce synthetic traces.

Due to the complexity of understanding human mobility and modeling it, models are created to reproduce few characteristics from real traces such as inter-encounter time (Cen, et al., 2008), regularity (Costa, et al., 2008) and community behavior (Eagle, et al., 2006; Cen, et al., 2008; Hsu, et al., 2009; Girvan, et al., 2002; Kotz, et al., 2005). In most cases, a synthetic trace is validated by comparing a few key characteristics against the real trace. This validation we think is not the best, as it does not test the application oriented parameters of the generated trace.

A new paradigm of protocols that relies on the human behavioral patterns has gained recent attention in DTN-related research. In these studies, researchers attempt to use social aspects of human mobility to derive new services and protocols (Chaintreau, et al., 2007; Fall, et al., 2002; Horn, et al., 1990; Hsu, et al., 2006a). As an example, researchers created a behavior-oriented service, called profile-cast, that relies on spatio-temporal similarities between the users(Ekman, et al., 2008). Profile-cast provides a systematic framework to utilize implicit relationships discovered among mobile users for interest- based message efficient forwarding and delivery in DTNs. Participatory sensing (Kumar, et al., 2010; Mtibaa, et al., 2010; Nelson, et al., 2007) provides a service for crowd sourcing using recruiting campaigns using mobile user profiles (Fall, et al., 2002). All these works rely on and utilize similarity of mobile user profiles. In all the above scenarios how can we be sure that a given mobility model can capture all the characteristics needed to test these protocols.

In an attempt to answer the above question, in this work, we have taken a novel approach for evaluating the mobility models; which is to compare the performance of routing protocols on synthetic trace and on real trace whose characteristics were utilized to create/validate the mobility model. This approach allows us to test and create mobility models while keeping the applicability of the generated traces. In this work, as a case study, we consider a complex mobility model TVC (Hsu, et al., 2009) (along with random direction mobility model (Hsu, et al., 2005)). This model generates the non- homogeneous behaviors of mobile users in both space and time. The traces generated by this model show (i) skewed location visiting preferences; (ii) time dependent periodical reappearance of mobile users as seen in WLAN measurements along with other encounter statistics such as average node degree and meeting time. It uses several real traces to validate the correctness of its design.

Table 1 Details of Wireless Measurements

Campus	*# Users*	*Duration*
Dartmouth	1500	Fall 2007
Infocom	42	3 days
IBM Watson	1366	Fall 2006
Univ. of Florida	3000	Fall 2008
USC	3000	Fall 2007

3 Data Set and Trace Analysis

In this section, we discuss the details of the traces used. One set of traces contain WLAN session logs. The other type of traces contain logs of bluetooth encounters among mobile devices.

3.1 Data Set

In order to realize the efficacy of mobility models matching real protocol performance and network dynamics, we intend to compare their output against wireless measurements. WLAN dataset from four university campuses are considered as shown in Table 2. We collect these datasets from the publicly available MobiLib(Helmy, et al., 2010) and Crawdad(Kotz, et al., 2005) repositories. Table 2 provides the detail of these WLAN measurements. We chose university campuses because they are extensive, have high density of active users and include location information. Also, these datasets have been used in previous studies of mobility modeling (Hsu, et al., 2009; Hui, et al., 2008; Kim, et al., 2007; Yoneki, et al., 2007). We perform Systematic Random Sampling (Schutt, et al., 2006) on the datasets to get an unbiased subset of mobile users from the population. Table 2 specifies the sampling frame that we use for this study. In the second step, we extract relevant statistics of mobile user spatio-temporal patterns. In the third step, for each mobile user we obtain normalized association matrix with time granularity of one day. On this matrix we apply SVD to extract the dominant trends. Finally, we compute the cosine similarity of all user pairs. We perform this process iteratively for four different time intervals: 1 week, 2 weeks, 3 weeks and 4 weeks. These traces are publicly available at (Kotz, et al., 2005) although we customized it in a format that suits us. Initially, we investigate mobile users' preferential attachment to certain locations and their time- dependent periodic behavior. Later on we investigate structural dynamics and compare them with the mobility models output.

Our second type dataset comprise encounter (radio contact between two mobile devices) traces from IEEE Infocom 2005 iMotes experiment(Scott, et al., 2009). This data is collected using Intel's iMote, which communicate on Bluetooth protocol and log contact information of all visible Bluetooth capable devices. Such a record contains three entities - MAC address, start time, end time that correspond to each encounter between the host and foreign device. As part of the experiment, these devices were distributed in conference settings to 41 participants for a period of three-four days. We transformed the gathered data for our need to study inter-meeting time and duration of meeting among mobile users and compare output between model and reality. This dataset is available at (Kotz, et al., 2005).

3.2 Encounter Traces

In order to pursue a study on the encounter statistics and dynamic routing in DTN, we need measurements that quantitatively depict the contact (a.k.a. encounter) between mobile users. An encounter occurs between a pair of nodes when they are in a radio communication range of each other. This is straightforward for the iMotes Bluetooth measurements that contain precise encounter information. However, the WLAN measurements are accumulated at the access point level and contain usage patterns. So, we need to convert these measurements in a way to get user encounters as well as maintaining their spatio-temporal footprints. We consider encounter in WLAN if two users connect to same access point and share online session time. For example, Alice and Bob are connected to access point AP-1 between 10:00 AM-02:00 PM. A counter argument can be established by saying that some WLAN devices may miss encounters beyond their coverage region of access points, but WLAN measurements have the advantage to obtain traces in much larger sizes with richer user presence. They also contain location information, which helps in spatio-temporal analysis. Mostly, a Bluetooth experiment has small set of user base for a limited time period.

4 Mobility Models Studied

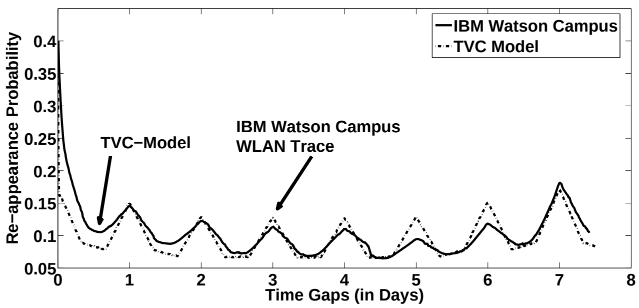
In this section, we discuss two mobility models used for evaluation. We use Random Direction Model (Royer, et al., 2001), which does not posses any spatial or temporal structure in mobility decisions, as an example from typical random mobility models. The lack of spatial and temporal structure leads to faster mixing of the mobile nodes, and sets the lower bound for delay and message delivery overhead. This, as we will show, deviates from realistic mobility traces significantly. We further consider Time Variant Community model (Hsu, et al., 2009) as an example of trace-based mobility models, which incorporate realistic mobility characteristics observed in real traces. Our goal is to evaluate whether such realistic mobility models lead to more realistic evaluation of routing performances. In the following text, we briefly describe these models and construct trace driven DTN scenarios to estimate routing performance.

