On the Relevance of Social Information to Opportunistic Forwarding

Abderrahmen Mtibaa Martin May Mostafa Ammar

Technicolor, Paris, France Georgia Institute of Technology

firstname.lastname@technicolor.com ammar@cc.gatech.edu

***Abstract*—**

In opportunistic ad-hoc networks, multi-hop data transfer over contemporaneous paths is unlikely since the devices are often disconnected from each other. However, data can still be stored and forwarded over time in an opportunistic hop-by-hop manner. Previous work has considered how the availability of various types of information such as social relationships can be used to guide forwarding algorithm to make better decisions and bring messages closer to the destination. This implicitly assumes that opportunistic contacts relate with the social property of node. However, the impact of such correlation between social and contact properties on social forwarding performances remains largely unexplored.

In this paper we argue that the relevance of such social information (*social inputs*) is as important as designing a new social forwarding algorithms. We examine multiple datasets to determine the impact of correlation, if any, between social information of individuals and their mobility patterns on the forwarding performances. We propose methods which process the social inputs to improve the relevance of such social information to forwarding. We show that our processing methods could improve the *success rate* performances of many social forwarding algorithms by more than 30%.

I. Introduction

With the emergence of a new generation of powerful mobile devices, new communication paradigms exploiting ad hoc data transfer becomes possible. In such ad hoc opportunistic network settings, end-to-end connectivity between devices using multi-hop data transfer is unlikely since the devices are often disconnected from each other. Devices in these networks [9], [5] have to decide whether or not to forward data to an intermediate node that they encounter. These forwarding decisions are typically guided by the desire to deliver the message as fast as possible while reducing the number of replicas of data items in the network.

In the recent past, the use of social networking information to optimize opportunistic forwarding decisions has become a hot, but also challenging, research topic. The idea is to use social interactions between participants to optimize message's forwarding decisions by anticipating the participant mobility.

There are basically two ways to capture social interactions between participants: using contact statistics and derive social interactions from them or by exploiting explicit social information. [10] for example, proposes a method to approximate the social interaction graph using contact statistics. With the emergence of online social network platforms and applications such as Facebook, Orkut, or LinkedIn, information about social interaction is readily available and easily modeled as a graph where two nodes are linked in the graph if they are "friends" in the social network. While this approach is applied and examined more and more often, the verification of the relevance of such information for social forwarding in general remains largely unexplored.

In this paper, we provide, to the best of our knowledge, the first assessment of the adequacy of explicit social information. We use multiple mobility traces that contain contact information as well as information about the social interaction between the participants. The social information includes very strong properties like declared friendships, but also considers common interests, common affiliations, and alike. We propose two metrics to assess the matching of the contact and the social graphs and to rank the quality of the information.

Instead of designing and evaluating a new social forwarding scheme, our work consists on verifying the impact of the relevance of such social information on the performance of difference forwarding algorithms. The question is now why if the social information used is not relevant to predict future contacts? Can we process these information in order to improve the final forwarding decisions?

The first insight in our work is to evaluate the impact of the relevance of social information on forwarding decisions and propose formal methods to improve theses decisions. We show that the performances of social forwarding schemes varies dramatically according to the nature of social interaction used. We proposed methods to process the original social information relying on either node's contact properties or other social properties in order to improve the relevance of such information to forwarding schemes. Our proposed methods may helps different social forwarding algorithms (considered in this paper) to outperforms theirs original *success rate* performances by more than 30%. However, we argue that when social inputs are totally irrelevant for prediction, forwarding decisions may be falsified; we generally observed, in this case, that even stateless probabilistic schemes outperforms social forwarding algorithms.

Despite the fact that social information are used as a good predictor for human mobility, social forwarding suffers in large networks where the social graph consists of different communities (*e.g.,* people in Paris could be socially connected to others in New York City but they are unlikely to meet each others physically). We propose a two step technique that could be integrated to most existing social forwarding schemes in order to improve theirs performances in large scale networks. It consists on (*i*) keeping the original social forwarding inside the communities, and (*ii*) circulating the messages among other communities using particular nodes which will operate as "bridges" connecting different communities.

The rest of the paper is organized as follows. We survey related work in Section section II. Section III overviews the basic idea behind our scheme and details our methodology. Section IV studies the relevance of social attributes and their impact on forwarding decisions. Section V investigates the impact of our data processing methods on social forwarding performances. In section VI, we expand our study for large scale networks. We conclude the paper in Section VII.

II. Related work

Mobile phones are no longer used only for calling or messaging. Nowadays, mobile phones provide constant Internet access and allow for a continuous maintenance of the online social network sites. Specifically, the opportunistic nodes may use information about the social network of the encountered participant.

Several research works recently considered the problem of designing mobile opportunistic forwarding schemes that are aware of social properties [10], [7], [13], [14], [8]. In these works, authors propose algorithms which make use of social metrics such as betweenness centrality or node degree in order to identify the set of node who are likely to relay the message to its destination. This implicitly assumes that opportunistic contacts relate with the social property that is used to design one algorithm. Our work does not propose new algorithms but it addresses the above issue more generally. It helps to understand what type of social information is the most relevant for forwarding, and analyze the impact of choosing a "good" and meaningful social information on social forwarding decisions.

