An Information-theoretic Framework for Similarity-based, Opportunistic Social Networks [*]

Mai ElSherief[*], Tamer ElBatt[†◇], Ahmed Zahran[†◇], Ahmed Helmy[‡]

[*]Department of Computer Science, University of California, Santa Barbara, USA.

[†]Wireless Intelligent Networks Center (WINC), Nile University, Giza, Egypt.

[◇]Faculty of Engineering, Cairo University, Giza, Egypt.

[‡]The Department of Computer and Information Science and Engineering, University of Florida, Gainesville, USA.

Abstract

In this paper we study similarity-based networks as a key enabler for innovative applications hinging on opportunistic mobile encounters. In particular, we quantify the, inherently qualitative, notion of user similarity and introduce a novel information-theoretic framework to establish fundamental limits and quantify the performance of knowledge sharing policies. First, we introduce generalized, non-temporal and temporal profile structures, beyond mere geographic location, in the form of a probability distribution function. Second, we analyze classic and information-theoretic similarity metrics using publicly available data. The most noticeable insight is that temporal metrics yield, on the average, lower similarity indices, compared to the non-temporal ones, due to incorporating the dynamics in the temporal dimension. Third, we introduce a novel mathematical framework that establishes fundamental limits for knowledge sharing among similar opportunistic users. Finally, we present numerical results characterizing the cumulative knowledge gain over time and its upper bound,

[*]This work was funded in part by a Google Faculty Research Award.

knowledge gain limit, using publicly available smartphone data for the user behavior and mobility traces, in case of fixed as well as mobile scenarios. The presented results provide valuable insights highlighting the key role of the introduced information-theoretic framework in motivating future research, diverse scenarios as well as future knowledge sharing policies.

Keywords: opportunistic, social networks, similarity, modeling, information theory, user traces.

1 Introduction

Recent studies, e.g., [1], point out a significant increase in the number of mobile subscribers, approaching 7 billion users worldwide. This surge in mobile devices, complemented by a plethora of wireless standards and new use cases, has inspired novel networking paradigms and services ranging from social [2, 3] to business. However, fully understanding and exploiting the social structure of mobile users remains a daunting challenge hindering network optimization and new services. Earlier social studies, e.g., Homophily [Lazarsfeld and Merton (1954)], have shown that people tend to have similarities with others in close proximity. In such clustered communities of interest, people tend to communicate, interact and trust each other [4]. Hence, smartphones can further enrich the mobile user experience via highly personalized applications, e.g., location-based services, targeted advertisement and social networking applications among many others.

The development of similarity-based, opportunistic social networking applications would typically involve the design of three core components, namely mobile user profiles, similarity assessment and knowledge sharing, if users are deemed similar. User profiles capture behavioral patterns relevant to the application of interest. The similarity assessment component judges, quantitatively, the similarity between the profiles of mobile users in proximity. Once two users are deemed similar, they may share knowledge and tips using policies that may depend on the service type and user preferences. For instance, two shoppers coming in proximity, in the same store (e.g., kids wear), would exchange their “anonymized” profiles to assess similarity. If similar, the smartphone application exchanges tips about stores ratings, special offers and other relevant information. Despite the fact that establishing trust in opportunistic settings [5] and profile anonymization are key components of the envisioned system, they are complementary to this work and are important subjects for future research. In this paper, we assume all users trust each other and focus on introducing the new mathematical framework instead.

Mobile user profiles proposed in the literature can be grouped based on different perspectives. Few are based on user location, e.g., [4,6], while others extend the profile to capture facets beyond location, e.g., [7,8]. From another perspective, profiles may be classified into vector (non-temporal) and matrix (temporal) profiles depending on whether the temporal dimension is captured or not. Similarity assessment depends on the profile type and application context. Classic metrics exist for vector-based profiles such as cosine and Pearson correlation [9]. Distance metrics from probability theory, e.g., Hellinger distance [10], can be leveraged to assess similarity between probability distribution profiles, like the ones proposed here. On the contrary, very few metrics are introduced for temporal profiles, e.g., singular value decomposition (SVD) based metrics [4].

In [11], the authors study the problem of content dissemination in opportunistic social networks. Their main result shows that high contact rate, non-social nodes (rarely found in "temporal communities") are mostly responsible for efficient content dissemination. However, unlike this work, the model adopted in [11] is not information-theoretic.

Information-theoretic models have been employed to other problems, e.g., cooperative data compression and distributed source coding for data gathering in multi-hop wireless sensor networks (WSNs) with spatial correlations, e.g., [12–15]. However, their prime focus is to eliminate redundancies among the possibly correlated sensor measurements [13,14]. The joint entropy of the random variables representing the individual sensors as data sources constitutes the lower bound on the traffic volume generated by the sensors, where source coding algorithms try to achieve. On the other hand, our objective in this work is fundamentally different. We establish fundamental limits, using basic information theoretic constructs on the maximum knowledge available for a user to reap in a given opportunistic encounter [?],. As defined later in the sequel, the knowledge gain limit of an arbitrary user constitutes the upper bound on the amount of knowledge a user can reap in a given opportunistic encounter and is characterized by the joint entropy of the random variables modeling the individual knowledge each user bears in diverse aspects of life. Furthermore, the nodes are stationary in data gathering WSNs and communications is multi-hop, whereas in our problem setting nodes are generally mobile and communications is limited to single-hop (pair-wise encounters), with the possibility of forwarding knowledge acquired from previous encounters.

Our main contribution in this paper is a novel information-theoretic framework for knowledge sharing in similarity-based, opportunistic social networks. First, we extend mobile user profiles, beyond mere location, to a generalized probability distribution function and study non-temporal and temporal versions. Second, we distill key insights about classic and proposed temporal and non-temporal similarity metrics, using publicly available data [16]. Third, we show the potential of the Hellinger distance to assess similarity between probability distribution user profiles and propose a novel temporal similarity metric, based on matrix vectorization, to capitalize on the richness in the temporal dimension while relying on lightweight computations. Fourth, we introduce the new notions of *Knowledge Gain* and its upper bound, *Knowledge Gain Limit*, per user. Finally, we establish fundamental limits with the aid of information theory and unveil key insights for diverse network topologies, sharing policies and mobility scenarios and validate our theoretical findings using publicly available user behavior and mobility traces.

The rest of this paper is organized as follows. We first motivate the vision and proposed framework in Section 2. In Section 3, we study mobile user similarity with emphasis on probability distribution profiles, using classic and novel metrics. In Section 4, we shift our attention to the novel information-theoretic framework to establish fundamental limits and quantify the performance of candidate knowledge sharing policies. We present key results based on realistic user mobility traces [17, 18] augmented with behavior traces from another data set [16]. Finally, conclusions are drawn and potential directions for future research are pointed out in Section 5.

2 Motivation

The wide proliferation of resource-rich smartphones renders them tightly coupled to their users, bearing a wealth of behavioral data, e.g., locations, social networks, online shopping, etc., inferring information about the user's preferences and interests. Thus, there has been growing interest in leveraging this data to open new frontiers and enrich the user's life experiences [19]. An instance of this interaction also prevails in crowd sourcing applications which may affect the user behavior at real-time, e.g., Waze and Google maps provide indicators for traffic congestion and road accidents which advise the mobile users to alter their routes.

