On Profiling Mobility and Predicting Locations of Campus-wide Wireless Network Users

Joy Ghosh, Matthew J. Beal, Hung Q. Ngo, Chunming Qiao
Department of Computer Science and Engineering
University at Buffalo, The State University of New York
201 Bell Hall, Buffalo, NY 14260-2000
Email:{joyghosh, mbeal, hungngo, qiao}@cse.buffalo.edu

***Abstract*— In this paper, we analyze a year long wireless network users' mobility trace data collected on ETH Zurich campus. Unlike earlier work in [9], [21], [35], we profile the movement pattern of wireless users and predict their locations. More specifically, we show that each network user regularly visits a list of places, such as a building (also referred to as "hubs") with some probability. The daily list of hubs, along with their corresponding visit probabilities, are referred to as a *mobility profile*. We also show that over a period of time (e.g., a week), a user may repeatedly follow a mixture of mobility profiles with certain probabilities associated with each of the profiles. Our analysis of the mobility trace data not only validate the existence of our so-called sociological orbits [13], but also demonstrate the advantages of exploiting it in performing *hub-level location predictions*. Moreover, such profile based location predictions are found not only to be more precise than a common statistical approach based on observed hub visitation frequencies, but also shown to incur a much lower overhead. We further illustrate the benefit of profiling users' mobility by discussing relevant work and suggesting applications in different types of wireless networks, including mobile ad hoc networks.**

***Index Terms*— WLAN mobility trace analysis, Sociological orbits, Mobility profiles, Location prediction, Mobile wireless networks**

I. INTRODUCTION

The mobility of users forming a mobile wireless network causes changes in the network connectivity and may even lead to intermittently connected networks. On one hand, nodal mobility may increase the overall network capacity [15]. On the other hand, it may make it challenging to locate users and route messages within the network.

Many researchers have tried to model practical mobility in various ways to achieve different goals. Earlier work on mobility modeling [8] was done mostly with Mobile Ad hoc NETworks (MANET) in mind. For example, some [27] used mobility pattern analysis to minimize radio link changes via appropriate selection of next hop within radio range. While the authors in [32], [36] performed physical location prediction via continuous short-term and short-range tracking of user movement, we had leveraged on our assumptions on "sociological orbits" to perform efficient routing within MANETs [13], [14]. More recently, Intermittently Connected Mobile Ad hoc Networks (ICMAN) (or in general, Delay Tolerant Networks (DTN)) have received a lot of interest. For example, researchers [7], [38] have suggested the concept of controlled mobility to aid in mobile ad hoc routing. Literature has also proposed several work on mobility trace analysis within campus-wide wireless networks, which we shall discuss at length in Section VI to highlight the novelty of our work.

Our study of user mobility traces is motivated by the need to extract practical mobility information, which may potentially benefit applications such as location approximation and routing within all types of networks such as MANETs, DTNs, wireless access networks, etc. More specifically, it is noted that wireless users belong to a larger social environment and as such, their movement behavior is subject to several location dependent sociological constraints (in addition to speed limits and specific walkways, as described in [4]). In particular, on any given day, each user may visit a list of places of some social importance (which we referred to as "hubs" in [13]) in some probabilistic manner, creating what we refer to in this paper as a "mobility profile".

In this work, we not only validate the existence of such mobility profiles via mobility trace analysis, but also show that in practice, a user is usually associated with more than one profile. In particular, we find evidence of a probabilistic mixture of such profiles that stay valid for a long period of time (i.e., several days or weeks), after which a different mixture of profiles will be in effect. The data analyzed in this paper is collected on the ETH Zurich campus and is similar in content to that available from the Dartmouth campus (both are obtained from Access Point system logs). However, compared to the most related (and yet much different) work in [9], this paper primarily focusses on the orbital parameters, in particular on the *user-centric* parameters like the user mobility profiles and its applications, whereas [9] focusses more on AP-centric parameters. Our mobility profile based hub-level location prediction is shown not only to be more precise than a common statistical method, but also to incur a much lower overhead. Note that although this work analyzes data from a campus-wide wireless access network (instead of a MANET, as data from former is more readily available), our mobility profiling and location prediction techniques are applicable to other types of networks as well, since the movement of users is ultimately influenced by their social environment.

The rest of this paper is organized as follows. In Section II, we discuss a sociological orbit framework and describe the major orbit (hub-centric and user-centric) parameters to be studied. In Section III, we study the distributions of the hub-centric parameters from the analyzed data. In Section IV, we study the user-centric parameters and in particular, we present a clustering algorithm using a *Mixture of Bernoulli's* to analyze user mobility profiles. In Section V, we highlight the advantages of profiling users' mobility by comparing profile based hub-level location predictions to predictions based on a general statistical method. In Section VI, we uphold the novelty of our work by comparing with other related work on the analysis of wireless network users' mobility traces, and also discuss potential applications that may benefit from our work. Finally, we conclude this work in Section VII.

II. SOCIOLOGICAL ORBIT FRAMEWORK

In this section, we briefly describe and enhance the sociological orbit framework we first proposed in [13]. In the real world, it is observed that users routinely spend a considerable amount of time at a few specific place(s), referred to as hub(s). For example, in a WLAN scenario a hub may be a floor within a building or, the entire building itself, depending on the scale of the network model. Although, it is hard (and may be even against privacy policies) to keep track of an individual at all times, one can still take advantage of the fact that most users' movements are within, and in between, a list of hubs.

Let us consider a graduate student who only has classes on Monday, Wednesday and Friday, when he is found on a school campus, spending most of his/her time in either his/her laboratory, a seminar room, or the cafeteria, each of which shall form a "hub" in this example (as shown in Figure 1). The actual list of hubs

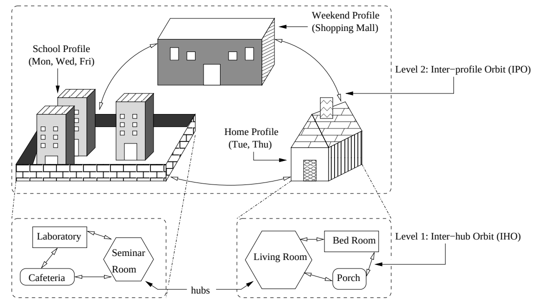


Fig. 1. A hierarchical view of sociological orbits

visited by the student on the same day is called a "hub list". Even if such hub lists may vary from one day to another, that variation is only marginal (as shown later in Section IV). In most cases, a number of hub lists over a period of days may be clustered together and represented by a single "weighted hub list", where the weight associated with each hub denotes the probability of the student visiting that hub within that period. In this work, we shall refer to such a weighted hub list to be a user's "Mobility Profile", and the movement in between the hubs within a profile as an "Inter-hub Orbit" (IHO). If one wishes to locate the student on a school day, knowing this *School Profile* within the campus shall be helpful, where one can most probably find him/her in either the laboratory, or the seminar room, or the cafeteria, without having to look all over the campus.

In real life, it is observed (and later verified from the analyzed data) that a user over long periods of time is usually associated with more than one mobility profile, mixed with certain probabilities. This is shown in Figure 1 as the *Weekend Profile* and the *Home Profile* to account for the student's remainder of the week. Such a movement in between multiple profiles at a higher level is referred to as the "Inter-profile Orbit" (IPO). Over different periods of time, this mixture of profiles may change, causing what we call an "IPO Timeout". The IPO and the IHO together constitute the hierarchical *sociological orbit* at two different levels. In this paper, such orbital mobility information is shown to be most helpful in predicting the hub-level location of a user with much more accuracy than a general statistical method at a much lower overhead.

To formalize the sociological orbit framework, we divide the orbital parameters into two categories: *Hub-centric*, and *User-centric*, as listed in Table I. On the *hub-centric* side, the *Hub Form* depends on the actual definition of a hub in the network being modeled; *Hub Visits* denotes the number of users visiting a hub in a given period; and the *Hub Stay Time* is the amount of

TABLE I
ORBITAL PARAMETERS

Category	Parameters
Hub-centric	Hub Form
	Hub Visits
	Hub Stay Time
User-centric	Mobility Profiles
	Hub List Size

time a user spends at one stretch within a hub. On the *user-centric* side, the *Mobility Profile Parameters* include a list of hubs and their corresponding weights, and the *Hub List Size* refers to the number of unique hubs visited by a user on a day. Although, the study in [9] is similar to our study of the hub-centric parameters, to the best of our knowledge this paper is the first to discuss and analyze the user-centric aspects related to sociological orbits.