Another series of papers are rather looking at the properties of the social graphs only. Social scientists in [3], and [2], make a difference between *self-reported* (*e.g.,* facebook social graph) and *aggregated* social graphs (*i.e.,* contact graph is aggregated into a social graph) for the same set of users. Their analysis show that self-reported social graphs, which are costless compared to aggregated social graphs, lead to satisfactory relevance to social forwarding. Our work addresses the above issue from a different viewpoint. Starting from an original social graph, we study the relevance of such social information to forwarding. Then, we consider a method to augment and improve the relevance of such social inputs to improve social decisions. We measure the impact of processed social inputs on three social forwarding algorithms [7], [13], [14]:

- Degree-Based Forwarding: consists on forward messages to socially well connected nodes. In fact, paths are constructed according to a non-decreasing social node degree rule (more details in [13]).
- Centrality-Based Forwarding: the main idea behind this algorithm is that central nodes in social graph are more likely to socialize with other people and then able to forward messages (more details in [13], [7]).
- PeopleRank : Fully distributed algorithm which rank nodes in the social graph similarly to what PageRank [4] does for web page - *i.e.,* it measures the relative "importance" of a node in the social graph. Then message forwarding decisions follows a non-decreasing rank rule (more details in [14]).

III. Motivation and Methodology

Social interaction between people was largely used as a good predictor for human mobility. Under such assumption, social information is used to optimize opportunistic forwarding decisions in Mobile Opportunistic Networks. Our main goals are *(i)* to study the impact of the nature of social information on forwarding performances, and *(ii)* discuss the different methods to process and improve such social information in order to offer better contact predictions to all social forwarding algorithms.

A. Network Model

We are interested in delivering data among a set of N mobile wireless nodes. Communication between two nodes is established when they are within radio range of each other. Data is forwarded from source to destination using these opportunistic *contacts*. We model the evolution of contacts in the network by a time varying graph $G(t) = (V, E(t))$ with $N = |V|$. We assume that the network starts at time t_0 and ends at time T (T can be infinite). We call this temporal network [11] the *contact graph*. Each $G(t)$ describes the contacts between nodes existing at time t. Such a time-varying graph model can be obtained from a mobility/contact trace[1] or from a mobility model along with knowledge of radio properties (*e.g.,* radio range).

We model social relationships between mobile nodes using a non-time varying graph, which we denote as $G_s = (V_s, E_s)$. In general, we assume that $V_s \supseteq V$, that is, some nodes in the social graph will not be part of our mobile network set. Social graphs reflect the interaction or interrelation between persons. Such information is available either in online social applications or could be extracted from the phone history or other sources. A link in the social graph between two nodes implies that these nodes are socially "connected" according to one or more *social attributes* (*e.g.,* friends in facebook or sharing a common interest). Note that in this paper social information is not deduced from contact properties between nodes as in [8] (*e.g.,* aggregation of contact graph into a social graph).

B. Social Forwarding algorithms

Let's consider a source node s which generated a message m for a destination node d at time t_0. We assume that each node $u \in G(t \geq t_0)$ can be a forwarder of a message m according to the store-carry-forward scheme. We define a *social forwarding algorithm* as a store-carry-forward algorithm which uses a social utility function $f(G_s)$ in order to identify the most likely nodes to relay m (*i.e.,* whether send m to the encounter node or

[1] http://www.crawdad.org, http://www.haggleproject.org

not). It spreads the message m among nodes (*relays*) who have specific social properties (relying on $f(G_s)$). Generally, m follows a social forwarding path according to a non-decreasing $f(G_s)$ (*e.g.,* non-decreasing betweenness centrality [10] or non-increasing social distance[2] to destination [13]).

In this paper, we are not proposing and evaluating a new social forwarding algorithm. Rather, our insights are that the relevance of social attributes is as important as designing a social algorithm to improve the efficiency of forwarding decisions. We study in this paper the impact of the quality of different social attributes on the performance of different social forwarding algorithms. Then, we propose different processing methods to augment such social information in order to improve forwarding decisions.

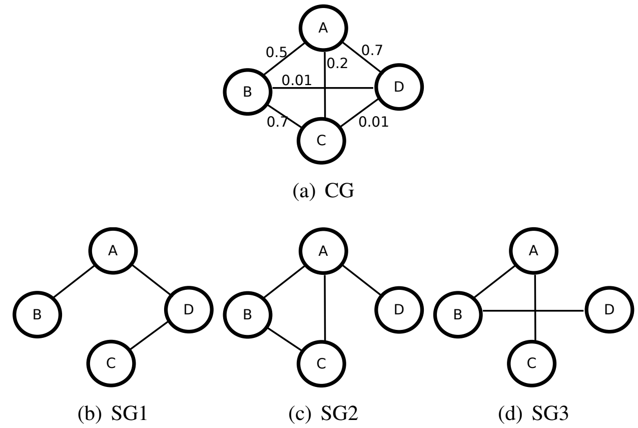
C. Motivation

First, we visualize using simplified examples how different social attributes could behave when used as inputs for social forwarding.

We establish a "stable" representation of the contact graph in order to visualize the structure difference between social and contact graph. Therefore, we define a frequency-dependent contact graph $G_c^\theta(V_\theta, E_\theta)$ where $e_{u,v} \in E_\theta$ if, and only if, the contact rate between u and v is at least θ (we define a contact rate between two nodes u and v in the contact graph as the average number of contacts made between the two nodes in a unit of time). Fig. 1(a) shows a simple example of a "stable" contact graph ($n = 4$) including different contact rates between the four nodes.