Inspired by the tight coupling between smartphones and users' behaviors, we pose the following fundamental question: Can we capitalize on the wealth of knowledge and life experiences of people we encounter throughout our lives and may have common interests, yet we do not know? The proposed framework caters to this question via an envisioned class of applications, coined *opportunistic recommendation systems* (ORS), whereby users capitalize on others' knowledge based on their mere co-existence and backed by homophily. The utility of ORS stems from extending our classic day-to-day "physical" recommendation exchanges, from people we know and encounter throughout the day to "cyber" exchanges with users we opportunistically encounter and do not know (yet may have things in common according to homophily) and even to users we have never encountered, through the concept of knowledge sharing/forwarding discussed later.

Finally, it is worth noting that similarity-based opportunistic social networks could serve as the basis for a variety of services, e.g., trust establishment, targeted advertisement, friend finders, and location/similarity-based services. Furthermore, ORS is expected to spur a plethora of novel smartphone applications serving large public venues, e.g., museums, theme parks, etc.

3 Pair-wise Mobile User Similarity

Similarity assessment is a classic problem in computer science, e.g., data mining, clustering and classification [20,21]. For instance, it has received considerable attention for recommendation systems in online social networks [22–25]. In [22], the authors propose a model to infer relationship strength based on profile similarity and interaction activity. In [23], similarity is computed based on users' ratings of items using heuristic measures such as cosine similarity and Pearson correlation. Similarity is also studied in various contexts, e.g., web users recommendation [24] and peer recommendation systems [25].

In mobile scenarios, similarity has received limited attention through exploiting the users' spatio-temporal proximity (i.e., being at the same place at the same time), e.g. [26–28]. In [28], similar users exchange ratings about touristic places they have previously visited. In [26] and [27], users can lookup who else is in proximity and depending on common interests may decide to communicate. To the best of the authors' knowledge, similarity of mobile users has been only investigated in [26,27,29,30]. However, the adopted user profile is solely based on location.

3.1 Generalized Mobile User Profiles

In this section, we introduce a profile structure for mobile users, beyond mere location. In addition, we explore non-temporal and temporal profiles. We assume V generic life categories, e.g., arts, sports, shopping, among others, chosen by the profile designer based on target application(s).

Thus, the non-temporal profile is a 1xV row vector where each element, C_i, captures the percentage of time spent by the mobile user, possibly online (*Interests*) or physical site visits (*Experiences*), in category i [31]. This vector can be viewed as a probability mass function (PMF) of the user profile random variable since $\sum_{i=1}^{V} C_i = 1$. The probability distribution definition of the user profile is not only convenient but also opens room for powerful mathematical tools to study similarity and knowledge sharing as discussed in Section 4.

On the other hand, inspired by [4] which proposed a temporal profile matrix for the user Wi-Fi Access Point connectivity pattern and the key observation that simple vector profiles hide important details about the user dynamics over time [31], we introduce probability distribution temporal profiles that capture facets other than location. Accordingly, profile vectors are captured over K time windows where each window could be a day, week, etc. depending on the dynamics of the user behavior and the time horizon of interest. This yields a KxV profile matrix where the K profile vectors are the rows of this matrix. Deciding the time granularity and horizon, K, is a key research issue which involves data mining and analysis techniques on real-life traces capturing the users' behavior dynamics over time and, hence, lies out of the scope of this work. For our comparative analysis, we rely on real smartphone traces from the LiveLab project at Rice University [16] where the window is taken to be one day and $K = 197$ days on the average.

Given the proposed PMF user profiles, we move next to similarity assessment.

3.2 Similarity Metrics

The choice of similarity metrics is highly dependent on the profile structure. For non-temporal profiles, cosine and Pearson correlation are widely used in the literature [9] taking values in the ranges $[0, 1]$ and $[-1, 1]$, respectively. These metrics are widely used due to their simplicity.

Inspired by the probability distribution structure of the proposed profiles, we examine distance metrics from probability theory, namely Hellinger distance, Canberra distance and Jensen Shannon Divergence [10]. The Hellinger distance is defined for two PMFs, A and B, as [10]

$$H(A, B) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^{V} (\sqrt{a_i} - \sqrt{b_i})^2}$$

where $H(A, B) \in [0, 1]$ and similarity can be easily defined as $Sim_{HL}(A, B) = 1 - H(A, B)$.

On the other hand, the Canberra distance and Jensen Shannon Divergence turned out to be problematic in our context since they yield infinite distances if there is one or more elements in the profile vector that are zero-valued. This is typical in practice and was frequently encountered in real-life traces, e.g., [16], where the users interests are clustered only in few categories.

On the other hand, temporal profiles lend themselves to two metrics. First, a metric based on Singular Value Decomposition (SVD) from linear algebra proposed in [4]. Second, we propose a novel, low-complexity vectorized cosine metric that is motivated by the limitations of SVD. SVD is an extension to classic cosine similarity and is defined for two profile matrices X and Y as

$$Sim_{SVD}(X,Y) = \sum_{i=1}^{Rank(X)} \sum_{j=1}^{Rank(Y)} w_{xi} w_{y_j} |V_{Xi} . V_{Yj}|, \tag{1}$$

which is essentially the weighted cosine similarity between the two sets of eigen-behavior vectors, where V_{Xi} and V_{Yj} are the ith and jth column of matrices V_X and V_Y, respectively. V_X and V_Y are the matrices obtained from the singular value decomposition (SVD) transformation [32] of X and Y, respectively, where $X = U_X \Sigma_X V_X^T$ and $Y = U_Y \Sigma_Y V_Y^T$.

On the positive side, SVD provides one provision for “anonymization’ since the users exchange only the elements of Σ and V, but not matrix U. This, in turn, prevents eavesdroppers from reconstructing the sender profile. On the down side, SVD similarity exhibits high computational complexity (scales quadratically with the history length K, for a fixed number of categories, V). Furthermore, similarity with oneself, $Sim_{SVD}(X, X)$, is maximum but not necessarily one, which causes problems while assessing similarity.

Motivated by the drawbacks of SVD and the simplicity of vector-based metrics, we propose a novel vectorized cosine (VCOS) metric with complexity scaling linearly with K. Thus, we transform the two KxV profile matrices, in question, to two 1x$K.V$ vectors via the vectorization operation in Linear Algebra [32] and then perform cosine similarity.

3.3 Similarity Metrics Performance Comparison

In this section, we compare the performance of different similarity metrics using a real data set from the LiveLab Project at Rice University [16]. This set offers traces for smartphone users and

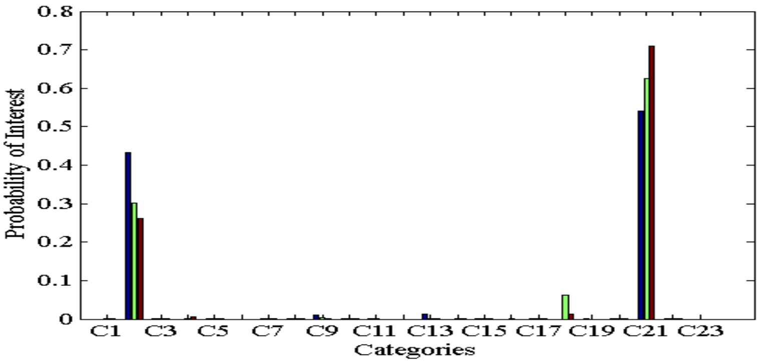

Figure 1: Sample profile PMFs for three users from LiveLab smartphone traces.