In the following sections, we analyze wireless network users' mobility trace data collected on ETH Zurich campus from 1st April, 2004 till 31st March, 2005. There were a total of $13,620$ users, 43 buildings, and 391 Access Points (AP). The data was obtained as system logs from the APs which recorded the *association, disassociation, missed polls*, and *roaming* events for users during the given period. First, to study the observed distribution of the *hub-centric* parameters of the framework, we setup an Oracle database with these traces and employed standard SQL queries[1]. Second, to analyze the *user-centric* parameters we employ a clustering algorithm using a *Mixture of Bernoulli's*. We also develop efficient methods to model and analyze mobility profiles to validate the existence of sociological orbits. Finally, we use the mobility profiles to do hub-level location predictions more precisely than a statistical method, and at much lower overhead.

III. HUB-CENTRIC PARAMETER DISTRIBUTIONS

In this section, we shall define and discuss the major parameters related to the concept of "hubs".

A. Hub Form

Hubs are generally defined as places of social importance. However, the hub form (e.g., size, shape) is mostly related to the type and scale of the network of interest. For instance, in a campus-wide wireless network (like the one we study in this paper), one could consider each AP, or room, or floor, or the entire building to be a hub. Different choices are driven by the specific interest in the granularity of the movement of the users. For example, if one is only interested in identifying one building out of many (which say are located far apart on the same campus) to locate a user, and not the specific room or, the floor within the building (which say has a local network connecting all the rooms), then each building should be considered as a hub. As such, this decision also affects the time spent in each hub, and the total number of hubs in the model. A reasonable approach to decide upon a suitable hub form is to maximize the benefit of a high level orbital information, without having to monitor/update short-term or, short-range mobility information.

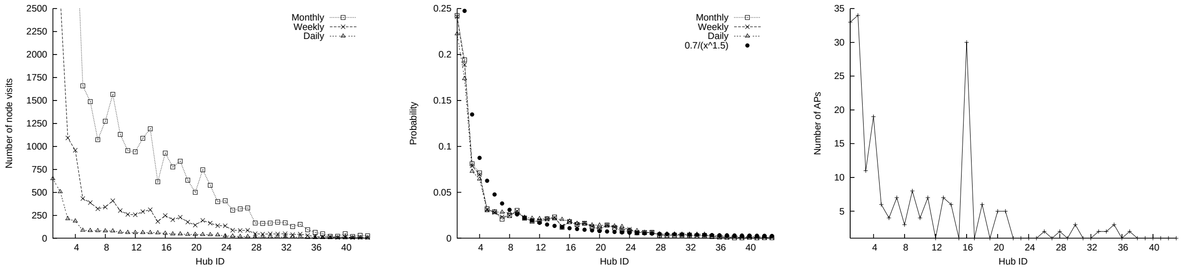
Note that in a broadly defined network of people, a hub is not required to have an AP (unlike in [9]), in which case a hub may still be identified in a variety of ways. The use of GPS service is the obvious first choice. Alternatively, in the broader contexts of *pervasive/ubiquitous computing* [1], and *Ambient Intelligence* (AmI) [26], localization in a cosmopolitan area will be even more readily available.

For the data we analyze here, we consider each building within the entire ETH Zurich campus to be a hub. There are a total of 43 buildings/hubs that we analyze in the traces. We assign an unique "Hub ID" to each hub. Some of these are academic buildings, libraries, cafeterias, etc. Each building, has a number of floors, and each floor has multiple rooms, some of which have APs within them. Accordingly, a hub in our case may be covered by multiple APs.

B. Hub Visits

In Figure 2(a), we plot the frequency with which network users visit each hub on a daily, weekly, and monthly basis to study the degree of social significance for each hub. As can be seen from Figure 2(a), only a few hubs (hub IDs 1 through 4) record large number of user visits, making them most "socially popular". We find the fraction of total node visits recorded by each hub to follow a power law distribution as shown in Figure 2(b). Also, the distributions for daily, weekly, and monthly visits are observed to be almost identical, indicating a temporal consistency of the social significance of each hub. For example, on either daily, weekly or, monthly basis, hub 2 is always seen to have a hub visitation probability of around 0.19. In Figure 2(c), we plot the number of APs in each hub. Interestingly, with the exception of hub 16, we observe the number of APs in each hub to be proportional to the number of its user visits. Larger number of APs could either

[1] We thank Nirmal Thangaraj for developing the SQL queries



(a) Hub Visit Frequency (b) Hub Visit Probability (c) AP Distribution within Hubs

Fig. 2. Hub visitations and AP distribution within hubs

indicate a larger size hub, or a hub with greater need to support higher network load and connectivity, all of which could indicate a hub's social significance. Overall, it is evident from Figure 2 that the number of popular hubs grows *very slowly* with the total number of hubs, and such information may be useful for efficient information dissemination within a network by directing information to one of these popular hubs first.

C. Hub Stay Time

The Hub Stay Time parameter signifies the absolute time in between an *association* event, and either a *disassociation* or, *missed poll* or, a *roaming* event as recorded by an AP for a user. We prune all values larger than 48 hours as they could indicate errors in the system logs, and plot all other values in Figures 3(a) and 3(b). As seen in Figure 3(a), hub stay time of 10 minutes occurred most often, and there is a sharp drop and gradual decrease for 20 minutes or more. Figure 3(b) plots the hub stay time distribution at the hourly level, which is shown to follow a power law distribution. Most users are seen to stay in a hub for 4 hours or less, with the percentage of people staying in a hub for more than 12 hours being significantly lower. This fits in well with the diurnal cycle of mobility of casual network users (i.e., work/roam in day, and stay at home at night). Figure 3(c) displays the average hub stay time recorded by each hub. It is interesting to note that the hubs that recorded a lower number of user visits in Figure 2(b) seem to have higher average hub stay times. This may be due to the fact that while some hubs (e.g., cafeteria) record high number of shorter visits, others (e.g., library) may record lower number of longer visits. Such social influences on mobility patterns of users around hubs may be efficiently leveraged upon by applications such as routing.

IV. Analysis of User-centric parameters

In this section, we shall analyze the *user-centric* parameters by examining individual network user's movement. To help select appropriate sample users, we first divide all the users in different user groups based on the number of days they are found to be "active" within the network (i.e., associated with at least one AP in the day). In Figure 4, we plot the fraction of total population vs. the number of their active days. The x-axis shows a range of values, i.e., 25 denotes up to 25 active days, 50 denotes anywhere between 26 and 50 active days, and so on. 80% of the total population is seen to be active for only 25 days or less in an year. Based on Figure 4, we

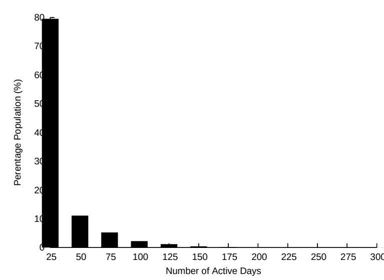
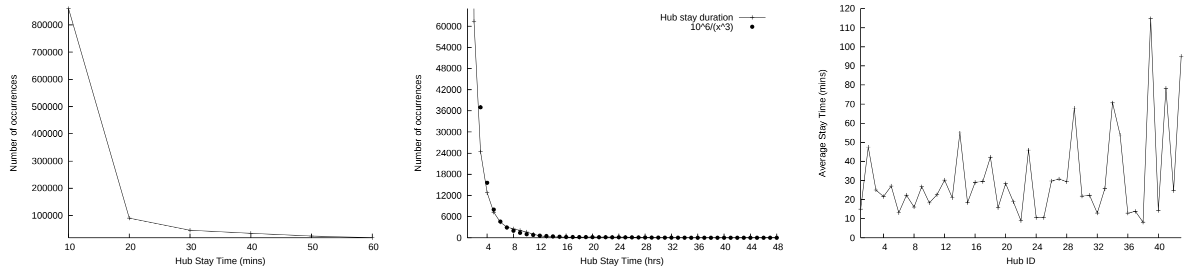


Fig. 4. Number of active days for users

categorize users in 6 different groups (G_j) as follows:

- G_1 : Active for 0 to 25 days
- G_2 : Active for 26 to 50 days
- G_3 : Active for 51 to 75 days
- G_4 : Active for 76 to 100 days
- G_5 : Active for 101 to 125 days
- G_6 : Active for 126 to 150 days

We wish to choose one user to represent each group who is the "most active" within that group (i.e., we wish to maximize both the number of active days and the the hub list size within each group). For a given group G_j, let $D^j_{\max}$ and $L^j_{\max}$ be the maximum number of active days and maximum average hub list size respectively, across all users in G_j. Let the pair D^j_i and L^j_i denote the number of active days and the average hub list size for a particular user i in G_j. Then, to represent group



(a) Hub Stay Time periods ≤ 1 Hour (b) Hub Stay Time periods ≤ 48 Hours (c) Average Hub Stay Time

Fig. 3. Hub stay time distributions

G_j we need to find a user who can *minimize*

$$\alpha.\frac{D^j_{\max} - D^j_i}{D^j_{\max}} + \beta.\frac{L^j_{\max} - L^j_i}{L^j_{\max}} \tag{1}$$

where, α and β are weights associated with each term. The results of using (1) with $\alpha = 1$ and $\beta = 1$ (we weigh both the number of active days and the hub list size equally) is summarized in Table II. The basic intuition behind selecting the "most active" user from each group is the availability of more statistically significant mobility data for such an individual. At the same time, studying users from different groups help represent most of the total population as seen in Figure 4. Alternately, one may also select the sample users from each group to find users that are either "least active" or, "active on average".