Note that, for example, if we consider θ equal to 0.7, $G_c^{0.7}(V_{0.7}, E_{0.7})$ consists only of $e_{A,D}$ and $e_{B,C}$. The resulting graph G^θ could be considered as a social graph as described in [8]- *i.e.,* a *contact graph* $G(t) = (V, E(t))$ is aggregated into a *social graph* $G_s = (V_s, E_s)$. This aggregation guarantees good correlation between social and contact graphs and then better prediction based on social information. However, in reality social attributes may not correlate at all with contact properties; this paper considers this general case. We consider also in Fig. 1 three social graphs connecting the four nodes in Fig. 1(b)-(d). One could notice that frequently seen nodes in the contact graphs have a strongly expected social relationship. For example, a higher contact rate between nodes A and D is translated into a social edge in Fig. 1(b), and Fig. 1(c). However, nodes A and D are not socially connected in Fig. 1(d) which also shows a social connection between nodes B and D although they do not meet frequently (contact rate between nodes A and D is equal to 0.01). We claim that it is critical to study empirically the commonality between social information defined by Online Social Networks and opportunistic contacts created by human mobility.

We define a *Closeness Error* that indicates how closely the social and the contact graph are matching. We measure the matching accuracy between two graphs $G(V, E)$ and



(a) CG

(b) SG1 (c) SG2 (d) SG3

Fig. 1. Examples of a contact graph, and three social graphs connecting 4 nodes

$G'(V', E')$ as follows:

$$CE(G, G') = \frac{|E \setminus E'| + |E' \setminus E|}{|E \cup E'|} \quad (1)$$

CE deals with the fraction of dissimilar edges in the two graphs G and G'. CE takes values between 0 and 1. If the graphs are identical, CE is 0. The more the graphs differ, the more the value of CE is close to 1.

	$\theta = 0.1$	$\theta = 0.5$	$\theta = 0.7$
$CE(CG^\theta, SG1)$	0.6	0.5	0.75
$CE(CG^\theta, SG2)$	0	0.2	0.5
$CE(CG^\theta, SG3)$	0.6	0.6	1

TABLE I
CLOSENESS ERROR

Table I shows the different closeness factors in function of the social graph used. One could notice that the social graph given by Fig. 1(c)) gives the best matching with the contact graphs. Indeed, this social graph is strongly correlated with the contact graph since frequently seen nodes in the contact graph are represented by a social connection in the contact graph. However, the two other social graphs have "anomalies". Note that, the structure difference between aggregated graphs (G^θ) and social graphs may consist of many different edges which could be translated to a bad forwarding prediction if such social attributes are used as inputs for forwarding algorithms.

To summarize, the inputs of social forwarding algorithms are as important as the design of the algorithm itself. In fact, social forwarding algorithms make forwarding decisions based on the social attributes (inputs) as illustrated in Fig. 2 described below. Forwarding algorithms consist basically on:

- The inputs: social information and contact properties.
- The output of forwarding algorithm consists on making a decision to allow (or not) message to be forwarded to an encounter.
- The algorithm itself selects the most likely forwarders of a message based on the inputs (predictors). In this paper, we consider the algorithm as a black box as described in

[2]the social distance defined between two nodes on the social graph (friends have distance 1, friends of friends have distance 2, etc.).

Fig. 2. Our focuses draw up the social information used as inputs for such algorithm (Fig. 3)

Fig. 2. Social Forwarding Algorithms

Our method consists on measuring the relevance of social information for forwarding and process such information in case of inefficiency in order to improve the inputs of the algorithm (as described in Fig. 3). Processing the social information consists on combining different social attributes or social and contact properties to emphasize strong social interaction between two participant.

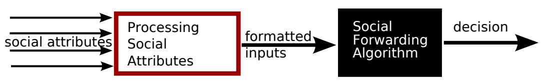


Fig. 3. Processing social data

IV. Priminary observations

In the previous section, we illustrated the impact of the nature of social attributes on forwarding decisions (Fig. 1). In this section, we further develop our first observations examining real mobility traces of a set of participants and their explicit social interactions (for more details, please refer to the Appendix section of this paper). We examine using a formal method to what extend the social attributes are good predictors for human mobility. Then, we verify the impact of different social attributes on social forwarding decisions.

A. Relevance of social attributes

Studing the relevance of social input to opportunistic networks returns to finding an answer the following question: "Is the social information shared by two individuals u and v a *good predictor* for their future physical meetings?"

Bayesian probability theory allows to verify such hypothesis. We apply the *Bayesian inference (BI)* rule to obtain the correlation between the contact frequency (given by the datasets) and the predictive distributions based on our social attributes (predictors). We approach this issue by expressing contact frequency and social relationships in terms of probability.

BI is the statistical inference in which observations are used to infer the probability that a hypothesis may be true. In this paper, we use *BI* to verify the inference of social predictors on human mobility. The Bayesian inference rule is given by:

$$P(I|C) = \frac{P(C|I)\ P(I)}{P(C)} \tag{2}$$

where

- I represents a specific hypothesis, that for two node u and v, $I(u, v)$ is true if and only if they are socially linked relying to this attribute I.
- $P(I)$ is the probability that two participants u and v are socially linked relying to I.
- $P(C|I)$ is a conditional probability that u and v are in contact if the hypothesis I is true. In technical terms,
$$P(C|I) = \frac{\Pi(u, v)}{\Pi} \quad |(u, v) \in E_s$$
where $\Pi(u, v)$ is the accumulated contact time between nodes u and v, and Π is the duration of the entire experiment.
- $P(C)$ is the marginal probability: the probability that u and v are in contact under all possible hypotheses. $P(E) = \sum P(C|I_i)P(I_i)$.

Note that, the higher the value of BI ($BI \in [0; 1]$) the more the contact information accounts for the social relation between the nodes. We apply the *BI* rule, described above in Eq. 2, to identify the most relevant social predictor for contact patterns.