Wi-Fi access points (APs) from 34 iPhone 3GS users, including 24 Rice University students from Feb. 2010 to Feb. 2011 and 10 Houston Community College students from Sep. 2010 to Feb. 2011. The relevant data is stored in two database tables. The first hosts the names and genre (category) of 2500 iphone apps, as defined by the App Store. These apps are grouped to only 23 interest categories, e.g., books, business, sports, travel, etc. The LiveLab data is particularly chosen as it readily captures categorized smartphone digital footprint logs for the mobile users as opposed to other traces in the literature which include only Wi-Fi AP connectivity traces that are not relevant to our study. The second table includes the app usage history log for the 34 users with the date and duration of access. Cross referencing these two tables, we can generate non-temporal and temporal profiles for each user.

A remarkable observation on the distilled PMFs is that the majority of the categories in most profiles are zero-valued and the user activity is concentrated in two to five categories as depicted in Fig. 1 and witnessed in real-life. This renders the LiveLab users "qualitatively" similar. This key finding hinders the use of some metrics, such as Canberra Distance and Jensen Shannon Divergence, with such sparse profiles due to the aforementioned infinite distance problem.

Thus, we focus on the cosine (COS), Hellinger (HLNG), SVD and Vectorized cosine (VCOS) similarity metrics[1] to evaluate the pair-wise similarity for the 34 LiveLab users, which yields

[1] Pearson correlation is not examined since it ranges from [-1, 1] and mapping for comparison to other metrics

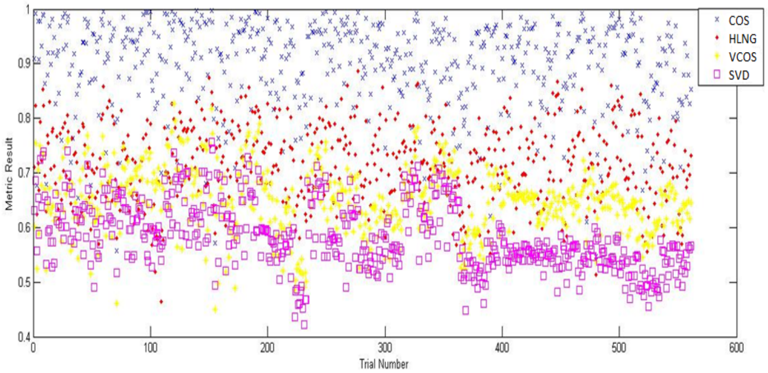

Figure 2: Metric indices for pair-wise similarity between LiveLab users.

Table 1: Percentage of similar users for all metrics for different similarity thresholds (T).

	T =0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
SVD	100	100	100	100	92.51	34.41	3.57	0	0	0
VCOS	100	100	100	100	98.93	80.93	18.36	0.3565	0	0
Hellinger	100	100	100	100	99.82	92.87	61.5	13.37	0	0
Cos	100	100	100	100	100	99.47	97.33	89.13	59.18	0

561 experiments. The outcomes of the four metrics are shown, vs the experiment index, in the scatter plot depicted in Fig. 2. First, it can be noticed from Fig. 2 that cosine and Hellinger similarity yield relatively higher metric values compared to the VCOS and SVD for the same pair of users. This interesting result confirms the intuition that temporal metrics are generally "more thorough" and, hence, conservative in assessing mobile user similarity. Thus, for a given threshold T between 0 and 1, two users may be perceived "similar" using a non-temporal metric, yet, are deemed "dissimilar" using a temporal metric. This is attributed to the fact that the temporal profiles are generally more thorough since they naturally bear more details and dynamics than the non-temporal ones. This result is shown quantitatively in Table 1. The table shows that VCOS and SVD yield a lower percentage of similar users than the cosine and Hellinger metrics, hence,

skew the similarity results.

they are more conservative. Second, the Hellinger metric may be perceived as a balance between both paradigms, since it is found to be closest to the average of the four metrics [31]. Although this demonstrates the potential of Hellinger similarity, it deserves further attention and analysis in future studies. Finally, the metrics studied and proposed here and the insights distilled open room for characterizing "actual" similarity, to serve as the ground truth in future work.

Based on the above observations, we envision two similarity assessment paradigms, namely macroscopic (non-temporal based) and microscopic (temporal-based), which can serve as building blocks for two-stage similarity assessment.

Macroscopic assessment: quantifies similarity between two vector-based, non-temporal profiles. Evidently, it is simpler and faster, with low-computational burden, yet, somewhat loose. Hence, it can serve as a first step "coarse-grained" similarity filter.

Microscopic assessment: scrutinizes similarity between two matrix-based temporal profiles. Unlike the first paradigm, it is more conservative in declaring similarity at the expense of more complexity and, hence, being slower. It serves as a second step "fine-grained" similarity filter.

In the next section, we shift our attention to knowledge sharing between similar users.

4 Knowledge Sharing in Opportunistic Social Networks

In the rest of the paper, we shift our attention to knowledge sharing between similar users. Our prime focus is to introduce a novel mathematical framework, establish fundamental limits, as opposed to designing and implementing specific knowledge sharing schemes, which constitute an interesting topic of future research. This framework lays the basis for assessing the merits of future knowledge sharing and delay-tolerant forwarding policies in opportunistic social networks.

In particular, we characterize, with the aid of information theory, the amount of knowledge a user can extract in an opportunistic encounter, coined knowledge gain (KG), and the maximum amount of knowledge available for a user in the network, coined knowledge limit (KL). The use of modeling abstractions to study the formation, dynamics and evolution of social networks is not new. For instance, graph theory has been employed extensively in social networks to model patterns of networks, clustering, homophily and basic concepts like centrality and connectedness,, e.g., [33, 34]. In addition, random graph theory has been employed to model social network formation, evolution and growth, e.g., [35, 36], among other topics. However, to the best of the authors' knowledge, employing information-theoretic tools to model knowledge sharing in opportunistic, mobile social networks has not been explored before.

4.1 Network Model and Assumptions

The notion of a "network" here, that is, nodes exchanging information, is established solely based on pair-wise similarity, according to Section 3. Thus, if a group of users in an opportunistic encounter, are all dissimilar, then there is no network, since no knowledge sharing will follow. The scenario of interest is the one that involves a subset of similar users which triggers tips exchange. Accordingly, we focus on a group of nodes where all nodes are pair-wise similar.

We model an opportunistic encounter of M similar users as a wireless ad hoc network. We assume that each user is similar to all other users in the network. Each user has its own non-temporal profile vector, or multiple row vectors across the temporal dimension, modeled as a probability distribution across different categories as described in Section 3.1. Each user is assumed to have a table of recommendations that stores *tips* (knowledge) for sharing with similar users, e.g., upcoming event(s), bestsellers, site visits, etc. We assume that the users leverage short-range wireless communication technologies with fixed transmission power (i.e. the circular disk model), e.g., Wi-Fi or Bluetooth, and, hence, medium access issues are resolved using these protocols.

In this section, we wish to address two fundamental questions pertaining to knowledge sharing:

1. For an arbitrary user i, what is the maximum amount of knowledge available for this user (fundamental limit) in a given similarity-based opportunistic encounter (network)?