TABLE II
SAMPLE USERS FROM ALL GROUPS

$Group$	MAC	D_i	L_i
G_1	0004.2396.92ab	24	2.29
G_2	0004.2398.82c0	71	4.08
G_4	0020.e089.9376	98	2.46
G_5	0004.2396.8ced	119	2.13
G_6	0005.4e41.cf1d	126	2.63

A. Model for Analysis of Mobility Profiles

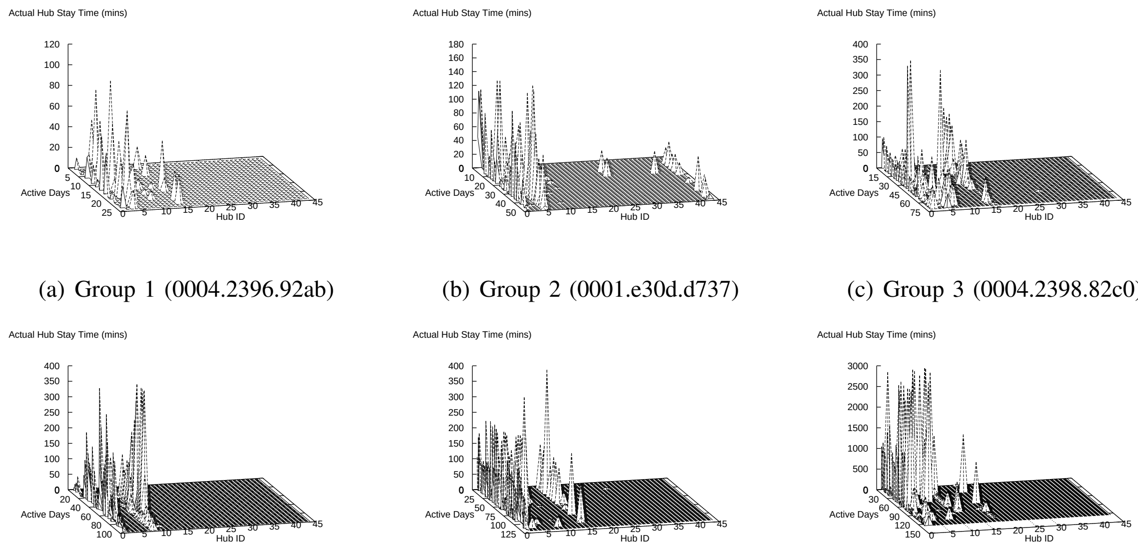
We now present a study on the mobility of these 6 sample users. We first plot their hub stay times in all the hubs during their active period, as shown in the 3-D Figure 5. To filter out noises (i.e., very brief hub stay durations), we run a horizontal plane parallel to the threshold value of 5 minutes across the plots shown in Figure 5 to obtain 2-D plots in Figure 6 showing only which hub(s) is(are) visited by a user (for more than 5 minutes) on a given day.

Next, we define an n dimensional space, where each dimension refers to a hub (i.e., $n = 43$ in our case). The hub list for a user in any given day (which is nothing but a binary vector of hub visits) may then be represented by a point in this space. For a particular user, similar hub lists on different days would generate several overlapping points whereas, two hub lists that differed only in terms of one or two hubs would generate points "close" to each other in this space. We use a clustering algorithm that helps define this concept of "closeness" by considering hub lists that say only differ in a maximum of 1 or, 2 hubs to be "close" and to belong to the same cluster. The mean of the cluster, which is a weighted hub list, then represents a mobility profile, as is described in more detail below.

B. Using a Mixture of Bernoulli's to Profile Mobility

A suitable choice to model the binary hub visitation vectors is a *Mixture of Bernoulli* distribution. In this mixture model there shall be more than one mixture component where, each component is considered an unique mobility profile represented by the component mean. Thus, a profile is nothing but a distribution over the hub visitation probabilities (i.e., a weighted hub list). We refrained from using the commonly used *Mixture of Gaussian* model because the domain of the Gaussian variable, being $(-\infty, \infty)$, is clearly not suitable for binary valued vectors. Assuming that the current mobility profile of a user is known, we model each hub visitation by a user as an independent event. On the other hand, if the current profile is not known, the general probability of a user visiting a hub is dependent on the probability associated with each mobility profile. The latter fact is crucial, since it allows for the knowledge of a user's hub visits to help infer the current mobility profile and therefore the probabilities of visits to other hubs on the same and future days, as shown later in Section V.

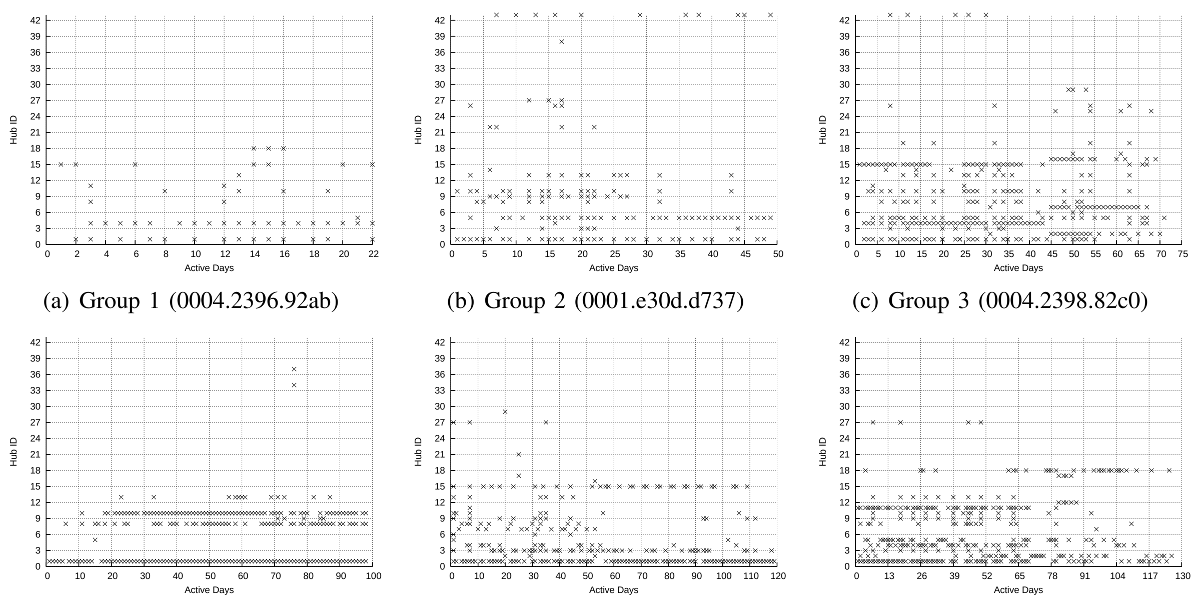
More formally, we use h ($1 \leq h \leq H$) to denote the unique hub id and i ($1 \leq i \leq n$) to denote the



(a) Group 1 (0004.2396.92ab) (b) Group 2 (0001.e30d.d737) (c) Group 3 (0004.2398.82c0)

(d) Group 4 (0020.e089.9376) (e) Group 5 (0004.2396.8ced) (f) Group 6 (0005.4e41.cf1d)

Fig. 5. Hub stay time distribution of all sample users



(a) Group 1 (0004.2396.92ab) (b) Group 2 (0001.e30d.d737) (c) Group 3 (0004.2398.82c0)

(d) Group 4 (0020.e089.9376) (e) Group 5 (0004.2396.8ced) (f) Group 6 (0005.4e41.cf1d)

Fig. 6. Daily hub visitation patterns of all sample users

day index, where H and n are the total number of hubs and days respectively. On each day i, we define a user's hub list to be a binary vector of hub associations $\mathbf{y}^{(i)} = [\mathbf{y}_1^{(i)}, \ldots, \mathbf{y}_H^{(i)}]$ where each element $\mathbf{y}_h^{(i)} \in \{0, 1\}$ such that $\mathbf{y}_h^{(i)}$ is equal to 1 if hub h was visited on day i, and zero otherwise. We denote the complete trace of hub visits across all n days with the symbol Y, which is the collection $Y = \{\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(n)}\}$. The total probability of Y is given by the product of a mixture of independent Bernoulli distributions as follows:

$$p(Y) = \prod_{i=1}^{n} p(\mathbf{y}^{(i)}) \ ,$$

where,

$$p(\mathbf{y}^{(i)}) = \sum_{j=1}^{k} p(j) \prod_{h=1}^{H} p(\mathbf{y}_h^{(i)} | \rho_{j,h}) \ .$$

Here, k is the number of mixture components (or, mobility profiles); $p(j)$ is the probability of following profile j; $\rho_{j,h}$ is the probability of visiting hub h on a day when following profile j. This framework is a generative Bayesian model in the sense that it defines a probability to every possible outcome, or pattern, that can be produced for Y.