4.1 Random Direction Model

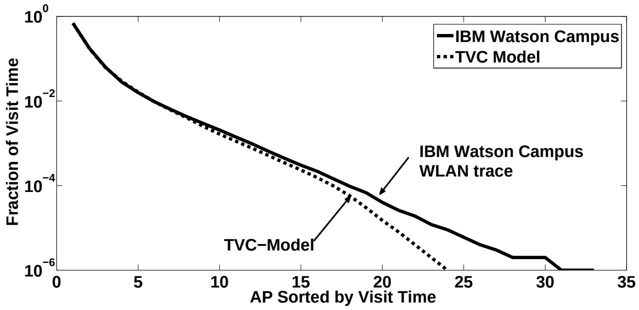
In random direction model, a mobile node makes random mobility decisions with respect to current time or location, independent of other nodes. A node randomly picks a movement direction, and takes straight-line movement towards that direction for a given distance. The node then stops for a given pause time before selecting a new direction to move. This model is more stable as compared to other random models and provides quantitatively even distribution of nodes in the simulation area. We setup this model to investigate the effect of random movements on DTN performance. We modify this model in two ways: (1) the baseline random direction model described above; (2) we add on/off behavior of mobile nodes (i.e., when a node is 'off', it cannot receive/transmit packets), which corresponds to the fact that mobile devices are not always turned on.

4.2 Time Variant Community Model

We choose the TVC model (Hsu, et al., 2009) as an example of trace-based mobility models that capture realistic features of human mobility. Specifically, the TVC model allows configurations to capture (1) spatial preference and (2) temporal periodicity in human mobility. With the setting of communities, preferred locations can be designated and mobile nodes visit such locations more often. The visits are further made periodically with the setting of time periods. TVC model

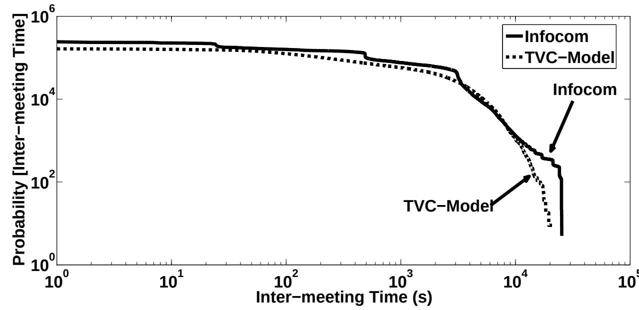


(a) Periodic re-appearance

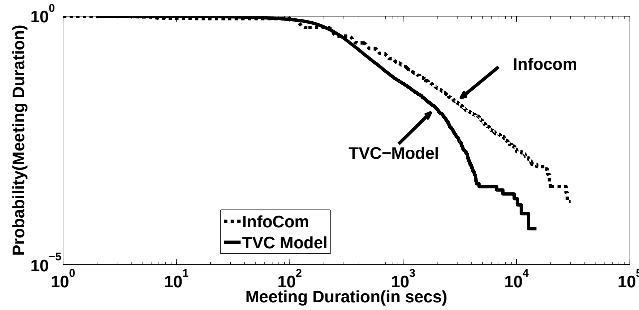


(b) Location visiting preferences

Figure 1 (a) Periodic re-appearances of mobile users in the IBM Watson Campus. TVC Model recreates similar preferences as observed in real traces. (b) TVC depicts skewed location visiting preferences as observed in IBM Watson WLAN traces.



(a) Inter meeting time



(b) Meeting duration

Figure 2 (a) TVC Model depicts meeting duration as measured in the real Infocom traces. (b) Inter-meeting time between mobile users are also similar for the TVC and real Infocom traces.

also includes on/off behavior of mobile nodes. It is shown that with careful community and time period setup, TVC model produces mobility characteristics that match with the real mobility traces better. Since the setup of TVC model is scenario specific, in this paper we have considered two instances of TVC model setup. We synthesize mobility traces from different settings of TVC (i) with matching location-visiting preferences and periodical visits to a trace collected at a research lab (ii) with matching encounter statistics at a conference. It is our goal in this paper to evaluate whether such improved realism in mobility characteristics translates to higher similarity in terms of routing performance to real traces, when we use TVC as opposed to random models.

5 Human Mobility Characteristics

In this section, we analyze set of metrics used to capture non-homogenous behavior of human mobility. It includes spatio-temporal preferences and encounter statistics. Later, we introduce the concept of *similarity* among mobile users and demonstrate its existence in the realworld traces and discover clusters of such users with high similarity index. Finally, we also evaluate current mobility models to capture similarity and clustering effects.

5.1 *Analysis of Spatio-Temporal Preferences*

The non-homogenous behavior of mobile users in space and time is captured by: (i) Skewed location visiting preferences (ii) Periodical reappearances. Studies carried out in (Eagle, et al., 2006; Hsu, et al., 2007, 2006; Scott, et al., 2009; Kim, et al., 2007) tell us that mobile user exhibit preferential attachment and periodic reappearances to few locations in DTNs. We assume, understanding these distributions aid to better message dissemination, prediction of information transmission and the message delivery in opportunistic setting.

We construct the TVC model to generate a month long synthetic trace for IBM Watson's 1366 nodes. In Figure 1 (a) and (b), we see that TVC model demonstrates realistically close location visiting and periodic reappearance properties. For brevity, periodic re-appearances are plotted for seven days only. The re-appearance of spikes demonstrates users visit the same location(s) with higher probability in a periodic fashion. A normalized curve of location preferences show nodes visit very few locations although spending significant amount of their online time. These two characteristics when combined results in better predicting the mobility and on/off patterns of mobile nodes. Furthermore, it can also help to identify hotspots and to measure an approximate delay in message reception. Next, we analyze the state space of encounters among mobile nodes.

5.2 Analysis of Encounter Statistics

In dynamic infrastructure-less mobile networks (like DTNs etc.), the routing is performed by data carrying mobile nodes. The exchange of information takes place when two nodes encounter (a.k.a meet) each other. Intuitively, we can improve routing mechanism given we understand the statistics of these encounter patterns. So, we analyze two encounter statistics: (i) Intermeeting Time, which is the time gap that separates two consecutive mobile encounters. (ii) Meeting Duration, which is the single uninterrupted meeting duration surrounded by intermeeting times. Thus, our statistics alternate between each other. These statistics as mentioned in (Karagiannis, et al., 2007) can have important implications on the performance of opportunistic forwarding algorithms in challenged networks.