2. For user i, what is the amount of knowledge gain that is achievable, i.e. the user can actually reap from similar users in the network, using a specific knowledge sharing policy?

Towards this objective, we introduce, next, terminology and mathematical definitions.

4.2 Knowledge Limit vs. Knowledge Gain

In this section, we introduce two new concepts that are fundamental to the analysis that follows, namely the knowledge limit and knowledge gain.

Definition 4.1. The Knowledge Limit (KL_i) is defined, for an arbitrary user i, as the maximum amount of knowledge that is available for user i to extract from similar users in a given network.

Definition 4.2. The Knowledge Gain (KG_i) is defined, for an arbitrary user i, as the amount of knowledge user i can gain from similar users, using a specific knowledge sharing policy.

It is straightforward to notice that $KG_i \leq KL_i$ since the knowledge limit constitutes the upper bound on the knowledge that can be reaped out of the network, irrespective of the sharing policy. Inspired by the probability distribution definition of the user profile, we argue that probability- and information-theoretic tools would prove useful for modeling and analyzing the system at hand.

Next, we introduce the formal definition of the knowledge gain per encounter. We assume that user tips (typically stored in a table) follow a probability distribution similar to the user profile. This does not only facilitate the mathematical analysis but is also a reasonable assumption, since users tend to have more tips in life categories they are more interested in.

4.2.1 The Knowledge Gain per encounter

We recall from information theory that the Entropy of a discrete-valued random variable X, denoted $H(X)$, represents a measure of the "uncertainty" which also represents the amount of information this random variable bears [37]. Given our assumption that the user recommendations/tips follow the same probability distribution as the user profile, tips can be modeled as a discrete random variable, X. Accordingly, $H(X)$ quantifies the amount of information (knowledge)[2] the user has. This model opens room for formally defining the newly introduced concepts

[2] We use the terms Knowledge and Information interchangebly in this paper.

of knowledge limit and gain.

First, we consider a toy example of an "opportunistic encounter" that involves only two users within the wireless communication range of each other. The two users have tips probability distribution vectors, denoted X and Y. Assume users X and Y meet opportunistically and are deemed similar[3] , according to Section 3. Thus, they start exchanging informative tips. Based on simple entropy relationships, we distinguish three types of "knowledge" quantities (tips):

1. Tips that user X has but Y does not: given by $H(X|Y)$.
2. Tips that user Y has but X does not: given by $H(Y|X)$.
3. Tips that both users have: given by $I(X;Y)$, the mutual information between X and Y .

Note that the first type of tips is the knowledge gain of Y. Second, the knowledge gained by user X from Y can be defined as

$$KG(X) = H(Y|X) = H(X,Y) - H(X) \tag{2}$$

where $H(X,Y)$ is the joint entropy of the two random variables representing the users tips probability distributions. The third type of tips (common to both users), characterized as the mutual information $I(X;Y)$, constitutes the "communication overhead" since it is exchanged over the air despite the fact that it does not contribute to increasing the knowledge of X or Y. This perfectly agrees with our assumption that the two users know nothing about each other, when they meet opportunistically, for the first time and, hence, this overhead is unavoidable.

It is worth noting that the knowledge limit for user X (or Y), in this toy example of two users, is equal to the knowledge gain.

4.2.2 The Knowledge Gain Limit

Based on the information-theoretic definitions of the KG and KL for a single encounter established in the previous section, we generalize the definition to characterize the knowledge limit for user

[3] We abuse notation and use the tips PMFs, X and Y, to refer to the users as well.

X_1, without loss of generality, in an opportunistic encounter with $M-1$ other users, deemed similar to X_1, as follows:

$$KL(X_1) = H(X_1, X_2, X_3, ..., X_M) - H(X_1) \tag{3}$$

which can be written as

$$KL(X_1) = H(X_2|X_1) + H(X_3|X_2, X_1) + ... + H(X_M|X_{M-1}, ..., X_1). \tag{4}$$

Thus, (3) asserts that the maximum amount of knowledge that user X_1 can extract from the network is simply the aggregate knowledge that all users have, after removing any redundant knowledge, which is represented by the joint entropy, $H(X_1, X_2, X_3, ..., X_M)$, less the knowledge user X_1 already has, that is, $H(X_1)$. We note that the KL characterization in (3) and (4) is general, valid for all network topologies and is independent of knowledge sharing policies.

4.3 Knowledge Sharing: Fundamental Limits and Policies

We utilize the basic definitions introduced in the previous section to establish the KL of an arbitrary user in diverse scenarios as well as its KG, under two sample knowledge sharing policies: i) Send my tips only, or "*Send Mine Only*", whereby a user sends only own tips to a similar, directly encountered user and ii) Forward my tips and others, or "*Forward Mine Plus Others*", whereby a user forwards own tips along with those acquired from others in previous encounters.

It is worth noting that these two policies are mere examples to illustrate the concept, however, other policies could be introduced and analyzed using the proposed model. For instance, a user could forward own tips along with a subset of others' tips based on some criteria. This gives rise to a family of knowledge sharing policies that deserves a comprehensive analysis, to assess their merits and potential trade-offs, which lies out of the scope of this work.

Next, we shift our attention to quantify the KL and KG achievable by the two aforementioned knowledge sharing policies, under a variety of opportunistic network configurations. In particular, we consider two connectivity scenarios, namely single-hop and multi-hop scenarios. In addition, we consider two mobility scenarios, namely fixed topology (i.e. stationary or quasi-stationary users) and time-varying topology caused by the user's portability within the same area.

4.3.1 Fixed Topology, Similarity-based Opportunistic Networks

A. Single-hop Networks

Under this setting, the users may be stationary, quasi-stationary or portable, yet, each node remains one-hop away, from all other nodes, all the time. For this setting, we can easily characterize the knowledge limit, as in (3), and, further, prove that the knowledge gain will attain the limit. This is in complete agreement with intuition since any node can take turn to exchange tips with all other nodes directly reachable. Thus, "all" knowledge that is available for any node in this network, can be fully reaped. This result is established for the *Send Mine Only* (SMO) policy in the following proposition.

Proposition 1. *For single-hop networks, an arbitrary node can achieve its knowledge gain limit using the SMO policy.*

Proof. Without loss of generality, we assume that node X_1 encounters other nodes in an ascending order of their IDs. Under the *Send Mine Only* policy, the cumulative knowledge gain for node X_1, $KG(X_1)$, after receiving tips from all other nodes $X_2, X_3, X_4, ..., X_M$ in turn, is given by $H(X_2|X_1) + H(X_3|X_2, X_1) + ... + H(X_M|X_{M-1}, ..., X_1)$, which is the same as $KL(X_1)$ in (4). The same argument can be applied to all other nodes in the network which proves the result. □

As indicated earlier, one of the fundamental issues in our study is how long does it take a user to attain the knowledge limit. This is directly related to the number of exchanges needed to attain the KL. Under the *Send Mine Only* policy and assuming that each node has at least one unique tip to contribute to the knowledge in the network, then it is straightforward to show that the worst-case number of exchanges needed for an arbitrary node to attain the KL is simply $(M - 1)$, that is, $O(M)$.

Next, we shift our attention to quantify the KG of single-hop networks, under the *Forward Mine Plus Others* (FMPO) sharing policy. Thus, a user shares not only its own tips but also tips collected from previous encounters, denoted by the subscript p. We prove in the following proposition that the knowledge limit is also attainable using the FMPO policy.