This mixture model is trained using the Expectation-Maximization (EM) algorithm of Dempster, Laird and Rubin [12]. By employing consecutive Expectation (E)- and Maximization (M)- steps, the probability of the entire data set Y is guaranteed to monotonically increase (or, remain the same). The E-step consists of computing the posterior probability of membership of a datum (or, hub list) across the k mixture components (or, mobility profiles). Intuitively, at this E-step we look at each hub list and try to guess the mobility profile being followed on that particular day. Formally, this corresponds to computing the *responsibilities* of each component in the mixture, denoted by $r_j^{(i)}$, such that $\sum_{j=1}^{k} r_j^{(i)} = 1$, and are found using Bayes' theorem:

$$\textbf{E-step} \qquad r_j^{(i)} \equiv p(j|\mathbf{y}^{(i)}) = \frac{p(j)p(\mathbf{y}^{(i)}|j)}{p(\mathbf{y}^{(i)})} \qquad (2)$$
$$\forall i = 1, \ldots, n \quad \text{and} \quad \forall j = 1, \ldots, k \ .$$

The M-step of the EM algorithm updates the parameters of each of the k components of the mixture model, in light of the responsibilities $r_j^{(i)}$ computed in the E-step. In other words, at this M-step we look at the probabilistic associations of the hub lists with each profile computed in the E-step, and update both the probabilities associated with each profile (i.e., mixing proportions), and the probabilities associated with each hub visitation within a profile. Thus, formally the parameters of the mixture model are: the mixing proportions, denoted by vector $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_k)$ where $\pi_k = p(k)$ such that $\sum_{j=1}^{k} \pi_j = 1$; and for each mixture component j, there is a vector of dimension H of probabilities of each hub being used, denoted by $\boldsymbol{\rho}^{(j)} = (\rho_{j,1}, \ldots, \rho_{j,H})$. Thus each component in the mixture represents a mode of a user's interaction with a subset of the H hubs available (i.e., each profile is nothing but a weighted hub list). The updates to the parameters in the M-step are as follows:

$$\textbf{M-step}, \boldsymbol{\pi} \qquad \pi_j = \frac{1}{n} \sum_{i=1}^{n} r_j^{(i)} \qquad (3)$$
$$\forall j = 1, \ldots, k \ .$$

and

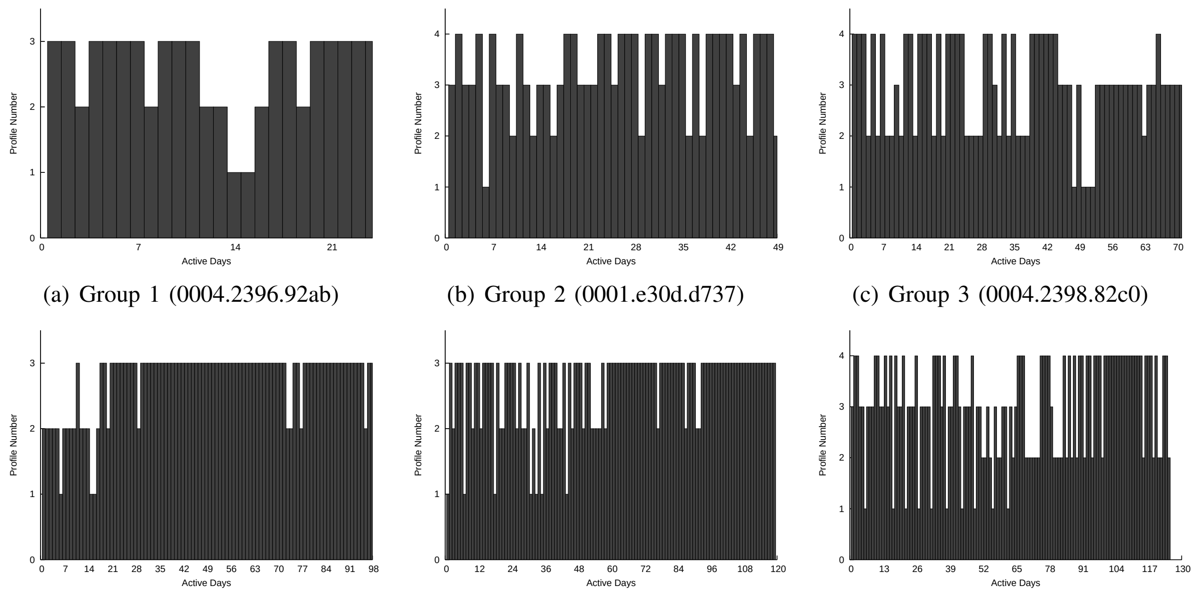
$$\textbf{M-step}, \boldsymbol{\rho} \qquad \rho_{j,h} = \frac{\sum_{i=1}^{n} r_j^{(i)} \mathbf{y}_h^{(i)}}{\sum_{i=1}^{n} r_j^{(i)}} \qquad (4)$$
$$\forall j = 1, \ldots, k \quad \text{and} \quad \forall h = 1, \ldots, H \ .$$

For each user, one may choose the number of components (i.e, profiles) for each mixture model by visual inspection of the data distribution. An alternate approach may include approximate Bayesian model selection techniques, e.g. via the Bayesian Information Criterion (BIC; [29]) or, other criteria. In this work, we run the clustering algorithm for each sample user multiple times with different number of randomly initialized cluster means (i.e., profiles), and select the one where each profile has moderate associativity with hub lists. Figure 7 shows the pattern of mobility profiles over all the days. Table III lists both the probability that a user is in a given profile, and the probability that a hub is visited when following a particular profile. As an example, from Figure 7(a) we find that the sample user from group G_1 is following his/her mobility profile 1 on day 14. From Table III we see that given profile 1 for that user, the hub visitation probabilities indicate *definite* visits to hubs $1, 4, 15$ and 18 on day 14, which may then be verified from his/her actual hub list distribution shown in Figure 6(a).

C. Hub List Size Distribution

The results in Table III may seem to indicate that several users tend to visit many hubs in any given day as their mobility profiles include multiple hubs. Hence, to study the distribution of the hub list sizes of our sample users we generate daily hub lists for each of them over their individual activation period based on their mobility profiles. More specifically, for each day we first choose one of their possible profiles at random following the mixing proportions, and then generate visits to each hub individually following the hub visitation probabilities in that chosen profile. We then obtain the aggregated (i.e., across all sample users) Hub List Size distribution and compare it with the actual distribution observed in the trace data. As seen in Figure 8(a), both the observed and the generated hub list sizes are distributed almost identically, and shorter (≤ 3) hub list sizes occur most often.

For a more comprehensive study, we also present the hub list size distribution observed for all the nodes in the analyzed data. Figure 8(b) shows the overall number of occurrences of different hub list sizes on a daily, weekly, and monthly basis. In these results below, each hub list size denotes the *unique* number of hubs a user visits in a day, week, or month. As seen in the figure, for all the time scales (i.e., daily, weekly, monthly), shorter hub lists sizes (≤ 3) occur most often, following which there is a



(a) Group 1 (0004.2396.92ab) (b) Group 2 (0001.e30d.d737) (c) Group 3 (0004.2398.82c0)

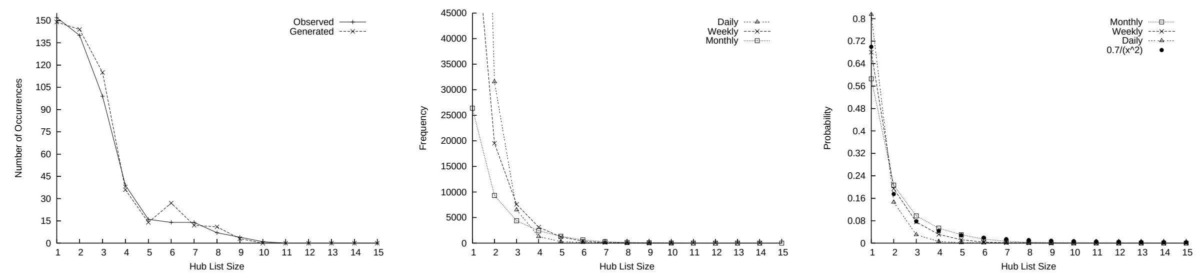
(d) Group 4 (0020.e089.9376) (e) Group 5 (0004.2396.8ced) (f) Group 6 (0005.4e41.cf1d)