The TVC model has ability to generate measurements to analyze encounter statistics. So, we configure it for Infocom setting to generate individual mobility traces for the same number of 41 nodes and for an equivalent duration of four days. The simulation area is modeled like a conference setting with flexibility to visit hotel rooms and outside locations. We later on process the generated traces and plot them along real measurements. The CDF plots in Figure 2 show that model significantly matches real encounter statistics. We see intermeeting time follows Powerlaw distribution up to a characteristic time period after which it decays exponentially. This made us to believe that TVC can also be used to model encounter patterns for unknown scenarios. We conclude, TVC model is statistically accurate on these metrics and closely follow observed realities. We hope that TVC will capture structural dynamics and will show close protocol performance as well, which we see next. In our case, we assume these metrics if captured via a model are vital in achieving identical performance in routing

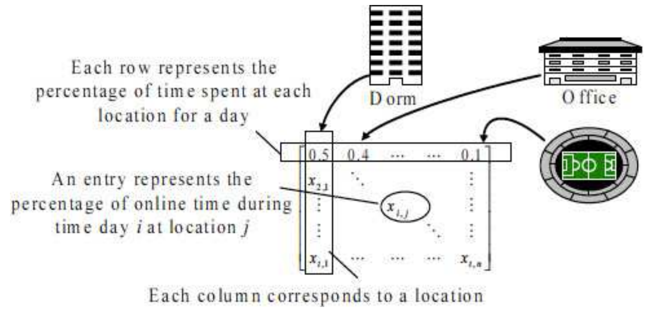
Table 2 Anonymized WLAN session sample

Mac ID	*Location*	*Start Time*	*End Time*
aa:bb:cc:dd:ee:ff	Loc-1	64400343	66404567
aa:bb:cc:dd:ee:ff	Loc-2	85895623	86895742
aa:bb:cc:dd:ee:ff	Loc-3	87444343	89404567
aa:bb:cc:dd:ee:ff	Loc-4	98846767	99878766

5.3 Similarity and Structural Dynamics

5.3.1 Similarity

The congregation of mobile agents with similar characteristic patterns naturally develops mobile societies in wireless networks (Eagle, et al., 2009; Hui, et al., 2007; Yoneki, et al., 2007). Upon reflection it should come as no surprise that these characteristics in particular also have a big impact on the overall behavior of the system (Costa, et al., 2008; Hui, et al., 2008; Mtibaa, et al., 2010; Musolesi, et al., 2008). Researchers have long been working to infer these characteristics and ways to measure them. One major observation is that people demonstrate periodic reappearances at certain locations (Eagle, et al., 2006; Hsu, et al., 2007; Kim, et al., 2007), which in turn breeds connection among similar instances (McPherson, et al., 2001). Thus, people with similar behavioral principle tie together. This brings an important aspect where, user-location coupling can be used to identify similarity patterns in mobile users. So, for the purpose of our study, to quantify similarity characteristics among mobile agents, we use their spatio-temporal preferences and preferential attachment to locations and the frequency and duration of visiting these locations. It is important to study similarity in DTN to develop behavioral space for efficient message dissemination (Hsu, et al., 2008) and design behavior-aware trust advisors among others (Kumar, et al., 2010). For efficient networking, it can help to quantify traffic patterns and develop new protocols and application to target social networking. Analysis of similarity can be used to evaluate the network transitivity, which helps to analyze macro-mobility, evolutionary characteristics and emergent properties. In this section, we introduce association matrix that captures spatio-temporal preferences and a statistical technique that use it to measures similarity among mobile users.



$$A = \begin{bmatrix} 0 & 0 & 0.5 & 0.3 & 0.2 \\ 0.2 & 0.3 & 0.4 & 0.1 & 0 \\ 0 & 0 & 0 & 0.6 & 0.4 \\ 0.2 & 0.3 & 0.1 & 0.2 & 0.2 \\ 0.1 & 0 & 0 & 0 & 0.9 \end{bmatrix}$$

Figure 3 (a) A prototype of Association Matrix. The columns represent locations (access point, building, etc) and rows represent time granularity (days, weeks, etc.). (b) A computed matrix A with 5 locations and time periods. Each entry represent the percentage online time spent at corresponding location column.

5.3.2 Capturing Spatio-Temporal Preferences

We use longitudinal wireless activity session to build mobile user's spatio-temporal profile. An anonymous sample is shown in Table-2 . Each entry of this measurement trace has the location of association and session time information for that user. The location association coupled with time dimension provides a good estimate of user online mobile activity and its physical proximity with respect to other online users (Hsu, et al., 2006; Kotz, et al., 2005). We devise a scalable representation of this information in form of an association matrix as shown in Figure 3. Each individual column corresponds to a unique location in the trace. Each row is an n-element association vector, where each entry in the vector represents the fraction of online time the mobile user spent at that location, during a certain time period (which can be flexibly chosen, such as an hour, a day, etc.). Thus for n distinct locations and t time periods, we generate a *t-by-n* size association matrix.

Representation Flexibility: The representation of spatio-temporal preferences in form of an association matrix can be changed to use each column for a building (where a collection of access points represent a building) and the time granularity can be changed to represent hourly, weekly or monthly behavior. For the purpose of our study, each row represents a day in the trace and column represents an individual access point.

5.3.3 Characterizing Association Patterns

For a succinct measure of mobile user behavior, we capture the dominant behavioral patterns by using Singular Value Decomposition (SVD) (Horn, et al., 1990) of the association matrix. SVD has several advantages:

- It helps to convert high dimensional and high variable data set to lower dimensional space there by exposing the internal structure of the original data more clearly.
- It is robust to noisy data and outliers.
- It can easily be programmed for handheld devices, which is our other on-going work.

The Singular Value Decomposition of a given matrix A can be represented as a product of three matrices: an orthogonal matrix U, a diagonal matrix S, and the transpose of an orthogonal matrix V.

$$A = U \cdot S \cdot V^T$$

where $U^T \cdot U = I = V^T \cdot V, U$ is $t - by - t$ matrix whose columns are orthonormal eigenvectors of AA^T, S is a $t - by - n$ matrix with r non-zero entries on its main diagonal containing the square roots of eigen values of matrix A in descending order of magnitude and V^T is a $n - by - n$ matrix whose columns are the orthonormal eigenvectors of $A^T A$. Thus the eigen behavior vectors of $V = \{v_1, v_2, v_3, ..., v_n\}$ summarize the important trends

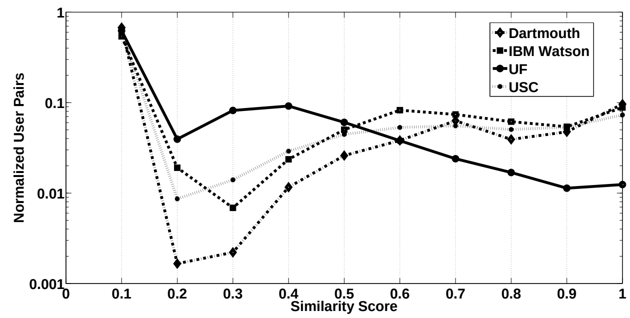


Figure 5 Log Normalized Similarity distribution of all four data sets is shown.

in the original matrix A. The singular values of $S = \{s_1, s_2, s_3, ..., s_r\}$ ordered by their magnitude $S = \{s_1 \geqslant s_2 \geqslant s_3, ... \geqslant s_r\}$. The percentage of power captured by each eigen vector of the matrix A is calculated by

$$w_i = \frac{\sum_{i=1}^{k} s_i^2}{\sum_{i=1}^{rank(A)} s_i^2}$$

It has been shown that (Hsu, et al., 2007) SVD achieves great data reduction on the original association matrix and 90% or more power for most of the users is captured by five components of the association vectors. By this result, we infer that users' few top location-visiting preferences are more dominant than the remaining ones.