Proposition 2. *For single-hop networks, an arbitrary node achieves the knowledge gain limit using the FMPO policy.*

Proof. Without loss of generality, we assume that each node starts off with its own knowledge only and node X_1 encounters all other nodes in an ascending order of their IDs. The knowledge exchange goes over multiple rounds whereby in the first round, for instance, the following exchanges take place in parallel: $X_1 \leftrightarrow X_2$, $X_3 \leftrightarrow X_4$, $X_5 \leftrightarrow X_6$, etc. Thus, $KG(X_1)$ based on encountering nodes $X_2, X_3, X_4, ..., X_M$ in turn, is given by

$$KG(X_1) = H(X_2, |X_1) + H(X_3, \vec{X_{3p}}|X_2, X_1) + H(X_4, \vec{X_{4p}}|X_3, \vec{X_{3p}}, X_2, X_1) + ...$$

$$+H(X_M, \vec{X_{Mp}}|X_{M-1}, \vec{X_{(M-1)p}}, ..., X_1) \qquad (5)$$

where $\vec{X_{ip}}$ are the previous encounters of node X_i. It is straightforward to notice that $\vec{X_{3p}} = X_4$ since X_1 pairs with X_3 after the first round of exchanges. By the same token, $\vec{X_{4p}} = X_3, X_5, X_6$, since X_1 pairs with X_4 after two rounds of pairing and so on. Thus, substituting in (5) after $M/2$ rounds yields

$$KG(X_1) = H(X_2, |X_1) + H(X_3, X_4|X_2, X_1) + H(X_4, X_3, X_5, X_6|X_4, X_3, X_2, X_1) + ...$$

$$+H(X_M, \vec{X_{Mp}}|X_{M-1}, \vec{X_{(M-1)p}}, ..., X_1) \qquad (6)$$

Using the chain rule of entropies and expanding all terms in (6) yields some zero terms due to acquiring the same (redundant) knowledge from previous encounters. This, in turn, yields the KL in (4) and proves the result. □

It should be noted that once the conditioning, in the conditional entropy terms in the RHS, accommodates all nodes in the network, the incremental gain becomes zero and the node achieves its knowledge limit. In essence, the role of the previous encounters (appearing in the conditional entropy terms) is the sole contributor to the FMPO policy attaining the KL faster, compared to the SMO policy, which will be shown in Section 4.4. Apparently, this does not come for free since there is a fundamental trade-off between the cumulative KG after a number of encounters and the associated communication overhead which warrants attention in future research, especially in multi-hop networks. We prove next that the communication overhead of FMPO is greater than or equal to SMO, in single-hop networks.

Proposition 3. *For single-hop networks, the communication overhead under FMPO is greater than or equal to SMO.*

Proof. We consider an encounter beween two users, X and Y. Generalizing to a sequence of encounters is straightforward. Denote the vector of previous encounters for X and Y by $\vec{X_p}$ and $\vec{Y_p}$, respectively.

Under the SMO policy, the communication overhead that X incurs is the common knowledge (mutual information) between what X sends (which is its knowledge only) and Y's knowledge so far which is given by

$$OH(X)_{SMO} = I(X; Y, \vec{Y_p}). \tag{7}$$

Similarly, the communication overhead from the perspective of user Y is $OH(Y)_{SMO} = I(Y; X, \vec{X_p})$.

Under FMPO, the communication overhead is the same for both users and is given by

$$OH(X)_{FMPO} = OH(Y)_{FMPO} = I(X, \vec{X_p}; Y, \vec{Y_p}). \tag{8}$$

The mutual information between two random variables A and B can be written as

$$I(A; B) = H(A) + H(B) - H(A, B). \tag{9}$$

Applying (9) on (7) yields

$$OH(X)_{SMO} = H(X) + H(Y, \vec{Y_p}) - H(X, Y, \vec{Y_p}). \tag{10}$$

Applying (9) on (8) yields

$$OH(X)_{FMPO} = OH(Y)_{FMPO} = H(X, \vec{X_p}) + H(Y, \vec{Y_p}) - H(X, \vec{X_p}, Y, \vec{Y_p}). \tag{11}$$

Subtracting (10) from (11) yields

$$OH(X)_{FMPO} - OH(X)_{SMO} = H(X, \vec{X_p}) - H(X) - H(X, \vec{X_p}, Y, \vec{Y_p}) + H(X, Y, \vec{Y_p}) \tag{12}$$

Since $H(A, B) = H(A) + H(B|A) = H(B) + H(A|B)$, (12) can be re-written as

$$OH(X)_{FMPO} - OH(X)_{SMO} = H(X) + H(\vec{X_p}|X) - H(X) - [H(\vec{X_p}|X, Y, \vec{Y_p}) + H(X, Y, \vec{Y_p})] + H(X, Y, \vec{Y_p}),$$

which reduces to

$$OH(X)_{FMPO} - OH(X)_{SMO} = H(\vec{X_p}|X) - H(\vec{X_p}|X, Y, \vec{Y_p}), \tag{13}$$

$$\geq 0, \tag{14}$$

where the inequality in (14) follows since conditioning reduces entropy. This proves the result. □

B. Fixed Topology Multi-hop Networks

Under this setting, we assume the network topology is connected and time-invariant where some nodes are multi-hop away from each other. In this case, the role of the knowledge sharing policy stands out and affects whether a user can/cannot attain the KL.

Next, we quantify the performance, and trade-offs, of the SMO and FMPO policies. In case of SMO, the knowledge gain achieved by node X_1 is limited by the neighborhood size, $N{<}M$ which renders the KG strictly less than the KL. The following proposition establishes this result.

Proposition 4. *For fixed topology, multi-hop networks, the SMO knowledge sharing policy is not guaranteed to attain the knowledge gain limit, that is, $KG(X_1) \leq KL(X_1)$ iff $N < M$.*

Proof. Without loss of generality, we assume node X_1 communicates with other nodes in an ascending order of their IDs. The cumulative knowledge gain for node X_1, $KG(X_1)$, according to exchanges with neighbors $X_2, X_3, X_4, ..., X_N$ is given by $H(X_2|X_1) + H(X_3|X_2, X_1) + ... + H(X_N|X_{N-1}, ..., X_1)$. It is worth noting that the summation of non-negative conditional entropy terms is limited to $N < M$ nodes. It misses other non-negative terms involving the $M - N$ non-neighbors to X_1. Hence, it directly follows that $KG(X_1) \leq KL(X_1)$, which proves the result. □

It is worth noting that the special case of $N = M$, for all nodes, reduces to the single-hop network case where we have shown in Section 4.3.1.A that the KL is achievable under both knowledge sharing policies. An interesting, and somewhat surprising insight, which will be discussed later, is that nodes mobility can be leveraged to achieve the knowledge limit, even if $N < M$.

Next, we focus on the performance of the FMPO knowledge sharing policy for fixed topology multi-hop networks. As expected, forwarding others tips opens room for a node to achieve its KL, even if $N < M$. The following proposition formally establishes this result.