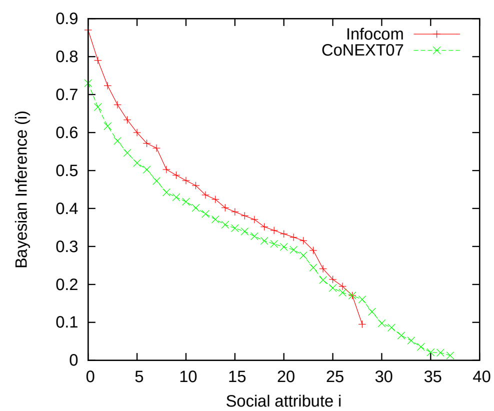


Fig. 4. Bayesian inference of social interests

In the following, we evaluate different forwarding algorithms by analyzing real traces. Specifically, we use the following experimental datasets: MobiClique, and Infocom06. Each dataset includes both, a mobility or contact trace and different social interaction graphs (more details in the Appendix of this paper).

Table II summarizes the characteristics of the used data sets. We differentiate the way the social interactions between participants are defined in each dataset. We define a social connection as *implicit* when the two persons share some common interests (deduced interaction). We define a social connection as *explicit* only when the two nodes declared (self reported) a direct connection - for example links in applications like Facebook.

Fig. 4 plots the Bayesian inference distribution with respect to the social attribute. Note that in each dataset, the social attributes are ranked in the x-axis such that the final distribution is decreasing. We observe that only few social attributes (around 10% to 20 % of the examined social attributes) lead to a good fit (BI greater than 0.5). Moreover, we notice that the most accurate fit in our data sets are given by explicit social attributes, and friend relationship lead to the best fit. Furthermore, more than 20% of social attributes can be

	MobiClique	Infocom(Int.)	Infocom(FB)
duration	3d	3d	3d
# connected nodes	27	65	47
# edges	115	835	219
average degree	9.5	25.7	9.3
median inter-contact	10mn	15mn	15mn
median contact time	240s	150s	150s

TABLE II
DATASET PROPERTIES

considered as inaccurate to predict contact properties of nodes (BI below 0.2). Such social attributes could falsify forwarding decisions instead of guiding to contact prediction.

Next, we measure the impact of the nature of such social attributes on the performances of forwarding algorithms. The performance of a forwarding algorithm is determined by two conflicting factors: (*i*) the message delivery delay; and (*ii*) the *success rate* normalized by the *success rate* of Epedimic forwarding. In the following, we assess the performance with regard to these two performance indicators.

B. Impact of social attributes on forwarding decisions

The trace collection is one part of the challenge. It is also difficult to prepare the data set for later analysis. Key here is to establish a benchmark, *i.e.,* a baseline performance used to compare the different methods. In our case, we have to determine the "optimal" forwarding paths given the mobility patterns and the connectivity properties. We establish this benchmark as follows: We compute offline, for all starting times (the time when the message is ready to be sent) and all source-destination pairs, the set of delay-optimal paths. Paths may use either a single contact, or a sequence of contacts in a time-respecting manner. In other words, we compute the optimal delivery time of a message sent by a node A to a node B ($A \neq B$) for all possible starting times. Then, we compute the success rate as the delivery probability with a given end-to-end delay bound following the same construction rules as described in [6].

We compare three social forwarding algorithms PeopleRank [14], Degree-Based Forwarding [13], and Centrality-Based Forwarding [13] (*e.g.,* Bubble Rap [10], SimBet [7]) using different social inputs. Degree-Based Forwarding, Centrality-Based Forwarding, and PeopleRank use a non-decreasing utility function relying on respectively node degree, node betweenness centrality and PeopleRank metric which was proposed in [14] as a relative importance of a node in a social network.

Fig. 5 plots the *success rate* of the three forwarding algorithms normalized by the *success rate* of flooding as a function of the message delivery delay (two contact traces and different social inputs). In Fig. 5(a)-to-(c), we plot the *success rate* performances of the three forwarding algorithms in the Infocom06 dataset using three different social inputs. In this figure, we examine the impact of social input selection on the forwarding performances.

Clearly, the algorithms perform differently from one social input to another. The friend relationship information improves the performance by roughly 15% for a 10-minutes timescale. Note that links in the social graphs formed by friend relationships (MobiClique experiment) could be considered as strong links because they have been confirmed explicitly by the two individuals (in contrast to interest-based relationship which is an implicit definition of social relationship between two individuals). In the MobiClique experiment, the two social graphs used for comparison rely on *explicit* social interaction which could explain the negligible performance difference for the MobiClique dataset in Fig. 5(d)-to-(f).

Second, one could show that PeopleRank algorithm outperforms the other algorithms according to the same social inputs. In fact, as described in details in [14] PeopleRank is implicitly combining social and contact properties thanks to its opportunistic update process. It favors the most frequently seen friends to update theirs rank more often, and so implicitly making difference between strong and weak ties in the social graph.