5.3.4 Calculating Similarity

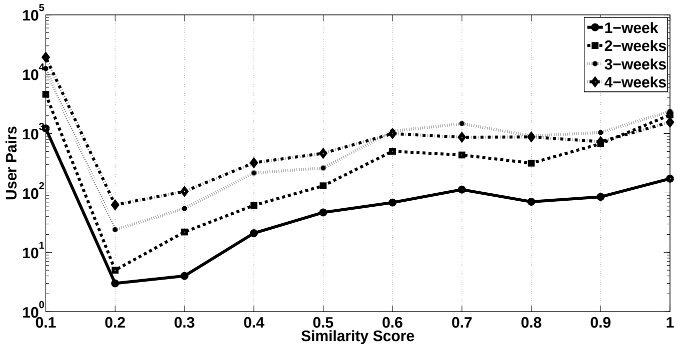
We use the eigen vectors of association matrix A to quantitatively measure the similarity between behavioral profiles of mobile user pairs. For a pair of users, with respective eigen-vectors as $X = \{x_1, x_2, x_3, ..., x_{r_x}\}$ and $Y = \{y_1, y_2, y_3, ..., y_{r_y}\}$, the behavior similarity can be calculated by the weighted sum of pair wise inner product of their eigen vectors as

$$Sim(X, Y) = \sum_{i=1}^{rank(X)} \sum_{j=1}^{rank(Y)} w_{x_i} \cdot w_{y_j} |x_i \cdot y_j|$$

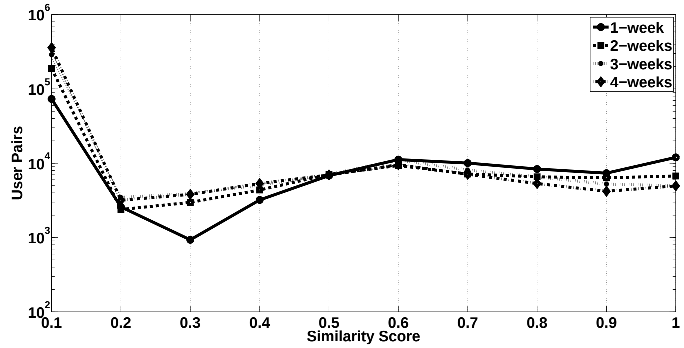
$Sim(X, Y)$ is quantitative measure index that shows the closeness of two users in spatio-temporal dimension. The value of similarity lies between $0 \leqslant Sim(X, Y) \leqslant 1$. A higher value is derived from users with similar association patterns. In this study, we are the first one to investigate the distribution of such a similarity metric among user pairs based on realistic data sets.

5.3.5 Simialrity Analysis

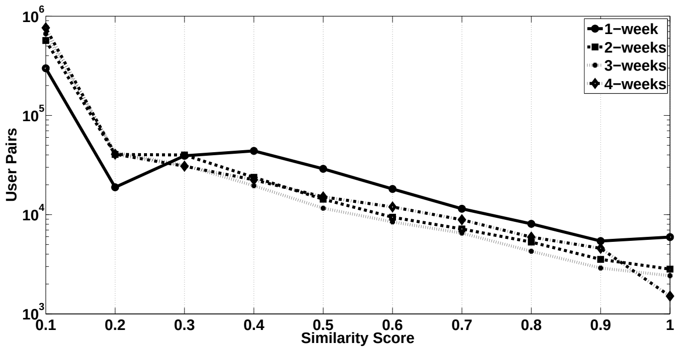
The distribution histogram of similarity scores for the campus datasets is shown in Figure 4. The figure shows number of user pairs as a function of similarity score that quantify the behavioral similarity between mobile



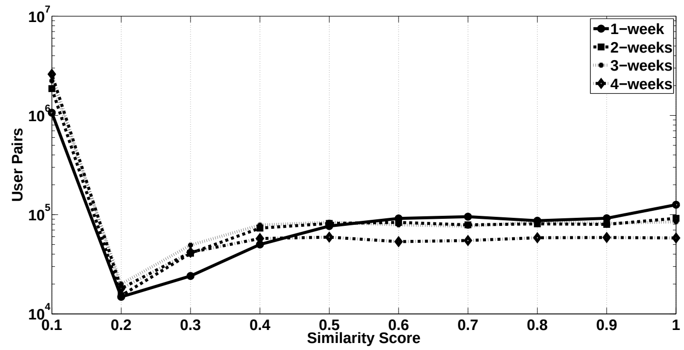
(a) Dartmouth



(b) IBM Watson



(c) UF



(d) USC

Figure 4 Similarity distribution histogram among user pairs is shown. All the four time intervals show near consistent user pair counts for a particular similarity score. Lowest similarity score (0.0 - 0.1) shows that users have very different spatio-temporal preferences. A fraction of the user pairs are also very similar with (0.9 - 1.0) similarity score.

users. We observe that: 1) mobile societies compose of users with mixed behavioral similarities, 2) For all four time periods there is a consistency and stability in the similarity score among mobile user pairs. The low similarity scores (0 - 0.1) in Figure 4 indicate a substantial portion of users is spatio-temporally very dissimilar. On the other hand, similarity scores of (0.9 1.0) suggest a statistically significant likelihood of high-density ties creating tightly knit groups. The variation in the middle shows partially similar and partially dissimilar user pairs. This is significant and provides an insight into the existence of mobile societies in the network with quite similar location visiting preferences. Overall, the curves show an assortative mixing of user pairs for all possible similarity scores. Figure 5 gives a normalized log plot to compare data sets from different campuses, and shows that similarity exists evenly across all the traces. Next, we briefly explain modularity and use a divisive algorithm to discover mobile societies in the traces.

5.3.6 Modularity

To understand the underlying structure of mobile societies (or communities), the similarity distribution is not sufficient. Therefore, we use a robust method to segregate user pairs that have high similarity score into tightly knit groups. To detect such communities in a graph like structure, a centrality-index-driven method (Newman, et al., 2006) is utilized. This measure to detect communities circumvents the traditional clustering notion to identify most central edges. Instead, a divisive algorithm is applied based on identifying least central edges, which connect most communities (via *edge betweenness*). First, the betweenness score of edges are calculated as the number of shortest paths between pair of vertices that run through it. Understandably, tightly knit communities are loosely connected by only few intergroup edges and hence shortest paths traverse these edges repeatedly, thereby increasing their respective betweenness score. If such edges are removed, according to a threshold, what we get are the groups of tightly knitted vertices known as communities. To identify a reasonable threshold value, modularity is used. Modularity is the difference of edges falling within communities and the expected number in an equivalent network with randomly placed edges (Newman, et al., 2006,a; Musolesi, et al., 2008).