Fig. 7. Daily distribution of mobility profiles

TABLE III
MOBILITY PROFILE PARAMETERS

Group 1 Profiles j	Mixing Proportions $\boldsymbol{\pi}_j$	Hub ID h (Hub Visitation Probability $\rho_{j,h}$)
1	0.08	1(1.0), 4(1.0), 15(1.0), 18(1.0)
2	0.31	1(1.0), 4(0.83), 8(0.27), 10(0.54), 11(0.27), 13(0.13), 18(0.13)
3	0.61	1(0.38), 4(0.81), 5(0.07), 15(0.34)
Group 2 Profiles j	Mixing Proportions $\boldsymbol{\pi}_j$	Hub ID h (Hub Visitation Probability $\rho_{j,h}$)
1	0.02	1(1.0), 9(1.0), 14(1.0)
2	0.14	1(0.14), 4(0.14), 5(0.49), 43(1.0)
3	0.31	1(1.0), 3(0.27), 5(0.8), 8(0.4), 9(0.53), 10(1.0), 13(0.6), 26(0.13), 27(0.2), 38(0.07), 43(0.07)
4	0.53	1(0.54), 3(0.08), 5(0.68), 8(0.04), 9(0.19), 13(0.08), 43(0.11)
Group 3 Profiles j	Mixing Proportions $\boldsymbol{\pi}_j$	Hub ID h (Hub Visitation Probability $\rho_{j,h}$)
1	0.06	1(1.0), 2(1.0), 5(0.75), 6(0.5), 7(0.75), 8(0.75), 10(0.5), 13(0.25), 16(1.0), 17(0.25), 29(0.25)
2	0.25	1(1.0), 3(0.22), 4(1.0), 5(1.0), 7(0.06), 8(0.66), 10(0.94), 11(0.06), 13(0.28), 15(0.89), 19(0.22), 26(0.11), 43(0.11)
3	0.32	1(0.09), 2(0.62), 4(0.05), 5(0.28), 7(0.75), 10(0.04), 15(0.05), 16(0.44), 17(0.04), 19(0.04), 25(0.18), 26(0.04), 29(0.09)
4	0.37	1(0.53), 3(0.12), 4(0.83), 5(0.18), 6(0.04), 8(0.04), 10(0.08), 14(0.19), 15(0.6), 43(0.11)
Group 4 Profiles j	Mixing Proportions $\boldsymbol{\pi}_j$	Hub ID h (Hub Visitation Probability $\rho_{j,h}$)
1	0.03	5(0.33), 8(1.0)
2	0.17	1(1.0)
3	0.80	1(1.0), 8(0.65), 10(0.9)
Group 5 Profiles j	Mixing Proportions $\boldsymbol{\pi}_j$	Hub ID h (Hub Visitation Probability $\rho_{j,h}$)
1	0.06	1(1.0), 3(1.0), 4(0.51), 5(0.14), 6(0.43), 8(0.29), 9(0.85), 10(0.58), 11(0.14), 13(0.58), 25(0.43)
2	0.20	1(0.14), 2(0.12), 3(0.22), 4(0.04), 7(0.54), 10(0.04), 15(0.26), 16(0.04), 30(0.04)
3	0.74	1(1.0), 3(0.36), 4(0.12), 5(0.01), 8(0.09), 9(0.06), 10(0.02), 15(0.23), 17(0.01), 21(0.01)
Group 6 Profiles j	Mixing Proportions $\boldsymbol{\pi}_j$	Hub ID h (Hub Visitation Probability $\rho_{j,h}$)
1	0.08	1(1.0), 3(1.0), 4(0.3), 5(0.4), 8(0.2), 9(1.0), 10(1.0), 11(0.9), 13(1.0), 18(0.1), 27(0.4)
2	0.21	1(0.03), 2(0.92), 12(0.19), 17(0.15)
3	0.26	1(0.87), 2(0.22), 4(0.7), 5(0.13), 8(0.31), 10(0.03), 11(1.0), 18(0.11)
4	0.45	1(0.56), 2(0.03), 3(0.045), 4(0.17), 5(0.35), 8(0.02), 10(0.05), 11(0.11), 18(0.36)

sharp decrease for hub list sizes of 4 or more. Figure 8(c) shows the fractional occurrences of the observed hub list



(a) Observed vs. Generated Hub List Size (b) Overall Hub List Size Frequency (c) Overall Hub List Size Probabilities

Fig. 8. Hub list size distributions

sizes, which are almost identical across the daily, weekly, and monthly values, and may be approximated as shown with a power law distribution. Based on these results, we may infer that the length of each hub list is determined more by the social routine of each user, than by the actual number of hubs present in the system, or the period for which the user is active.

V. Mobility Profile based Location Predictions

In this section, we highlight another important contribution of our work by showing how the mobility profiles may be useful in making hub-level location predictions with more accuracy than a general statistical method and at a lower overhead. More specifically, we first show an efficient way to apply the clustering algorithm described in Section IV-B and identify the right mixture of mobility profiles with lower overhead than compared to that in a statistical method. We then focus on two types of profile based predictions: *Unconditional Prediction*, where given the hub visit information over a window of n days, we wish to predict the hub visit patterns for the next window of n days; *Conditional Prediction*, where given that we can identify the current mobility profile of a user (based on available information about a hub a user either visited, or plans to visit), we wish to find the probability of that user visiting another hub in that same day.

A. Determining a Mixture of Mobility Profiles

From Figure 7, it becomes lucid that the seemingly random movements of a user as seen from Figure 6 can now be systematically described via a mixture of mobility profiles over a period of time. However, since this mixture will eventually change, we still need an efficient method to identify the right mixture of profiles describing the user's movement pattern over a given period. One may use the mobility traces of hub visits collected over 7 days (i.e., a week) to determine the possible mobility profiles and their corresponding mixing proportions using the *Mixture of Bernoulli's* described in Section IV-B. It is then possible to identify the appropriate mixture to include all the profiles with a corresponding mixing proportion greater than some specified threshold (which is 10% in our case, as shown in Figure 9). One may then choose to only consider this specific mixture for the next 7 days, when the next mixture update is performed. Considering the sample user from group G_6 as an example, we find that a mixture update (due to a change in the mixture from one window to the next) is required on days $14, 21, 56, 70, 77, 84, 91, 112,$ and 119, which amounts to only 9 updates for a 125 days activation period. Later in this section, we show that even with such infrequent updates (i.e., low overhead) our mobility profiles are able to predict daily hub-level locations with more accuracy than a common statistical method, which in contrast would require a hub visitation information update *every day*.

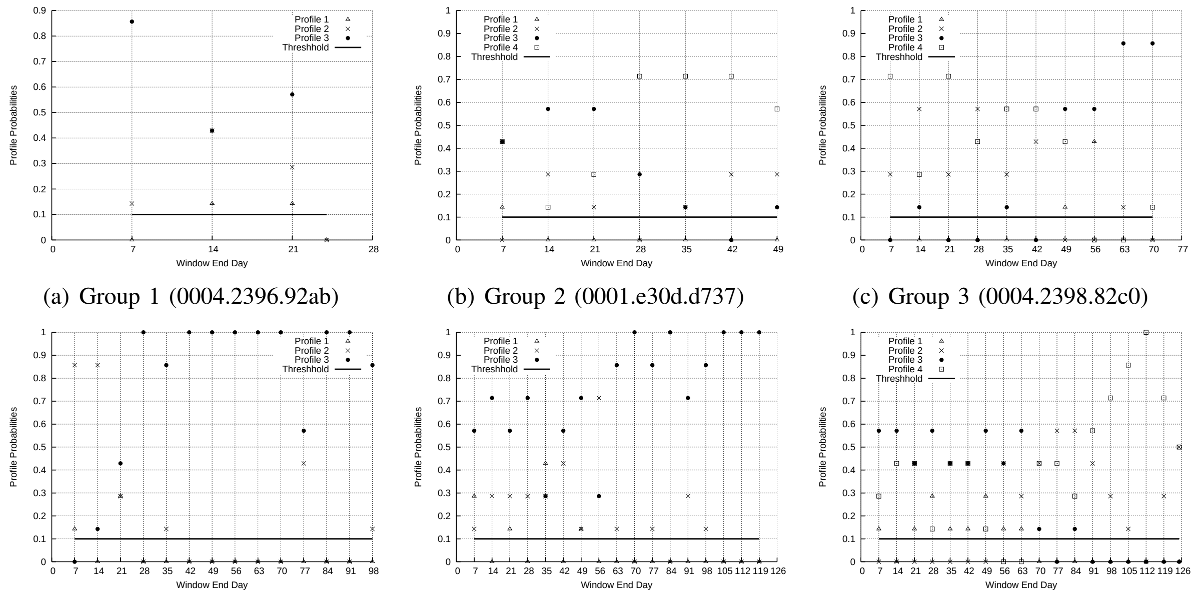
B. Unconditional Prediction

In this part, we study the accuracy of the unconditional profile based hub-level location prediction, and compare it with that made from statistical observation alone. We again consider only the sample users.