Proposition 5. *For fixed topology, multi-hop networks, an arbitrary node can achieve the knowledge gain limit using the FMPO knowledge sharing policy.*

Proof. We provide an outline of the proof due to space limiations. Without loss of generality, we assume that each node starts off with its own knowledge only and X_1 encounters its single-hop neighbors in an ascending order of their IDs, while other neighbors have pair-wise encounters with other nodes in the network. The cumulative knowledge gain for node X_1, $KG(X_1)$, after meeting neighboring nodes $X_2, X_3, X_4, ..., X_N$ is given by

$$\begin{aligned} KG(X_1) = H(X_2, |X_1) + H(X_3, \vec{X_{3p}}|X_2, X_1) + H(X_4, \vec{X_{4p}}|X_3, \vec{X_{3p}}, X_2, X_1) + ... \\ + H(X_N, \vec{X_{Np}}|X_{N-1}, \vec{X_{(N-1)p}}, ..., X_1). \qquad (15) \end{aligned}$$

At this point, two cases arise. First, if the previous knowledge vectors, $\vec{X_{ip}}\ \forall i$, bear the "forwarded" tips from all non-neighboring nodes, namely $X_{N+1}, X_{N+2}, ..., X_M$, then it can be shown that the cumulative knowledge gain of X_1 becomes

$$KG(X_1) = H(X_1, X_2, X_3, ..., X_M) - H(X_1) = KL(X_1) \tag{16}$$

which proves the result. Second, if the previous encounters do not cover knowledge from all non-neighbors, then this implies that node X_1 needs more time to attain $KL(X_1)$. Backed by network connectedness and unconstrained delay, X_1 can attain its KL almost surely via re-pairing with neighbors it has already parid with (to reap new knowledge acquired over time) until it acquires all missing knowledge from nodes out of its communication range. This proves the result. □

4.4 User Traces and Performance Results

In this section, we support our theoretical findings with numerical results based on smartphone user profile traces [16] and real-life user mobility traces, gathered at Infocom 2005 [17, 18].

4.4.1 Single-hop Networks

In this section, we rely on real user traces, either for user behavior or mobility. For user behavior, we utilize digital footprint traces (interests) for 20 smartphone users, over $V = 24$ life categories, from the LiveLab project [16]. In order to quantify the knowledge limit and gain for an arbitrary user, we need to pre-process a huge amount of six month worth of interests data in two steps. First, we compute the joint probability mass function for the 20 users under investigation over the period from September 2010 to February 2011. Thus, we monitor the users' activities categorized under 24 categories[4], each second, and record their concurrent activities. Afterwards, we divide by the total duration of the six months to get the joint PMF. Next, we show the cumulative knowledge gain increase as the node under investigation encounters more nodes over time.

First, we present the performance results for a single-hop network under the SMO knowledge sharing policy. For the network of $M = 20$ nodes discussed earlier, an arbitrary user can achieve the knowledge limit within $M - 1 = 19$ encounters. This is shown in Fig. 3 for three arbitrary users, namely $B00$, $B04$ and $D03$. It can be noticed that the cumulative knowledge gain is a

[4] The 24th category captures the case when the smartphone is off or not running any application.

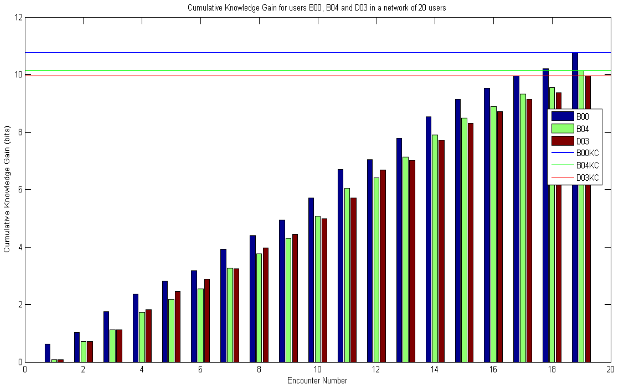

Figure 3: Cumulative knowledge gain for three users in a single-hop network under SMO.

non-decreasing function with time. On the other hand, the knowledge gain limit is shown as a horizontal solid line that is generally different from one user to another.

Next, we consider the same network setting, yet, employing the FMPO policy. Based on Proposition 2, we show that, for this type of networks, all nodes achieve the knowledge limit using the FMPO policy, yet, faster than SMO, i.e. in less encounters due to sharing the tips of others. This valuable insight is confirmed for users $B00$, $B04$ and $D03$ in Fig. 4.

4.4.2 Fixed Topology, Multi-hop Networks

As indicated earlier and established in Proposition 4, achieving the KL of an arbitrary user using SMO is fundamentally limited by the single-hop neighborhood size of this node, denoted N. To

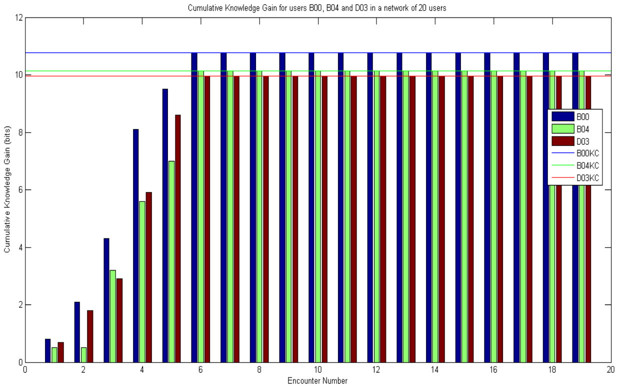

Figure 4: Cumulative knowledge gain for three users in a single-hop network under FMPO.

this end, we generate 20 randomly generated topologies of uniformly distributed users whereby each user has 6-7 single-hop neighbors, out of 20 nodes, on the average. It can be noticed that $B00$ does not achieve the KL available for it in this network, as established in Proposition 4 and shown here using real smartphone user behavior traces in Fig. 5. Thus, the maximum KG node $B00$ can achieve is only 43% of its KL. Similarly, users $B06$ and $D00$ have single-hop neighbors strictly less than $M = 20$ and, hence, cannot achieve their respective KLs.

Finally, we analyze the KG and KL performance of the FMPO policy in multi-hop topologies. In this case, the FMPO policy is expected to overcome the limited neighborhood problem due to sharing others' knowledge (tips) and, hence, the nodes could achieve the knowledge limit as shown in Proposition 5. The results here are based on 20 randomly generated topologies. The cumulative knowledge gains for users $B00$, $B06$ and $D00$ are depicted in Fig. 6. We notice that the KL is achievable for the three shown users after 8, 9 and 10 encounters, respectively.