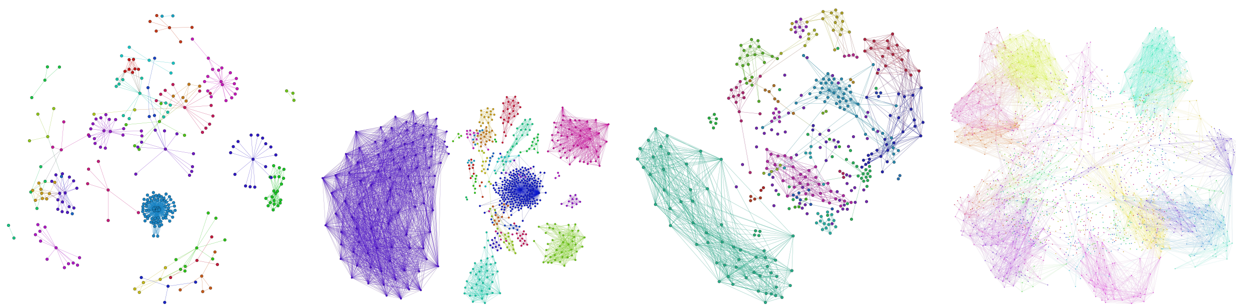
5.3.7 Detection of Mobile Societies

Human networks are known to exhibit a multitude of emergent properties that characterize the collective dynamics of a complex system (Cen, et al., 2008; Steinhaeuser, et al., 2008; Bastian, et al., 2009). Their ability to naturally evolve into groups and communities is the reason they show non-trivial clustering. Here, we consider the spatio-temporal preferences and cosine

Table 3 Network Analysis of Datasets on three different metrics

Dataset	*Clustering Coefficient*		*Average Path Length*		*Modularity*	
	Orig	*Rand*	*Orig*	*Rand*	*Orig*	*Rand*
Dartmouth	0.89	0.05	0.10	2.47	0.63	0.2
IBM Watson	0.92	0.05	0.40	2.12	0.79	0.14
UF	0.78	0.051	0.30	2.605	0.67	0.24
USC	0.91	0.05	0.19	2.0	0.46	0.11

* *Orig* = Original Dataset Graph   * *Rand* = Random Graph



(a) Dartmouth (b) IBM Watson (c) UF (d) USC

Figure 6 Shown are the structural and spatio-temporal dynamics of Mobile Societies as function of weighted cosine similarity score, produced from highly positive modularity values. Note: this figure is best viewed in color.

similarity of mobile users as a relative index to generate emergent structures, which we call mobile societies. The network transitivity structures of mobile nodes for various campus datasets are shown in Figure 6. We use mutual similarity score of mobile nodes to produce a connected graph and applied random iterations of modularity (Newman, et al., 2006,a) and betweenness algorithm to infer the mobile societies. A set of visibly segregated clusters validates their detection and presence in mobile networks.

5.3.8 Modularity Analysis for Mobile Societies

Statistically, modularity greater than 0.4 is considered meaningful in detecting community structure. For our dataset, we also find high modularity index as compared to an equivalent random graph. The comparison is shown in Table-3. Henceforth, the heterogeneity in dataset has tightly knitted Mobile Societies. This analysis further helped us to investigate the possibility of existence of different clusters of users based on their proximity in similarity score values.

5.3.9 Network Analysis for Mobile Societies

We compute the average clustering coefficient and the mean- shortest path length of these clusters. We compare the results with a random graph of the same size to understand the variation and capacity to depict small world characteristics. Table-3 delineates network properties and average modularity that provide details of the structure of mobile societies against same size random graph. The comparative values in the table clearly show that mobile societies can exhibit small world characteristics. However, we leave such small world study for future work.

Based on the above analysis, we find that similarity not only exists among mobile users, but its distributions seem to be stable for different time periods. Furthermore, this trend is consistent in all four traces, which highlights similarity clustering as an important characteristic to capture using mobility models.

5.3.10 Similarity in Model-The Missing Link

In this section, we evaluate existing mobility models and contrast their output against real trace results. Trace based mobility models (Bhattacharjee, et al., 2004; Daly, et al., 2009; Ekman, et al., 2008; Hsu, et al., 2005; Kim, et al., 2006; Lee, et al., 2006; Lelescu, et al., 2006; Rhee, et al., 2008) are a close approximation of realistic human movements and their non- homogenous behavior. They focus on vital mobility properties like nodes' on/off behavior, connectivity patterns, spatial preferences under geographical restrictions, contact duration, inter-meeting and pause time, etc. We consider two mobility models, the random direction model (a widely used classic mobility model) and Time Variant Community Model (Hsu, et al., 2009) (due to its capability to capture spatio-temporal mobility properties). In the ensuing text, we briefly describe the TVC model and use it to generate realistic movements. Finally, we compare its result against the similarity characteristic found in real measurements.

We setup the TVC model for two university campuses (IBM Watson and USC) to statistically evaluate the similarity metric established previously. Our goal is two folds:

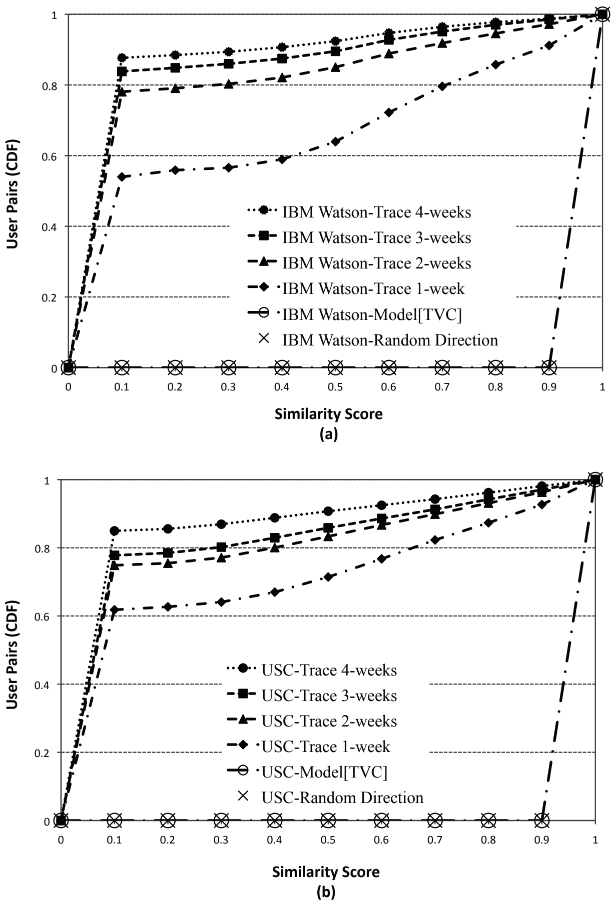


Figure 7 Cumulative distribution function of distances for the similarity score of mobile users. Real trace curves show a conformance with user pairs for different values of similarity score, while TVC and Random Direction Model has all users pairs in the 0.9 score range.

- As proposed by the TVC model, we seek to maintain the skewed location visiting preferences and time dependent mobility behavior of users.
- To analyze whether TVC model successfully captures similarity among mobile users and quantitatively simulate the distribution that we have seen in the real measurements.

5.3.11 Construction of TVC Model for Campuses

Initially, we determine the number of communities that nodes should periodically visit. We determine that top 2-3 communities capture most skewed location visiting preferences. Then we employ a weekly time schedule to capture the periodic re-visits to these major communities. To keep fair comparison against the real measurements, we configure the TVC model with same number of mobile nodes and generating measurements equivalent to one month time period with one-day granularity. Finally, for WLAN measurement we assume mobile users are stationary while being online (Hsu, et al., 2009).