1) Statistical based prediction: In this method, one initially collects the mobility traces of a sample user for n days, and then determines the user's hub visit probabilities based on those n days. Using this probability distribution, one can then predict the user's hub list for day $n+1$, and compare it with the observed hub list for day $n+1$ to compute the daily Statistical based Prediction Error (SPE) rate as

$$SPE = \frac{Incorrect\ number\ of\ hub\ predictions}{Total\ number\ of\ hubs} \quad (5)$$

After day $n+1$, the sample user's hub visit probabilities are recomputed based on the past $n+1$ days and then



(a) Group 1 (0004.2396.92ab) (b) Group 2 (0001.e30d.d737) (c) Group 3 (0004.2398.82c0)

(d) Group 4 (0020.e089.9376) (e) Group 5 (0004.2396.8ced) (f) Group 6 (0005.4e41.cf1d)

Fig. 9. Windows of 7 days to find mixture of mobility profiles

used to predict the hub list for day $n + 2$, and so on till the end of the activation period for the sample user is reached.

2) Profile based prediction: In this approach, one initially collects a sample user's mobility traces for a window size of n days, and then applies the clustering algorithm described in Section IV-B to find out a mixture of mobility profiles and their associated probabilities (as was shown in Figure 9). Based on this profile information, one not only predicts the hub list for the next day (i.e., day $n + 1$, similar to the statistical based prediction), but also for the entire next window of n days (i.e., day $n + 1$ till $2n$). To be more precise, for each day, one first randomly chooses one mobility profile out of the mixture of profiles based on their mixing proportions and then predicts the day's hub list based on that chosen profile. For each day within the window, one compares the hub visit predictions with the observed hub visit values to compute the daily Profile based Prediction Error (PPE) rate similar to that shown for SPE in (5).

After the next window of n days, one re-computes the mixture of mobility profiles based only on the hub visit information of the last n days (unlike in the statistical method, where the entire past history is considered) and uses the new mixture information to predict the hub lists for days $2n + 1$ till $3n$, and so on till the end of the user's activation period is reached. *Note that the number of re-computations required for each sample user in the profile based method is thus only $1/n$ times that in the statistical method.*

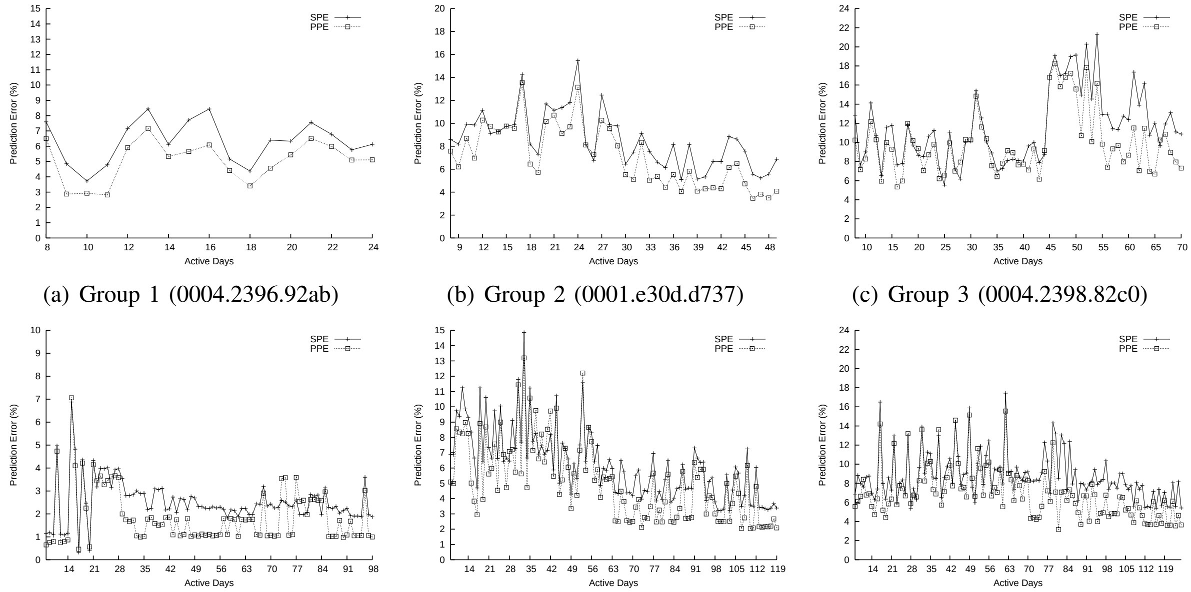
In our experiment, we choose a window size of $n = 7$ to be consistent with the results shown in Figure 9. In Figure 10, we plot the percentage values for SPE and PPE. For both SPE and PPE, each error value is an average over 1000 runs. As seen in Figure 10, PPE has lower values than SPE in almost all cases. The few cases where PPE has higher values than SPE is mostly attributed to a substantial change in the mixture of mobility profiles, where the old and new set of profiles had very different hub associations. To quantify the improvement in location prediction achieved by our profile based method over that by the statistical method, we define

$$Prediction\ Improvement\ Ratio(PIR) = \frac{SPE - PPE}{SPE}$$

and present its distribution parameters in Table IV. As seen, the mean values (considering the standard errors) are all positive, indicating a much better overall performance of our profile based hub-level location predictions as compared to the statistical approach and at a much lower overhead. ($1/7$ times that of statistical method in this case). This is one of the most critical contributions of our concept of profiling mobility based on sociological orbits.

C. Conditional Prediction

In this section, we show how the current mobility profile information may improve the performance of certain hub-level predictions. The authors in [9], [10]



(a) Group 1 (0004.2396.92ab) (b) Group 2 (0001.e30d.d737) (c) Group 3 (0004.2398.82c0)

(d) Group 4 (0020.e089.9376) (e) Group 5 (0004.2396.8ced) (f) Group 6 (0005.4e41.cf1d)

Fig. 10. Unconditional prediction comparisons

TABLE IV
THE DISTRIBUTION OF PIR (%)

$Group$	Mean $\pm$ Standard Error
G_1	20.6 ± 2.3
G_2	18.9 ± 2.0
G_3	12.9 ± 2.0
G_4	27.2 ± 3.0
G_5	21.5 ± 1.4
G_6	21.2 ± 1.5

have shown that a common statistical approach (similar to the one described in Section V-B) is capable of keeping track of a user's visits to different locations (via the system logs on APs). Consequently, it is possible to provide a probabilistic view of finding the user in any location at any time based on the past history of that user's hub visits. Taking the user from group G_2 as an example, we find that he/she visited hub 43 on 11 days in a 49 day activation period as seen in Figure 6(b). If we were to assume that this mobility pattern over 49 days is going to repeat itself, then the general probability of finding that user in hub 43 would be $\frac{11}{49} = 0.22$. From within our profile based framework, we not only are capable of providing similar general information but, given the current mobility profile, also can be much more specific. For instance, given the same example and assumption as above, the general statistical probability $P(h)$ of finding the user in hub h on any given day may be calculated equivalently through our approach as

$$P(h) = \sum_{j=1}^{k} \boldsymbol{\pi}_j * \rho_{j,h} \tag{6}$$

Using (6) and the data in Table III, the general probability of finding the user from group G_2 in say "target hub" $H_t = 43$ on day $D = 16$ of his activity would be given as: $(0.02) * (0.00) + (0.14) * (1.0) + (0.31) * (0.07) + (0.53) * (0.11) = 0.22$ (which is the same as that noted before). However, as soon as the user ventures into say "identifier hub" $H_i = 4$ on day 16 (see Figure 6(b)), our method shall identify the current profile (P_{now}) to be 2, as it is the only one with hub 4 in it. With this additional knowledge, our approach would then be able to re-compute the probability of finding the user in hub 43 on day 16 to be $\rho_{2,43} = 1$. From Figure 6(b), we find that the user did indeed go to hub 43 on day 16 (i.e. $\mathbf{y}_{43}^{(16)} = 1$), which makes our profile based prediction more precise. Several similar cases for each user type are listed in Table V, where we find that the conditional probability ρ_{j,H_t} (obtained based on mobility profiles) is closer to the actual event $\mathbf{y}_{H_t}^{(D)}$ than the general probability $P(H_t)$ (obtained from the common statistical approach). In particular, as seen in the cases for the users from groups G_2, G_3, G_4 and G_6 our predictions are completely accurate, whereas those from the statistical method are far from correct.