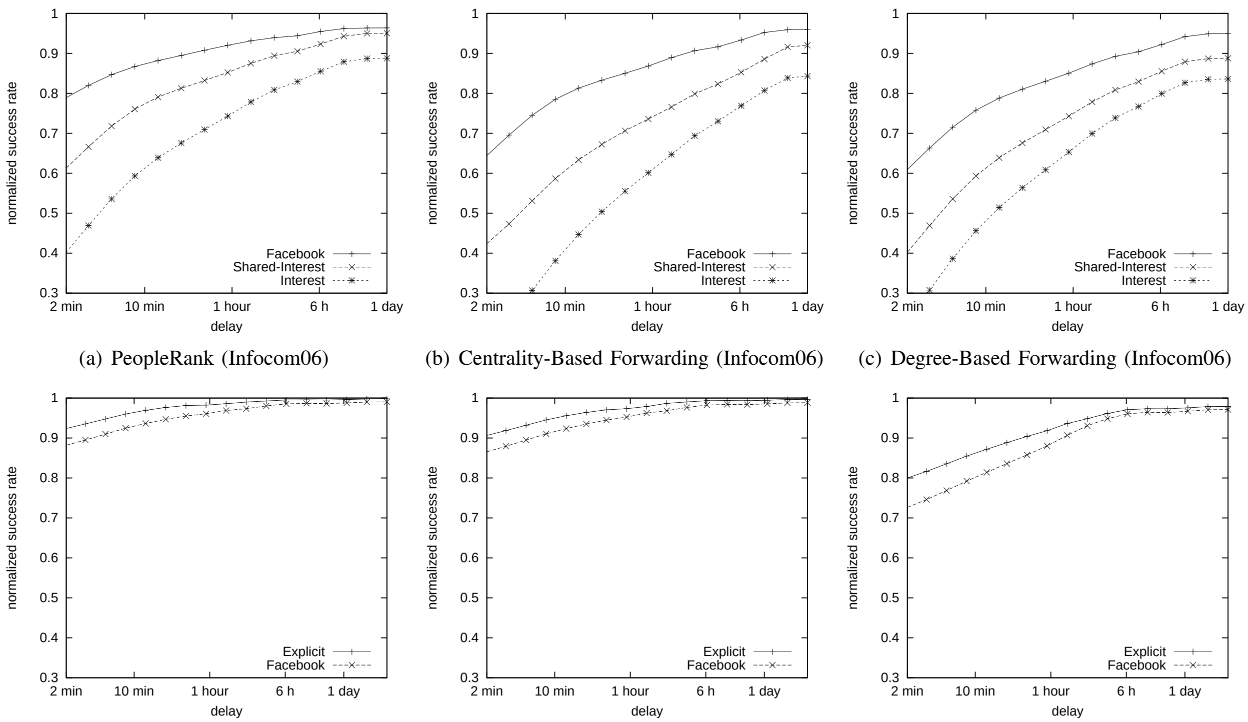
To summarize, the contact graph may be approximated by an "implicit" social graph or an "explicit" social graph. However, "explicit" social inputs are more relevant for forwarding in opportunistic networks. As we showed in the first part of the analysis, there is a structural difference between the two graphs. However, some social attributes are more relevant for forwarding than other. In case of irrelevancy, social inputs may falsify the forwarding prediction. Next, in the remainder of this paper we try to exploit this by processing the different social inputs in order to improve the forwarding decisions.

V. PROCESSING THE SOCIAL INFORMATION

We describe methods to process and improve the social inputs and study the impact of such processing on social forwarding performances (end-to-end delay and *success rate*).

A. Processing the Social Information

In order to improve the matching of the social and the contact graph, one has to identify the most relevant links in the social graph. There are basically two ways of achieving this: either one reduces the number of edges in the graph such that only the strongest links remain in the graph, or the processing step has to put more weight on the contacts that represent strong interaction. Although the first technique is simpler, we believe that the second technique is more efficient since the weights are used to rank different social connections to neighbors. So, forwarding algorithms may use this rank to



(a) PeopleRank (Infocom06) (b) Centrality-Based Forwarding (Infocom06) (c) Degree-Based Forwarding (Infocom06)

(d) PeopleRank (MobiClique) (e) Centrality-Based Forwarding (MobiClique) (f) Degree-Based Forwarding (MobiClique)

Fig. 5. Impact of social attributes on forwarding

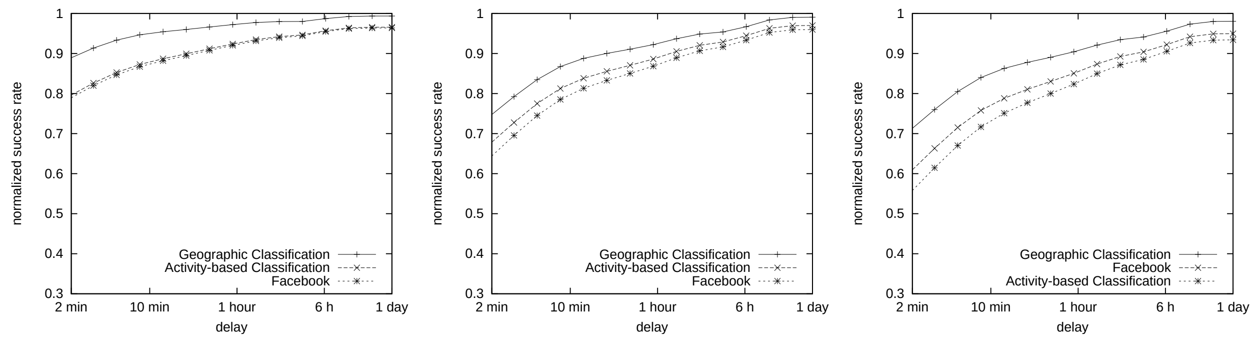
identify the set of nodes that are more likely to forward the message to the destination. To implement the second approach, we consider a weighted social graph G_s, where we associates a weight $w_{(i,j)}$ to every edge $e_(i,j)$ in the graph. Weights are computed according to the additional social links (if there are any). In this paper, we consider additional social attributes such as interests, affiliation, nationality, etc., as well as contact properties of nodes such as contact rate, contact time, and inter-contact time to compute the weight of a social interaction between two nodes in the social graph.