5.3.12 Similarity Evaluation

TVC model accurately demonstrates location visiting preferences and periodic reappearances for both campuses(Thakur, et al., 2010). Surprisingly, it is unable to accurately capture the richness in similarity distribution on spatio-temporal basis. For all values of similarity score except 0.9, TVC and Random Direction model yields no user pairs. Figure 7 shows similarity distribution CDF curves for both campuses. We clearly observe a discrepancy between the curves from actual traces and the two mobility models (TVC and random direction). In addition, dendrograms in Figure 8 shows the result of hierarchical clustering based on users mutual similarity scores. Here, in real traces we find clusters at different similarity scores. In Figure 8(a), the average distance of 2.0 has close to 18 small clusters and Figure 8(c) shows 16 small clusters of mobile users. However, corresponding TVC dendrograms in Figure 8(b) and 8(d) show only one cluster of mobile users at a distance of 2.0. A possible explanation is that the community assignment in TVC model creates a homogeneous user population where all nodes are assigned the same communities. While it captures the location visiting and periodic preferences, it fails to differentiate among mobile nodes with different behaviors. What is missing here is a mechanism to assign different locations as the communities to different nodes, in a way that reproduces the social structure (clusters) observed in the traces.

Results in this section show that although TVC model is able to capture location visiting preferences and periodic reappearances, it does not capture the similarity metric distribution and the clusters with different behaviors in the traces. Random direction model also fails on this front in a similar way. This study realizes us that current mobility models are not fully equipped to handle behavioral metrics and community behavior of users that form mobile societies. It compels us to revisit mobility modeling in the attempt to capture both individual and community behavior of mobile users, which is part of our on-going study.

6 Routing Protocol Analysis

In this section, we compare routing protocol performance between realworld and mobility model generated traces. Our implementation of epidemic routing input time varying mobile encounter sessions. Essentially, they serve a basis for intermittently connected dynamic network topology setting where each encounter is viewed as an opportunity to receive and forward messages. We run epidemic routing against the IBM Watson and Infocom traces to measure the performance on two parameters:

1. Rechability: The percentage number of nodes that could be reached in multiple hops by a given source averaged over all nodes in the scenario.

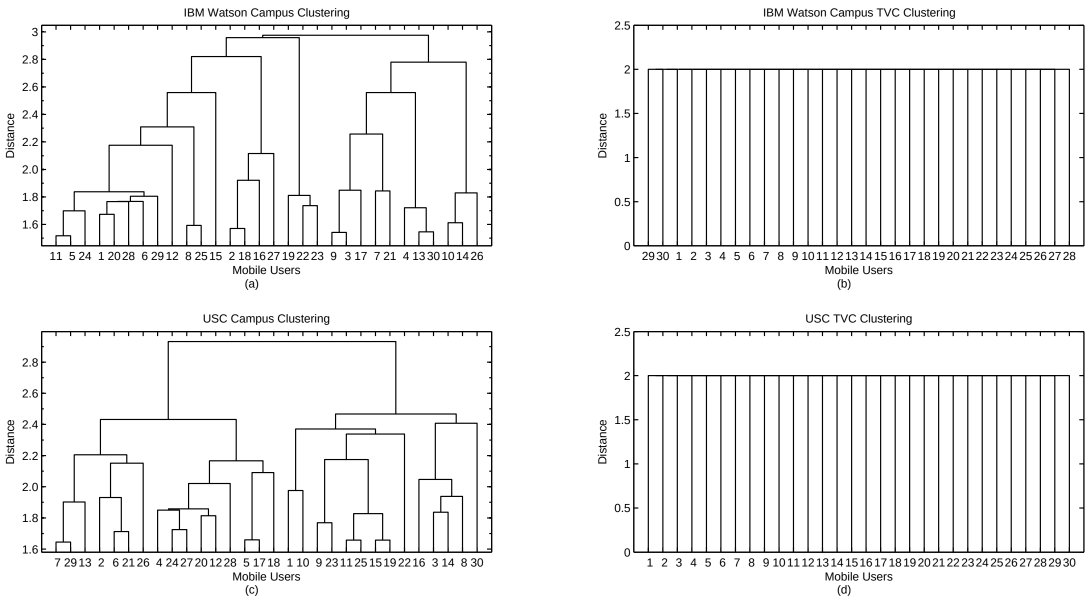
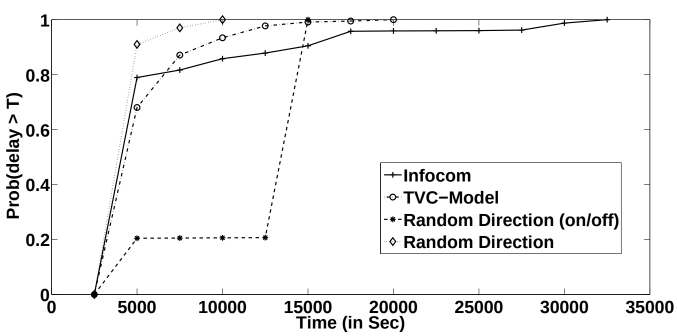
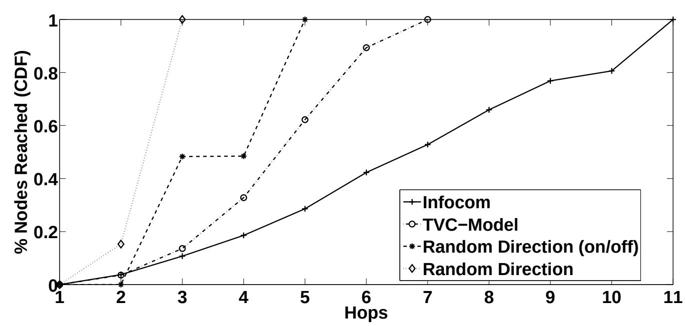


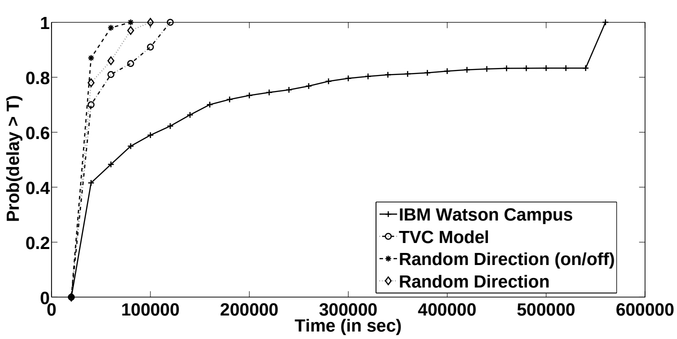
Figure 8 Dendrograms giving visual representation of two-dimensional hierarchical clustering for real and TVC model generated traces for USC and IBM Watson campus mobile users. Real traces (Figure a & c) show an incremental built-up of component based on the similarity score strength between mobile user. TVC Model (Figure b & d), output only one cluster containing all mobile users. Invariably, TVC treats all mobile users to have same preferences.