Essentially, the mobility profiles help us group the hubs in separate (but, potentially overlapping) sets of hubs on the basis of visits occurring to them within

TABLE V
CONDITIONAL PREDICTION COMPARISON

Group	H_t	D	$P(H_t)$	H_i	P_{now}	ρ_{j,H_t}	$\mathbf{y}_{H_t}^{(D)}$
G_1	11	13	**0.08**	8	2	**0.27**	**1**
G_2	43	16	**0.22**	4	2	**1**	**1**
G_3	7	7	**0.3**	14	4	**0**	**0**
G_4	1	15	**0.97**	5	1	**0**	**0**
G_5	7	53	**0.11**	2	2	**0.54**	**1**
G_6	3	63	**0.1**	9	1	**1**	**1**

the same period of time (i.e., following some mobility pattern), unlike in the statistical method where all the hubs are treated independently and identically. Note that in practice it may not always be possible to uniquely identify the current mobility profile based on the hubs visited so far (i.e., *identifier hubs*), as one hub could belong to 2 (out of say 4) profiles. However, as shown earlier in Section V-B, as long as the *identifier hub* is able to suggest a proper subset (or, a mixture) of the user's mobility profiles for a given period, we are able to predict hub visits more precisely than a common statistical method and at a lower overhead.

VI. OTHER RELATED WORK AND APPLICATIONS

In this section, we discuss relevant work (in an approximately decreasing order) to highlight the novelty of our sociological orbit aware approach on users' mobility pattern analysis and also discuss related applications of our approach.

In earlier work [13], [14], we intuitively assumed a basic sociological orbit framework and proposed orbit aware routing protocols within MANETs. However, we did not provide any empirical evidence to validate the existence of such sociological orbits. This current paper makes unique contributions via the orbit validation through empirical data analysis, mobility profiling and hub-level location prediction methods. One of the earliest attempt to understand the 802.11-based network access patterns, traffic load and user preference of wireless over wired networks was made by the authors of [22], [33], [34]. However, the main focus of these projects were not on the mobility pattern of individual users.

An extensive campus-wide 802.11 based wireless network testbed was setup in Dartmouth College, and its traces has been studied by the authors in [17], [20], [21], focusing more on AP-centric parameters. In particular, they studied amongst many things, the number of periodic visits to a particular AP or building, the length of such periods, and the frequency with which any sequence of 2 locations were visited in succession. However, for such information on periodic and sequential visits to location pairs to be useful in statistical location prediction for a set of n hubs, one would need all ${}^nP_2 = n * (n-1)$ permutations of visit sequences. In contrast, our proposed methods for profiling mobility based on sociological orbits only requires information on up to n hubs to be collected with a much lower frequency of updates as mentioned earlier in the paper.

There have been several other related work on analyzing 802.11-based wireless access patterns at ETH Zurich [35], University of North Carolina at Chapel Hill [10], University of California at San Diego [5], [24], MIT [6] and University of Saskatchewan [28], just to name a few. Most of these work focussed on topics such as: training mobility models for use in network simulation; taking wireless network load measurements for general resource allocation and capacity planning; understanding the user mobility in terms of periodic visits to APs; improving web-caching, QoS-routing, etc.

However, none of these work explored the notion of sociological orbits and in particular, mobility profiles. As a result, when it comes to predicting the next AP to be accessed for example, they often employ common statistical approaches whereby access to different APs is assumed to be independent. It is worth mentioning that much earlier work exists on understanding cellular users' mobility patterns, with primary objectives being triggering location updates based on user movement and improving paging services for example [3]. However, they dealt with a set of cells that have a much larger area than hubs, and whose physical boundary may not have special social implications as hubs do.

Note that our sociological orbit aware mobility profiles can also be helpful in improving realistic mobility modeling, QoS routing, and resource allocation. In fact, better understanding of users' mobility profiles will result in a greater benefit to many applications than the existing approaches. Example applications may be those that need to "push" a large amount of data (such as critical software updates, on-line movie rental, etc.) directly to a mobile user's laptop and/or PDA, or those requiring a small amount of data (e.g., event alerts) to be sent, or real-time audio/video or images from surveillance cameras to be streamed to concerned individuals (e.g., security personnel). Knowing a user's current mobility profile will enable such an application to afford simulcasting the data to a small set of hubs for reduced delivery time and better security. Conceptually, this generalizes the simple techniques used to download files and stream multimedia to vehicles on a highway equipped with roadside base stations that are spaced a few miles apart from each other, whose mobility patterns may easily be modeled following the work in [11] for example.

In addition to the above work on campus-wide wireless access networks, there has also been similar effort in studying how mobility (controlled or not) affects routing protocols and performance (e.g., network capacity) in ad hoc networks, including sensor networks with mobile sinks (or base stations), and delay tolerant networks (DTNs) or intermittently connected mobile ad hoc networks [7], [15], [30], [38]. However, they did not deal with specific user mobility patterns. In [18], [19], [23], [31], the main focus was on the so-called "contact probability" of two nodes, which is oblivious to the specific locations (or "hubs") they visit.

Our sociological orbit-aware mobility profiles can also be applied to ad hoc networks. For instance, knowing that two users will visit a common hub may help infer the contact probability. Moreover, it may allow one user to store a message within the common hub (e.g., using a stationary device) so as to be picked up later by another user without having any actual "contact" with each other. Example applications of the users' mobility profiles may be to monitor air (or, water) quality in an infrastructure-less environment and its impact on the health of the people who live or work there, or to detect and control the spread of a flu virus [37]. In these applications, people can wear tiny sensors with limited transmission range. During any given day, only a few people (say "carriers") may travel to a site as a part of their "social routine", where an access point is present, for uploading the sensor data and downloading control messages. A majority of others will only be able to send data to a person sharing the same "hub" as a part of its orbital pattern, and that person in turn will forward the data to another person until one of these "carriers" is reached. After the collected data is processed at a remote center, it is possible that certain symptoms are detected but definitive diagnosis is still not possible. In such cases, additionally, more intensive data collection may be needed only at selected locations and/or by selected persons. Knowing the orbital patterns of the persons will certainly help target the right subset of people and thereby reducing unnecessarily flooding of the request for data collection, and also cutting down the energy/bandwidth in collecting uninterested data, whereas knowing the contact probability alone may not be sufficient. Even in network security, authors in [16] had profiled mobility sequences based on location coordinates for reducing false alarms in anomaly-based intrusion detection.

There also exists a few studies on the social aspects related to wireless networks. For example, the work in [2] analyzed the NTT DoCoMo's i-mode users' email partners to infer their social relationships. Similarly, the authors of [25] studied the social influence on wireless networks as we did. However, they proposed mobility models based on *Social Network Theory*. In particular, they introduced a "sociability factor" to define the social relationships between users, and used that to characterize user groups that may have correlated mobility. Our work is different in that we profile users' mobility in terms of the locations they visit, instead of modeling their physical movement in between, or within those locations in terms of velocity and direction of the movement as in [25], [35].

VII. Conclusion

Knowing users' mobility patterns is crucial to the efficient design and operation of many wireless networks and applications that need to be scalable and QoS-capable.

In this paper, we have analyzed the year-long mobility trace data of 13,620 WLAN users collected on the campus of ETH Zurich with 391 Access Points (APs). We not only validate the so-called sociological orbits exhibited by mobile wireless users, but also profile the user movements to help in location prediction. Unlike previous work on analyzing similar mobility trace data which focus on AP-centric parameters, our focus has been on hub-centric parameters (where a hub is a place of social interest and thus can be served by several APs) such as hub staying time, and in particular, user centric-parameters such as the number of hubs visited by a user in a day (or hub list size) and mobility profiles (i.e., a probabilistic list of hub visitations) of a user.

We note that it should not be surprising that sociological orbits do exist in users' mobility patterns. Nevertheless, one of the contributions of this work is that for the first time, it has been found from analyzing the traces that a user does exhibit sociological orbits at multiple levels on different time scales: a hub-level orbit consisting of a number of hubs during a day, and a mobility profile-level orbit consisting of a few profiles during a week or longer period. In addition, the results of our analysis will enable researchers to model users' mobility in more realistic way using the appropriate values and distributions of user-centric parameters we have identified via analysis.

Another, perhaps a more important contribution is that, although intuitively, it is beneficial to exploit the knowledge of the sociological orbits in users' movement, this work is the first that proposed an efficient method to determine the main mobility profiles of a user using a mixture of Bernoulli's as the clustering algorithm, and then make either unconditional or conditional predictions on which hubs a user might visit. More specifically, our results use only a short history (e.g., the past 7 days) of mobility trace data, but is shown to predict around 10% to 30% more accurately than a general statistical approach that relies on daily collection and computation of the trace data. This illustrates the strength of our sociological orbit aware approach, and in particular, the usefulness of the mobility profiles of a user.