4.4.3 Time-varying Topology (Mobile) Networks

A. User Profiles and Mobility Traces

After an extensive search for mobile user traces on publicly available data repositories, e.g., CRAWDAD [17, 18] and the alike, we did not find traces that include both, user behavior and mobility traces. Moreover, most of the mobility traces are university campus Wi-Fi access patterns as

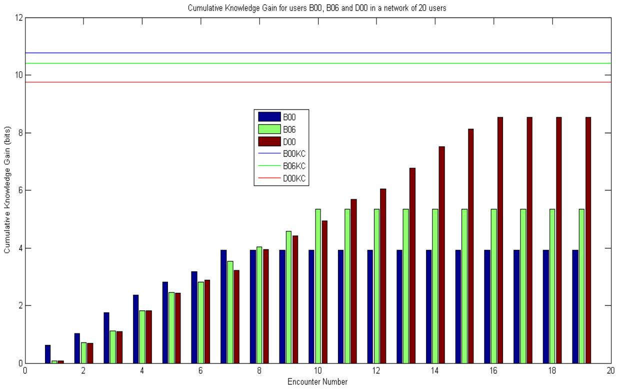

Figure 5: Cumulative knowledge gain for three users in a fixed multi-hop network under SMO.

opposed to mobile user encounters. In order to proceed with the performance evaluation based on real data, we resort to jointly leveraging traces from two different data sets, for user behavior and mobility. First, user profiles are constructed based on the LiveLab project data [16] described earlier. On the other hand, mobility traces are based on a "conference encounter" data, namely Infocom 2005 [17, 18]. For the Infocom 2005 experiment, the data set is relatively small whereby participants are 50 attending the student workshop. Nevertheless, it constitutes a reasonably sized set for our performance evaluation purposes. The students were given iMotes on March 7^{th}, 2005 between lunch time and 5 pm and collected on March 10th, 2005 in the afternoon. Two iMotes were lost while seven did not deliver useful data due to an accidental hardware reset. Contacts with these nine iMotes were discarded from the traces of others to avoid any effect on the results. The first six hours are discarded since they were attending the same workshop. We consider the contacts of 20 nodes only to match the number used from the LiveLab user profiles data. Thus, we associate the profiles of 20 randomly chosen users from the LiveLab data set to the mobility traces of 20 iMotes from Infocom 2005 and monitor them for half a day. This enables us to conduct our knowledge sharing analysis and collect the sought performance results.

Despite the fact that user profile and mobility traces are brought from two totally independent data sets, we find it a very useful attempt towards evaluating our policies, due to the lack of the

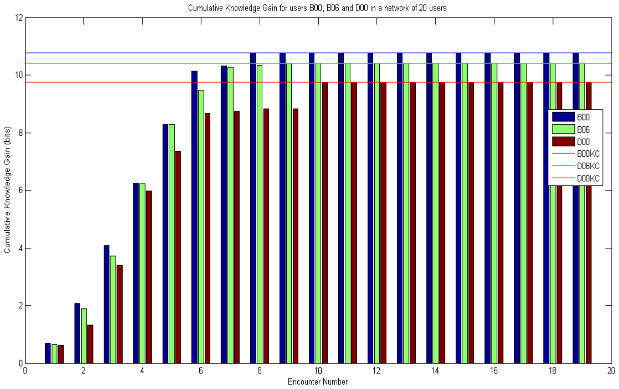

Figure 6: Cumulative knowledge gain for three users in a fixed multi-hop network under FMPO.

sought data in the public domain. This constitutes a strong motivation for the mobile networking and computing community to focus on the social dimension as well as the mobility and wireless connectivity dimensions, which already have several data sets in the public domain.

B. Performance Results

In this section, we quantify the knowledge limit and gain of time-varying topology (mobile) multi-hop networks, under the two sharing policies. Intuitively, users' mobility would play a key role in whether a node can achieve its KL and, if so, how much time this would incur. The gathered results are shown in Table 2. We compare the KG acquired by sample nodes using SMO in two cases, namely the stationary case where a snapshot is taken at time $t = 0$ and the mobile case over half a day. At time $t = 0$, all nodes, except for node $B07$, are disconnected yielding KG of zero. Node $B07$ is initially connected to $D06$ and reaps a KG of 0.64 as shown. The intriguing observation here is that mobility does help some nodes to approach the knowledge limit. On the other hand, some nodes, e.g., $B06$, do not benefit from mobility since they remain disconnected throughout the experiment lifetime. This insight agrees with intuition since the mobility patterns of some nodes could assist them in encountering the "knowledge hot spots" of the network. On the other hand, the mobility of other nodes could give rise to encounters with very slim/no KG benefits, e.g. nodes $B02$ and $B06$. Finally, we highlight that FMPO achieves KG no less than SMO, over the same period of time, which agrees with our theoretical findings.

Extensive studies of user encounter patterns in campus WLANs, e.g., [18, 38], have shown that, on the average, a user encounters only 2% of the population in a month and pointed out the heavy clustering of a user's behavior (spending 90% of their online time within only five APs (out of 600). In the following proposition, we establish the result based on an ideal mobility model guaranteeing encounter with all other nodes, which may not be valid for a whole campus scenario according to [38]. Nevertheless, for local encounters and mobile communities, the encounter ratio tends to be quite high (vs. 2% for the whole population) and our model is likely to be valid for

Table 2: Cumulative knowledge gain for nine mobile users after half a day.

	B00	B02	B03	B04	B05	B06	B07	B08	B09
Knowledge Limit (in bits)	10.76	10.24	10.44	10.13	10.22	10.4	10.46	10.09	10.34
KG Using SMO for stationary nodes (t=0) (in bits)	0	0	0	0	0	0	0.64	0	0
KG using SMO (in bits)	7.12	0.63	7.44	6.94	9.03	0	8.82	7.78	8.36
KG using FMPO (in bits)	9.12	9.05	7.93	7.62	9.03	0	8.82	8.45	8.97

realistic mobility scenarios.

Based on the seminal work on the effect of mobility on the throughput and delay in wireless ad hoc networks [39], the following proposition proves that the knowledge limit in mobile, delay-tolerant, multi-hop social networks is always achievable, under idealistic assumptions and loose delay constraints. Under those assumptions, an arbitrary node will encounter all other nodes in the network, almost surely. Nevertheless, modeling realistic mobility and characterizing the conditions under which the KG is improved by mobility is an interesting subject of future research.

Proposition 6. *For a time-varying topology network, an arbitrary node achieves its knowledge limit under loose delay constraints, almost surely, in case each node moves according to an independent two-dimensional random walk in a fixed area.*

Proof. In case of loose delay constraints and independent two-dimensional random walks, an arbitrary node will encounter all other nodes in the network, almost surely (follows from Lemma 6 in [39]). Hence, without loss of generality, we assume that node X_1 encounters all nodes in an increasing order of their node IDs. Under SMO, the cumulative knowledge gain for node X_1, $KG(X_1)$, based on encountering nodes $X_2, X_3, X_4, ..., X_M$ is the same as (4). Similar arguments can be employed to prove the same result using the FMPO policy, which proves the proposition. □

5 Conclusion

In this paper we propose a novel information-theoretic framework for similarity-based opportunistic social networks. We first introduce generalized, non-temporal and temporal profiles in the form of a probability distribution function. Second, we analyze classic and information-theoretic similarity metrics using publicly available data. We observe that temporal metrics yield, on the average, lower similarity indices, compared to the non-temporal ones, due to incorporating the dynamics in the temporal dimension. Third, we introduce a novel mathematical framework that establishes fundamental limits and insightful results for sample knowledge sharing policies among similar opportunistic users. Finally, we present numerical results characterizing the cumulative knowledge gain over time and its upper bound, knowledge gain limit, using publicly available traces for user behavior and mobility, in case of fixed and mobile scenarios. This work can be extended along multiple directions, e.g., novel similarity metrics capitalizing on the strengths of non-temporal and temporal profiles, examine Hellinger and vectorized cosine similarity with diverse users and scenarios, leverage the proposed mathematical framework to analyze novel knowledge sharing policies and, finally, establish fundamental limits for realistic mobility scenarios.

References

[1] ITU, "ITU releases 2014 ICT figures," http://www.itu.int/net/pressoffice/press_releases/2014/23.aspx/, accessed: 02/19/2015.