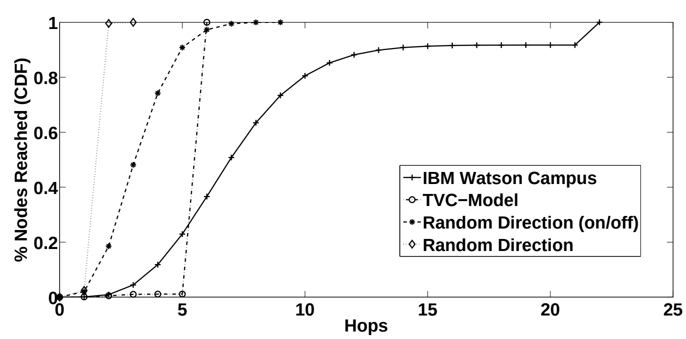
(a) Delay-Infocom



(b) Reachability - Infocom



(c) Delay - IBM



(d) Reachability-IBM

Figure 9 (a,b) Show Epidemic routing results for the Infocom settings. (c,d) Show Epidemic routing results for the IBM Watson settings. As seen, largely deviate in their network performance for delay and reachability compared to real measurement results.

2. Delay: The percentage of number of nodes that are reached in a given amount of time.

We plot the routing performance in Fig-9 to validate real and synthetic traces in all cases of reachability and delay. Surprisingly, despite models claim to exhibit vital mobility characteristics, they dramatically deviate in network routing performance benchmarks. Alongside, a quantitative report is shown in Table-4. We observe epidemic results on Infocom experiment trace takes an

Metric	Infocom iMote				IBM Watson Campus			
	O	***T***	***R1***	***R2***	***O***	***T***	***R1***	***R2***
Reachability	11	7	4	3	22	4	7	3
Delay	89	24	34	0.17	66	37	2.4	0.18
O = Real Traces; T = TVC Model; R1 = Random Direction (on/off); R2 = Random Direction.								

Table 4 Summary of performance measurement for epidemic routing.

average of 11 hops to deliver message to all other nodes; while it takes only seven in case of TVC traces and even less in case of Random Models. Meanwhile for the delay, there are at least two folds of difference between real measurement and synthetic trace. The TVC and Random models take much less time in delivering messages compared to the observed ones in the real scenario. We find similar results in case of IBM Watson traces as well. The difference in reachability is more than 15 hops between real and TVC model performance in delivery the message.

7 Discussion

Mobility models are designed with a particular scenario in mind. However, in this study we would like to question the efficacy those metrics that are widely adopted or expected to be vital in closing the performance gap between modeling and reality. Now, there is an immense need to identify them and a perception should be made to use them for correct estimation. Our results show current metrics like spatio-temporal preferences and encounter statistics are inadequate; because models completely miss out on the structure and performance criteria. We believe it is important for the researchers to search for fundamental characteristics that drive the dynamics in challenged networks. Not only we should maintain the current characteristic but also look out for structural semblance and topological realisms between simulation and similarity. A good research direction would be look into measures that affect globally in a similar way routing decision are made.

8 Conclusion

In this paper, we show that existing model demonstrate spatio-temporal and encounter statistics seen in real traces. We analyzed the spatio-temporal behavioral similarity profiles among mobile users. We define mobility profiles based on users association matrices, and then use a SVD-based-weighted-cosine similarity index to quantitatively compare these mobility profiles. Analysis of extensive WLAN traces from four major campuses reveals rich similarity distribution histograms suggesting a clustered underlying structure. Application of modularity based clustering validated and further quantified the clustered behavior in mobile societies. Similarity graphs exhibit an average modularity of 0.64, and clustering coefficient of 0.86, which indicates potential for further small world analysis.scrutinize mobility models on routing performance benchmarks. We compared similarity characteristics of the traces to those from existing common and community based mobility models to capture similarity. Surprisingly, existing models are found to generate a homogeneous community with one cluster and thus deviate dramatically from realistic similarity structures. We also testify that despite models capture realistic human behavioral patterns; their routing performance deviate from reality. We used the same synthetic traces to run epidemic routing and measure performance. By doing so, we find that mobility models performance is not analogous to reality. The synthetic mobility traces' indeed carry no structural similarity. These dramatic deviations from realism indicate serious flaws in the existing models and their inadequacy as testbed tools for any kind of performance evaluation purposes. In this paper, we limit our work in verifying epidemic routing against two well-known mobility models. In future, we are looking to test other routing protocols and models that parameterize. We elaborated on the presence of similarity among mobile users and the detection of collective behavior via community detection in wireless networks. We showed the gap between reality and current mobility models in demonstrating collective behavior. In our on-going work, we are developing a multi-dimensional mobility framework that helps scientists to develop mobility metrics and verify current models against realistic settings and provide guidelines to develop new models. We are looking into a global perspective of clustering and mobility coefficient and maintaining structural properties and performance by revisiting mobility modeling, which is vital for the evaluation and design of next-generation behavior-aware protocols.

9 Acknowledgement

This research is partially funded by Cisco Inc. and by NSF awards 0832043 and 0724658.

References

Bai, F. and Helmy A. (2008) A Survey of Mobility Modeling and Analysis in Wireless Adhoc Networks, *Wireless Ad Hoc and Sensor Networks", Kluwer Academic Publishers.*

Fan, B., Narayanan, S. and Helmy A. (2008) The IMPORTANT framework for analyzing the Impact of Mobility on Performance Of RouTing protocols for Adhoc NeTworks, *Ad Hoc Networks*, Volume 1, Issue 4, November 2003, Pages 383-403.

M. Balazinska and P. Castro. (2003) CRAWDAD data set ibm/watson (v. 2003- 02-19). http://crawdad.cs.dartmouth.edu/ibm/watson

Bhattacharjee, S. D., Rao, A., Shah, C., Shah, M. and Helmy, F. A. Empirical modeling of campus-wide pedestrian mobility, *IEEE Vehicular Technology Conf (VTC)*

Bastian, M., Heymann, S. and Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. *International AAAI Conf. on Weblogs and Social Media*

Burleigh, S., Hooke, A., Torgerson, L., Fall, K., Cerf, V., Durst, B., Scott, K. and Weiss, H. (2007) Delay-tolerant networking: an approach to interplanetary Internet., *IEEE Communications Magazine*, vol.41, no.6, pp. 128-136.

Camp, T., Boleng, J., Davies, V. 2002) A survey of mobility models for ad hoc network research, *Wireless Communications and Mobile Computing*.

Chaintreau, A., Hui, P., Crowcroft, J., Diot, C., Gass, R. and Scott, J. (2007) Impact of Human Mobility on Opportunistic Forwarding Algorithms, *IEEE Transactions on Mobile Computing*, Pages 606-620.

J., C., M.C., G., P., W., T., S., G., M. and A.-L., B. Uncovering individual and collective human dynamics from mobile phone records. *Jrnl. of Physics*.

Costa, P., Mascolo, C., Musolesi, M. and Picco, G. P. Socially-aware Routing for Publish-Subscribe in Delay-tolerant Mobile Ad Hoc Networks. *IEEE Jrnl. Communications*

Daly, E. M. and Haahr, M. Social Network Analysis for Information Flow in Disconnected Delay-Tolerant MANETs. *IEEE Tran. on Mob. Comp*, Pages 606-621

Eagle, N. and Pentland, A. (2006) Eigenbehaviors: Identifying Structure in Routine. *Proc. Roy. Soc. A*

Eagle, N. and Pentland, A. Eigenbehaviors: Identifying Structure in Routine. *Proc. Roy. Soc. A*

Eagle, N., Pentland, A. and Lazer, D. Inferring friendship network structure by using mobile phone data. *PNAS*.