Note that although this work is based only on the mobility trace data from ETH Zurich, it is expected that the data analysis, mobility profiling and location prediction techniques we have developed, as well as the conclusions we have drawn in this paper that validate the existence and usefulness of the sociological orbits are in general applicable to other university and corporate campuses, as well as other public/private environments (there certainly isn't a sufficient amount of mobility trace data available except from a couple of places). In addition, we expect that this work will inspire additional innovative work on social influence aware and user-centric designs and operations of not only wireless access networks, but also mobile ad hoc and peer-to-peer networks, as well as intermittently connected or, delay tolerant networks.

REFERENCES

[1] ABOWD, G., AND MYNATT, E. Charting past, present, and future research in ubiquitous computing. *ACM Transaction on Computer Human Interaction 7*, 1 (March 2000), 29–58.

[2] AIDA, M., ISHIBASHI, K., TAKANO, C., MIWA, H., MURANAKA, K., AND MIURA, A. Cluster structres in topology of large-scale social networks revealed by traffic data. *Proceedings of IEEE Globecomm '05, St. Louis* (November 2005).

[3] AKYILDIZ, I. F., HO, J. S. M., AND LIN, Y.-B. Movement-based location update and selective paging for pcs networks. *IEEE/ACM Transactions on Networking (TON) 4*, 4 (August 1996), 629–638.

[4] BAI, F., SADAGOPAN, N., AND HELMY, A. Important: a framework to systematically analyze the impact of mobility on performance of routing protocols for adhoc networks. *Proceedings of IEEE INFOCOM '03 2* (March 2003), 825–835.

[5] BALACHANDRAN, A., VOELKER, G., BAHL, P., AND RANGAN, P. Characterizing user behavior and network performance in a public wireless lan. *In Proceedings of ACM Sigmetrics, '02* (March 1998).

[6] BALAZINSKA, M., AND CASTRO, P. Characterizing mobility and network usage in a corporate wireless local-area network. *In Proceedings of ACM MobiSys '03* (May 2003), 303–316.

[7] BURNS, B., BROCK, O., AND LEVINE, B. N. Mv routing and capacity building in disruption tolerant networks. *In Proceedings of IEEE INFOCOM '05 1* (March 2005), 398–408.

[8] CAMP, T., BOLENG, J., AND DAVIES, V. A Survey of Mobility Models for Ad Hoc Network Research. *Wireless Communications and Mobile Computing (WCMC): Special issue on Mobile Ad Hoc Networking: Research, Trends and Applications 2*, 5 (2002), 483–502.

[9] CHEN, G., HUANG, H., AND KIM, M. Mining frequent and periodic association patterns. *Dartmouth College, Coomputer Science and Engineering, Tech Report: TR 2005-550* (July 2005).

[10] CHINCHILLA, F., LINDSEY, M., AND PAPADOPOULI, M. Analysis of wireless information locality and association patterns in a campus. *In Proceedings of IEEE INFOCOM '04 2* (March 2004), 906–917.

[11] CHOFFNES, D. R., AND BUSTAMANTE, F. E. An integrated mobility and traffic model for vehicular wireless networks. *Proceedings of the 2nd ACM international workshop on Vehicular ad hoc networks* (September 2005), 69–78.

[12] DEMPSTER, A., LAIRD, N., AND RUBIN, D. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B 39*, 1 (1977), 1–38.

[13] GHOSH, J., PHILIP, S. J., AND QIAO, C. Poster abstract: Sociological orbit aware location approximation and routing (solar) in manet. *Presented as a Poster in ACM Mobihoc '05, Champaign, IL* (May 2005).

[14] GHOSH, J., PHILIP, S. J., AND QIAO, C. Sociological orbit aware location approximation and routing in manet. *In Proceedings of IEEE Broadnets '05, Boston, MA* (October 2005), 688–697.

[15] GROSSGLAUSER, M., AND TSE, D. N. C. Mobility increases the capacity of ad hoc wireless networks. *IEEE/ACM Transactions on Networking 10*, 4 (August 2002), 477–486.

[16] HALL, J., BARBEAU, M., AND KRANAKIS, E. Using mobility profiles for anomaly-based intrusion detection in mobile networks. *Pre-conference wireless and mobile security workshop, ISOC NDSS '05* (2005).

[17] HENDERSON, T., KOTZ, D., AND ABYZOV, I. The changing usage of a mature campus-wide wireless network. *In Proceedings of ACM MOBICOM '04* (September 2004), 187–201.

[18] HUI, P., CHAINTREAU, A., SCOTT, J., GASS, R., CROWCROFT, J., AND DIOT, C. Pocket switched networks and the consequences of human mobility in conference environments. *Presented at the Workshop on Delay Tolerant Networking during SIGCOMM 2005, Philadelphia, PA*.

[19] JUANG, P., OKI, H., WANG, Y., MARTONOSI, M., PEH, L. S., AND RUBENSTEIN, D. Energy-efficient computing for wildlife tracking: design tradeoffs and early experiences with zebranet. *Proceedings of the 10th International Conference on Architectural support for programming languages and operating systems (ASPLOS)* (October 2002), 96–107.

[20] KIM, M., AND KOTZ, D. Classifying the mobility of users and the popularity of access points. *Proceedings of the International Workshop on Location- and Context-Awareness (LoCA), Germany* (May 2005).

[21] KIM, M., KOTZ, D., AND KIM, S. Extracting a mobility model from real user traces. *To appear in IEEE INFOCOM'06, Barcelona, Spain* (April 2006).

[22] LAI, K., ROUSSOPOULOS, M., TANG, D., ZHAO, X., AND BAKER, M. Experiences with a mobile testbed. *In Proceedings of The Second International Conference on Worldwide Computing and its Applications (WWCA98)* (March 1998).

[23] LEGUAY, J., FRIEDMAN, T., AND CONAN, V. Dtn routing in a mobility pattern space. *Proceeding of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking* (August 2005), 276–283.

[24] MCNETT, M., AND VOELKER, G. M. Access and mobility of wireless pda users. *In Mobile Computing Communications Review 9*, 2 (April 2005), 40–55.

[25] MUSOLESI, M., HAILES, S., AND MASCOLO, C. An ad hoc mobility model founded on social network theory. *Proceedings of ACM/IEEE MSWIM '04, Venezia, Italy* (October 2004).

[26] REMAGNINO, P., AND FORESTI, G. Ambient intelligence: A new multidisciplinary paradigm. *IEEE Transactions on Systems, Man and Cybernetics 35*, 1 (Jan 2005), 1–6.

[27] SAMAL, S. Mobility pattern aware routing in mobile ad hoc networks. *MS Thesis, Virginia Polytechnic Institute and State University* (May 2003).

[28] SCHWAB, D., AND BUNT, R. Characterizing the use of a campus wireless network. *In Proceedings of IEEE INFOCOM '04 2* (March 2004), 862–870.

[29] SCHWARZ, G. Estimating the dimension of a model. *The Annals of Statistics 6* (1978), 461–464.

[30] SHAH, R., ROY, S., JAIN, S., AND BRUNETTE, W. Data mules: Modeling a three-tier architecture for sparse sensor networks. *IEEE SNPA Workshop* (May 2003).

[31] SPYROPOULOS, T., PSOUNIS, K., AND RAGHAVENDRA, C. S. Spray and wait: an efficient routing scheme for intermittently connected mobile networks. *Proceeding of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking* (August 2005), 252–259.

[32] SU, W., LEE, S.-J., AND GERLA, M. Mobility prediction and routing in ad hoc wireless networks. *International Journal of Network Management 11*, 1 (February 2001), 3–30.

[33] TANG, D., AND BAKER, M. Analysis of a metropolitan-area wireless network. *In Proceedings of ACM MOBICOM '99* (August 1999), 13–23.

[34] TANG, D., AND BAKER, M. Analysis of a local-area wireless network. *In Proceedings of ACM MOBICOM '00* (August 2000), 1–10.

[35] TUDUCE, C., AND GROSS, T. A mobility model based on wlan traces and its validation. *In Proceedings of IEEE INFOCOM '05 1* (March 2005), 664–674.

[36] WANG, W., AND AKYILDIZ, I. F. On the estimation of user mobility pattern for location tracking in wireless networks. *Proceedings of IEEE Globecom '02* (November 2002), 619–623.

[37] WANG, Y., AND WU, H. Dft-msn: The delay/fault-tolerant mobile sensor network for pervasive information gathering. *To appear in IEEE INFOCOM'06, Barcelona, Spain* (April 2006).

[38] ZHAO, W., AMMAR, M., AND ZEGURA, E. Controlling the mobility of multiple data transport ferries in a delay-tolerant network. *In Proceedings of IEEE INFOCOM '05* (March 2005).