1) Combining multiple social attributes: In the previous section (more specifically in Fig. 4), we have shown that some social inputs lead to poor contact prediction and impact negatively the final forwarding decision. In this section, we describe a technique to augment such social information by combining multiple social sources. There are many ways to combine the available social information. In the following we describe the two methods considered in this paper:

- **Geographic Classification**: Since the dartmouth campus area is roughly 1300x1300 square meters, people attending the campus every day are mostly visiting the same places. Usually, these places are selected in the way that minimizes the walking distance. To capture this classification, we looked at splitting the Dartmouth campus area into regions (Northwest *NW*, Northeast *NE*, Southwest *SW*, and Southeast *SE*)[3]. A node i belongs to a region R if it has been connected to more access points belonging to the corresponding region compared to the other regions.
- **Activity-Based Classification**: The Dartmouth College campus has over 160 buildings. Usually people visiting the campus are interested in few buildings. People could be classified relying on their activity interests. For example, the campus contains more than a dozen athletic facilities and fields. Most of them are located in the southeast corner of campus. Athletic people are more likely to meet each others and be classified in an athletic community. We consider people more connected to athletic building's access points as part of the athletic community. Similarly we define academic and residential communities.

Fig. 6 plots the normalized *success rate* of the three forwarding algorithms after processing the social inputs with the two methods described before (Geographic, and Activity-based classification). One may notice that the Activity-Based Classification does not improve considerably algorithms' performances. It even decreases the *success rate* of the Degree-Based Forwarding by 3% for the 10 minute timescale (see Fig. 6(c)). Indeed, combining did not lead always to a performance improvement, and it was shown that combing the two most relevant social inputs even decreased the forwarding performance. However, Fig. 6 shows that the second method improves the matching of the social information and the performances of our three forwarding algorithms. The forwarding algorithms outperform the original performance (*i.e.,* without combination) by about 10 to 20% for the 10 minute timescale. We find that with the Geographic Classification-, the processed information enforces the common edges (strong ties) in the social graphs and extends it with complementary edges which

[3]http://www.dartmouth.edu/~maps/campus/close-ups/index.html



(a) PeopleRank (b) Centrality-Based Forwarding (c) Degree-Based Forwarding

Fig. 6. Impact of combining social attributes on forwarding (Infocom06)

are considered as weak ties in the social graph.

2) Combining social and contact properties: The "pure" social forwarding algorithm described above favors socially well connected nodes. However, such social information defined by online social network applications does not always lead to a very accurate prediction of physical contact opportunities. In fact, people interact with each other in different ways; some have regular physically meetings, some are connected only in a virtual online environment, and some are connected without ever communicating with each other. Since the purpose of this study is to propose a solution for improved message forwarding, we came up with the idea to combine the use of contact information and social networking information. Specifically, we consider the contact rate as a weight of social interaction between two nodes; $w_{(i,j)}$ is the fraction of times node i and node j were in contact with each other. More formally:

$$w_{(i,j)} = \frac{\pi(i,j)}{\sum_{j \in F(i)} \pi(i,j)} \quad (3)$$

where $\pi(i,j)$ denotes the number of times node i and node j are in contact (*i.e.,* they get close to each others) with each other, and $F(i)$ is the set of social neighbors that links to node i.

In Fig. 7, we plot the normalized *success rate* of the three forwarding algorithms with reference to the combination of social and contact properties in the Infocom06 data set. We notice that by augmenting the social inputs with contact properties, the performance of the three considered algorithms is improved by 10 to 20% (in term of *success rate* compared to the performance of the original social input within the 10 minute timescale). Moreover, the combined solution leads to a better performance than the pure combination of social information (1 to 3% improvement of the *success rate*).

Note that PeopleRank favors implicitly frequently seen social neighbors in its update process. It uses implicitly contact properties in addition to social inputs to make forwarding decisions.

B. Randomize social inputs

As discussed before, in some settings, the social graph does not at all match the contact graph of the participants. In those cases, some randomness in the structure of the social graph helps to improve the forwarding performance and to overcome the inaccuracy of the social graph.

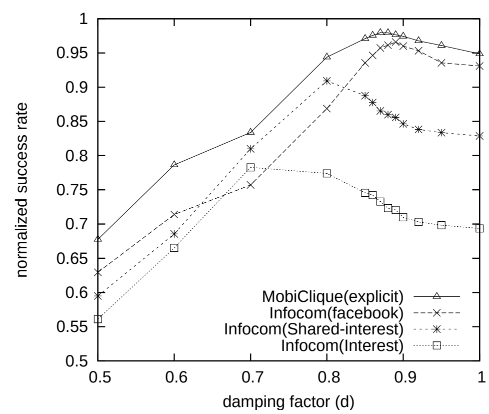
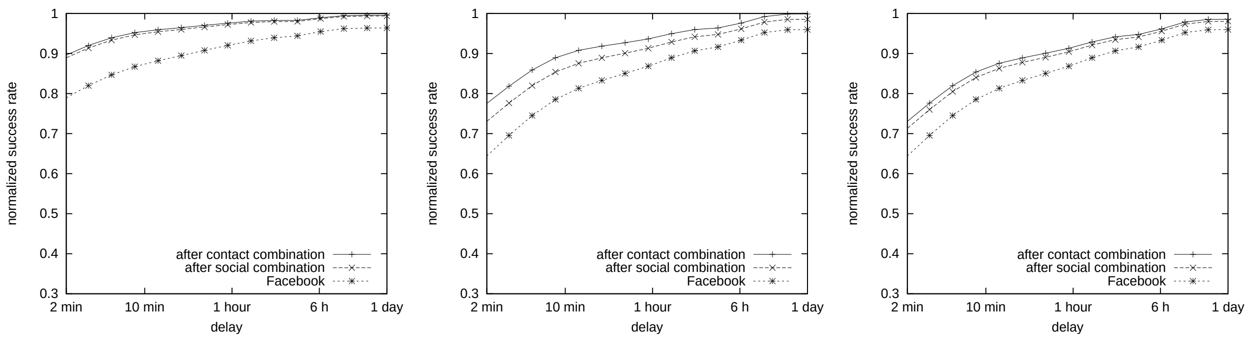