[2] I. Smith, "Social-mobile applications," *Computer*, vol. 38, no. 4, pp. 84–85, 2005.

[3] Y.-J. Chang, H.-H. Liu, L.-D. Chou, Y.-W. Chen, and H.-Y. Shin, "A general architecture of mobile social network services," in *Convergence Information Technology, 2007. IEEE International Conference on*, 2007.

[4] W. Hsu, D. Dutta, and A. Helmy, "CSI: A paradigm for behavior-oriented profile-cast services in mobile networks," *Ad Hoc Networks*, vol. 10, pp. 1586–1602, 2012.

[5] S. Trifunovic, F. Legendre, and C. Anastasiades, "Social trust in opportunistic networks," in *IEEE INFOCOM Worshop.*, 2010, pp. 1–6.

[6] W. Hsu, D. Dutta, and A. Helmy, "Profile-cast: Behavior-aware mobile networking," in *Wireless Communications and Networking Conference, 2008. WCNC 2008. IEEE*, 2008, pp. 3033–3038.

[7] J. Ramer, A. Soroca, and D. Doughty, "Mobile user profile creation based on user browse behaviors," Oct. 30 2007, US Patent App. 11/929,129.

[8] S. Moghaddam, A. Helmy, S. Ranka, and M. Somaiya, "Data-driven co-clustering model of internet usage in large mobile societies," in *Proceedings of the 13th ACM international conference on Modeling, analysis, and simulation of wireless and mobile systems*, 2010, pp. 248–256.

[9] W. Jones and G. Furnas, "Pictures of relevance: A geometric analysis of similarity measures," *Journal of the American Society for Information Science*, vol. 36, pp. 420–442, 1987.

[10] S. Cha, "Comprehensive survey on distance/similarity measures between probability density functions," *City*, vol. 1, no. 2, p. 1, 2007.

[11] A.-K. PietilÃďnen and C. Diot, "Dissemination in opportunistic social networks: the role of temporal communities," in *ACM Mobihoc*, 2012.

[12] S. Pradhan and K. Ramachandran, "Distributed source coding: symmetric rates and applications to sensor networks," in *IEEE Data Compression Conference*, 2000.

[13] D. Marco, E. Duarte-Melo, M. Liu, and D. Neuhoff, "On the manyto-one transport capacity of a dense wireless sensor network and the compressibility of its data," in *IEEE/ACM IPSN*, 2003.

[14] S. Pattem, B. Krishnamachari, and R. Govindan, "The impact of spatial correlation on routing with compression in wireless sensor networks," *ACM Transactions on Sensor Networks (TOSN)*, vol. 4, no. 4, 2008.

[15] T. ElBatt, "On the trade-offs of cooperative data compression in wireless sensor networks with spatial correlations," *IEEE Transactions on Wireless Communications)*, vol. 8, no. 5, 2009.

[16] E. T. Z. A. ElSherief, M. and A. Helmy, "An information-theoretic model for knowledge sharing in opportunistic social networks," in *The 8th IEEE International Conference on Social Computing and Networking (SocialCom 2015)*, 2015.

[17] C. Shepard, A. Rahmati, C. Tossell, L. Zhong, and P. Kortum, "Livelab: measuring wireless networks and smartphone users in the field," *ACM SIGMETRICS Performance Evaluation Review*, vol. 38, no. 3, 2011.

[18] "CRAWDAD trace set cambridge/haggle/imote (v. 2009-05-29)," Downloaded from http://crawdad.cs.dartmouth.edu/cambridge/haggle/imote.

[19] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on opportunistic forwarding algorithms," *IEEE Transactions on Mobile Computing*, vol. 6, no. 6, pp. 606–620, 2007.

[20] N. Eagle, A. Pentland, and D. Lazer, "Inferring friendship network structure by using mobile phone data," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 36, 2009.

[21] R. T. Ng and J. Han, "Efficient and effective clustering methods for spatial data mining," in *Proc. of the 20th International Conference on Very Large Data Bases*, 1994, pp. 144–155.

[22] P. Berkhin, "A survey of clustering data mining techniques," in *Grouping multidimensional data.* Springer, 2006, pp. 25–71.

[23] R. Xiang, J. Neville, and M. Rogati, "Modeling relationship strength in online social networks," in *Proceedings of the 19th ACM International Conference on World wide web*, 2010, pp. 981–990.

[24] I. Konstas, V. Stathopoulos, and J. M. Jose, "On social networks and collaborative recommendation," in *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, 2009, pp. 195–202.

[25] M. Pazzani, "A framework for collaborative, content-based and demographic filtering," *Artificial Intelligence Review*, vol. 13, no. 5, pp. 393–408, 1999.

[26] A. Tveit, "Peer-to-peer based recommendations for mobile commerce," in *Proceedings of the 1st ACM International Workshop on Mobile commerce*, 2001, pp. 26–29.

[27] S. Brings, "Location-based dating to the iphone," http://venturebeat.com/2009/01/21/skout-brings-location-based-dating-to-the-iphone/, accessed: 01/04/2012.

[28] I. Trestian, S. Ranjan, A. Kuzmanovic, and A. Nucci, "Measuring serendipity: connecting people, locations and interests in a mobile 3g network," in *Proceedings of the 9th ACM SIGCOMM Internet measurement conference*, 2009, pp. 267–279.

[29] A. De Spindler, M. Norrie, M. Grossniklaus, and B. Signer, "Spatio-temporal proximity as a basis for collaborative filtering in mobile environments," *Proceedings of UMICS*, pp. 912–926, 2006.

[30] M.-J. Lee and C.-W. Chung, "A user similarity calculation based on the location for social network services," in *Database Systems for Advanced Applications.* Springer, 2011, pp. 38–52.

[31] G. S. Thakur, A. Helmy, and W.-J. Hsu, "Similarity analysis and modeling in mobile societies: the missing link," in *Proceedings of the 5th ACM workshop on Challenged networks*, 2010, pp. 13–20.

[32] M. ElSherief, T. ElBatt, A. Zahran, and A. Helmy, "The quest for user similarity in mobile societies," in *The 2nd International Worskhop on Social and Community Intelligence, in conjunction with IEEE Percom.*, 2014.

[33] G. Strang, "The fundamental theorem of linear algebra," *American Mathematical Monthly*, pp. 848–855, 1993.

[34] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social networks*, vol. 1, no. 3, pp. 215–239, 1979.

[35] J. A. Barnes, "Graph theory and social networks: A technical comment on connectedness and connectivity," *Sociology*, vol. 3, no. 2, pp. 215–232, 1969.

[36] P. Erdos and A. Renyi, "On the evolution of random graphs," *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, vol. 5, pp. 17–61, 1960.

[37] M. E. Newman, D. J. Watts, and S. H. Strogatz, "Random graph models of social networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. Suppl 1, pp. 2566–2572, 2002.

[38] J. Thomas and T. Cover, *Elements of Information Theory, 2nd Edition.* John Wiley & Sons, Inc., 2006.

[39] W.-j. Hsu and A. Helmy, "On nodal encounter patterns in wireless lan traces," *IEEE Transactions on Mobile Computing*, vol. 9, no. 11, pp. 1563–1577, 2010.

[40] A. El Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Throughput-delay trade-off in wireless networks," in *IEEE INFOCOM'04*, 2004.