Ekman, F., Keranen, A., Karvo, J. and Ott, J. Working day movement model. *ACM workshop on Mobility models*

Fall, K. (2002) A delay-tolerant network architecture for challenged Internets. *ACM SIGCOMM*.

Garbinato, B., Miranda, H. and Rodrigues, L. Middleware for Network Eccentric and Mobile Applications. *Springer*.

Girvan, M. and Newman, M. E. J. (2002) Community Structure in Social and Biological Networks. *PNAS*

NOMADS Lab. MobiLib: Community-wide Library of Mobility and Wireless Networks Measurements

Horn, R. A., Matrix Analysis. *Cambridge Univ. Press*.

Wei-jen Hsu and Helmy, A. On Modeling User Associations in Wireless LAN Traces on University Campuses. *IEEE Int.l Workshop on Wireless Network Measurements*.

Hsu, W., Merchant, K., Shu, H., Hsu, C. and Helmy, A. Weighted waypoint mobility model and its impact on ad hoc networks. *ACM SIGMOBILE*, Pages 59-63.

Hsu, W., Dutta, D. and Helmy, A. (2008) CSI: A Paradigm for Behavior- oriented Delivery Services in Mobile Human Networks. *IEEE ToN*

Hsu, W., Dutta, D. and Helmy, A. (2007) Mining behavioral groups in large wireless LANs. *ACM MobiCom*, Pages 338-341

Hsu, W. and Helmy, A. (2006) On nodal encounter patterns in wireless LAN traces. *IEEE WiNMee*

Hsu, W., Spyropoulos, T., Psounis, K. and Helmy, A. (2009) Modeling spatial and temporal dependencies of user mobility in wireless mobile networks. *IEEE/ACM Trans.Networking*, Vol17, 5, Pages 1564-1577

Hui, P ., Crowcroft, J. and Yoneki, E. (2008) Bubble rap: social-based forwarding in delay tolerant networks. *MobiHoc,* , Pages 241-250

Hui, P., Yoneki, E., Chan, S. Y. and Crowcroft, J. Distributed community detection in delay tolerant networks. *ACM MobiArch*.

Karagiannis, T., Le B., J. and V. Milan. (2007) Power law and exponential decay of inter contact times between mobile devices, *ACM MobiCom*

Kim, M. and Kotz, D. (2007) Periodic properties of user mobility and access- point popularity. *Personal Ubiquitous Comput.er*, Vol 11, 6, Pages 465-479

Kim, M. and Kotz, D. Periodic properties of user mobility and access-point popularity. *Personal Ubiquitous Computing*.

Kim, M., Kotz, D. and Kim, S. Extracting mobility model from real user traces. *IEEE INFOCOM*.

Kotz, D. and Henderson, T. (2005) CRAWDAD: A Community Resource for Archiving Wireless Data at Dartmouth. *IEEE Pervasive Computing*

Kotz, D. and Essien, K. Analysis of a Campus-wide Wireless Network. *Wireless Networks*

Kumar, U., Thakur, G. and Helmy, A. PROTECT: Proximity- based Trust-advisor using Encounters for Mobile Societies. *ACM IWCMC*

Lee, J. and Hou, J. C. Modeling steady-state and transient behaviors of user mobility: formulation, analysis, and application *ACM MobiHoc*

Lelescu, D., Kozat, U. C., Jain, R. and Balakrishnan, M. Model T++, *ACM MobiHoc*

McPherson, M., Smith-Lovin, L. and Cook, J. M. Birds of a Feather: Homophily in Social Networks. *Rev. of Sociology*.

Mtibaa, A., May, M., Diot, C. and Ammar, M. PeopleRank: Social Opportunistic Forwarding. *IEEE INFOCOM*.

Musolesi, M., Hui, P., Mascolo, C. and Crowcroft, J. Writing on the Clean Slate, *Intl Workshop on Auto. Opp. Comm.*

Musolesi, M. and Mascolo, C. (2006) A community based mobility model for ad hoc network research. *REALMAN*, Pages 31-38

Nazir, F., Prendinger, H. and Seneviratne, A. Participatory mobile social network simulation environment. *ACM MobiOpp* .

Nelson, S. C., Harris, I.,Albert F. and Kravets, R. (2007) -driven, role- based mobility in disaster recovery networks. *In Proceedings of the 2nd ACM workshop on Challenged networks (ACM Mobicom CHANTS 07)*

Newman, M. E. J. Modularity and community structure in networks. *PNAS*.

Newman, M., Barabasi, A. and Watts, D. J. The Structure and Dynamics of Networks: *Princeton University Press*.

Reddy S., Estrin, D. and Srivastava, M., Recruitment Framework for Participatory Sensing Data, *IEEE ICPC*.

Rhee, I., Shin, M., Hong, S., Lee, K. and Chong, S. On the Levy-Walk Nature of Human Mobility, *IEEE INFOCOM*.

Royer, E. M., Melliar-smith, P. M. and Moser, L. E. An Analysis of the Optimum Node Density for Ad hoc Mobile Networks. *IEEE ICC*

Scott J. , Gass R. , Crowcroft J., Hui, P., Diot, C. and Chaintreau, A. (2009) CRAWDAD data set cambridge/haggle (v. 2009-05-29), Downloaded from http://crawdad.cs.dartmouth.edu/cambridge/haggle

Schutt, R. K. Investigating the Social World: The Process and Practice of Research. *Pine Forge Press*.

Shilton, K., Ramanathan, N., Reddy, S., Samanta, V., Burke, J. A., Estrin, D., Hansen, M. and Srivastava, M. B. Participatory Design of Sensing Networks: Strengths and Challenges. *Participatory Design Conference, Bloomington*.

Spyropoulos, T., Psounis, K. and Raghavendra, C. S. Performance analysis of mobility-assisted routing. *ACM MobiHoc*.

Spyropoulos, T., Psounis, K. and Raghavendra, C. S. (2006) Performance analysis of mobility-assisted routing. *ACM MobiHoc*, Pages 49-60

Steinhaeuser, K. and Chawla, N. V. Community Detection in a Large Real-World Social Network. Social Computing, Behavioral Modeling, and Prediction, *Springer*.

Thakur S. G, Helmy, A., and Hsu, W. (2010) . Similarity analysis and modeling in mobile societies: the missing link. *In Proceedings of the 5th ACM workshop on Challenged networks (ACM CHANTS 10)*

Vahdat, A. and Becker, D. (2000) Epidemic Routing for Partially Connected Ad Hoc Networks. *CS-200006. Duke University*

Yoneki, E., Hui, P. and Crowcroft, J. Visualizing community detection in opportunistic networks. *ACM CHANTS*.

Zachary, W. W. An information flow model for conflict and fission in small groups. *Jr. of Anthr. Res.*, Pages 452-473.

Zhang, X., Kurose, J., Levine, B. N., Towsley, D. and Zhang, H. (2007) Study of a Bus-Based Disruption Tolerant Network: Mobility Modeling and Impact on Routing, *ACM Mobicom*, Pages 195-206.