Fig. 8. Impact of damping factors

As described in [14], PeopleRank is a social forwarding algorithm where a damping factor d is used to control the amount of randomness introduced in the forwarding decision. Specifically, PeopleRank is based on the following formula to determine the social rank of every node:

$$PeR(i) = (1-d) + d \sum_{j \in F(i)} \frac{PeR(j)}{|F(j)|}$$

where $F(i)$ is the set of social neighbors that links to node i. When d is equal to 0, all nodes are used for forwarding with the same probability, values close to 1 will prefer the socially best connected nodes. Fig. 8 plots the *success rate* of PeopleRank normalized by the *success rate* of flooding within a delay of ten minutes for different damping factor values. It can be seen from this plot that optimal damping factor values change with the different traces. We conjecture that the reason for this difference lies in the accuracy of the social inputs used for PeopleRank forwarding. Fig. 8 also shows that the curves decrease for a damping factor close to 1. This effect can be explained by the fact that a damping factor d equal to 1 considers exclusively socially connected nodes and could miss other potential forwarders such as familiar



(a) PeopleRank (b) Centrality-Based Forwarding (c) Degree-Based Forwarding

Fig. 7. Impact of combining contact properties on forwarding (Infocom06)

strangers introduced in [12]. This means that even when social inputs considered exhibit an optimal correlation to contact graph, some randomized forwarding decisions are beneficial.

Inspired by the damping factor idea for PeopleRank , one could define a probabilistic social graph $G_s^d(V, E^d)$ such that edges between two nodes i and j exit with a probability d if i and j are socially connected ($e(i, j) \in G_s$) and with a probability $1 - d$ otherwise.

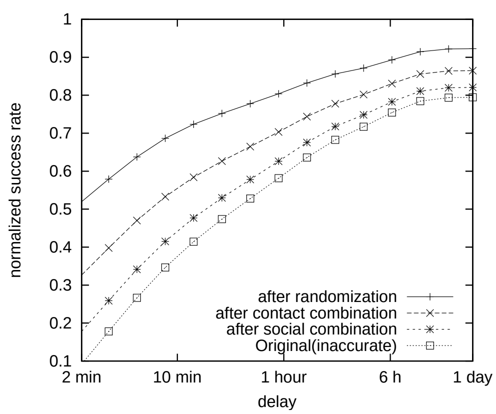


Fig. 9. Impact of social randomization technique on the centrality-based algorithm

To evaluate the randomization technique, we consider the most inaccurate social attribute for the Infocom dataset (given by Fig. 4). In this specific case, we compare the previous combination techniques to randomization in Fig. 9. As expected, the original input lead to the worst performances where the normalized *success rate* of the centrality-based algorithm remains below 35% for the 10 minute timescale. Moreover, combination techniques described previously does not improve the *success rate* performance by more than 10 to 20% within the same timescale (which correspond to a *success rate* of less than 50%). Indeed, the original social graph may contain inaccurate social edges that lead to wrong contact prediction. Previous techniques consist on emphasizing the strong ties from the weak ones. However, when the original social information is mainly inaccurate, the previous processing techniques are not able to deal with an improvement of the forwarding performance. Fig. 9 shows that randomization is an option in case of matching inaccuracy. It helps the forwarding algorithm to outperform its original performance by more than 35%.

C. Limitations

In large networks, the transmission of messages through the most socially important people will ultimately consume most of theirs resources. Moreover, it is hard to defend the assumption that a subset of socially well ranked nodes will meet physically all other nodes in very large networks. Next, we verify this assumption in a large network with multiple social communities. We highlight the limitation of some social forwarding algorithms in large scale networks. Then, we propose a technique which spreads the messages among multiple communities.

We have used the WiFi access network of Dartmouth college [1] to consider a larger experimental data set. We assume that two nodes are able to communicate if they are attached at the same time to the same access point (The Dartmouth college wireless network is composed of about 550 access points). The Dartmouth college covers multiple student residences, sport infrastructures, administrative buildings, and academic buildings.

Since here we are measuring the scalability of social forwarding algorithms, we considered an optimal scenario where the social inputs are well correlated to the contact patterns (collected from real mobility data set). We artificially created multiple social graphs with regard to the contact rates between the nodes. We have seen in the previous sections that such accurate social inputs lead to a good forwarding performance in "single community" data sets.

Fig. 10 plots the normalized *success rate* of the three forwarding algorithms with respect to the Dartmouth data set. One could notice that within a 10 minute timescale 25% to 55% more losses than with Epidemic forwarding despite the fact that social inputs match with the contact properties of nodes. In fact, due to the social forwarding algorithm, many contact opportunities are not used as the nodes are not in the same social network. In next section, we present a two step technique that solves the scaling and this performance issue.

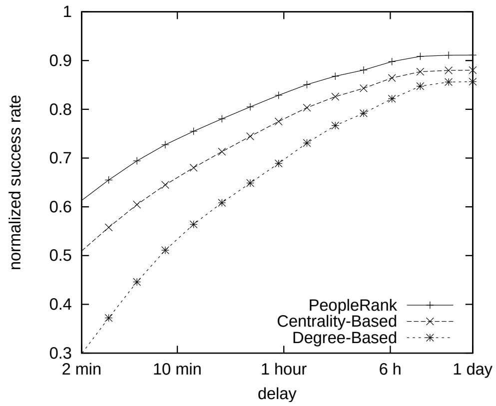


Fig. 10. Scalability issues

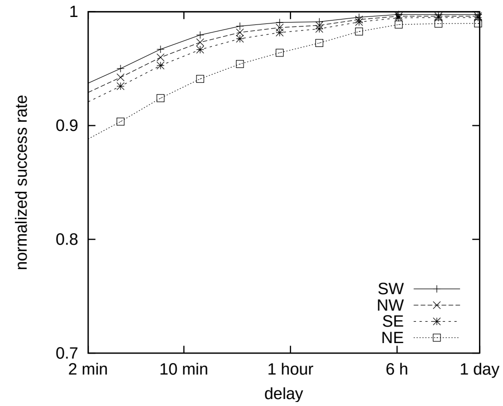


Fig. 11. Forwarding performances per community

VI. MULTI-COMMUNITIES SOCIAL FORWARDING

Forwarding in mobile adhoc networks faces extreme challenges from potentially very large number of mobile nodes, and limited communication resources such as bandwidth and energy. Such conditions make forwarding more challenging in large scale networks. In previous section (Fig. 10), we have noticed that using social inputs in large scale areas may present weaknesses. Our main guess is that in large scale networks when multi-communities (let's assume that communities represents cities in real world) may exist social prediction present limitation and two people socially connected may not meet frequently because they could be long away from each others.

Fist of all, we verify the performances of social forwarding in a single community. To do so, we consider different communities in the Dartmouth data set (described in details in the previous section). We take into consideration two community classifications:

- Geographic classification: we look at splitting the Dartmouth campus area into regions (Northwest *NW*, Northeast *NE*, Southwest *SW*, and Southeast *SE*)[4]. A node i belong to a region R if it has been connected to more access points belonging to the corresponding region compared to the other regions.
- Activity based classification: we consider people more connected to residential building's access points as part of the residential community. Similarly we define academic and athletic communities.

Fig. 11 plots the normalized *success rate* of PeopleRank per geographic community in Dartmouth data set. PeopleRank acheave above 92 to 97% of the optimal *success rate* given by flooding within 10 minutes timescales (it represents 20% more *success rate* compare to the multi-communities performance given by Fig. 10).

We were motivated by satisfactory performances of social forwarding within single community, we propose a two step forwarding mechanism. Such mechanism could be applied easily to almost all social forwarding algorithms, and consists of:

- Step1: Social forwarding algorithms operate normally within the same community. Indeed, messages will be forwarded socially toward nodes which belong to the same community.
- Step2: Particular nodes will operate as a bridge and circulate the message to the other communities, and within these communities messages will be forwarded socially (Step1). Bridge nodes are characterized by high mobility in the dataset, and may belong to many communities (*e.g.,* according to the first definition of community, bridges are nodes moving around the center of the campus).

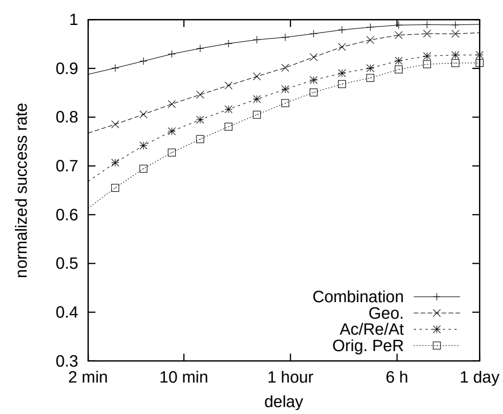


Fig. 12. Two steps forwarding

Fig. 12 plots the normalized *success rate* of two steps PeopleRank algorithm in Dartmouth data set. One could notice that the two step algorithm outperforms the original one for all timescales (5 to 30% of *success rate* improvement). Furthermore, such improvement differs from one definition of communities to another. Geographic definition gives better performances because in the Dartmouth campus building are not regrouped geographically[3], and then message could take time to reach other member of the same community. However, combining the two definition of communities leads with the best *success rate* and more than 30% improvement compared to the second definition.

[4]http://www.dartmouth.edu/~maps/campus/close-ups/index.html

VII. Concluding Remarks

The rise of online Social Network platforms and applications such as Facebook, Orkut, or MySpace, makes information about the social interaction of user "easily" accessible (online, or directly through the information stored on mobile devices). In opportunistic networks, such information can then be used to predict future encounters of participating devices. In this paper, we have studied how different definitions of social connections between participants impact the performances of social forwarding algorithms.

We have measured the matching accuracy between participants' contact patterns and their social relations, considering cases where social information is defined explicitly by the user and cases where links between participants is implicitly deducted using common interests. Our results are based on human contact data sets and the associated social information of the experimentalists.

To summarize our findings, social inputs may be considered to guide and improve forwarding decisions in mobile opportunistic networks. However, in some cases social interaction information alone is not sufficient for forwarding and needs to be processed in some way with additional information. We have proposed and tested different rules to process these social attributes in order to augment their relevance for opportunistic forwarding. Indeed, we have shown that processing social inputs may improve the original *success rate* performance by 40%.

However, despite the fact that social information are used as a good predictor for human mobility, social forwarding suffers in very large networks where multiple communities and multiple social graph within the communities exist. We have proposed a two step technique which could be integrated to most social forwarding algorithms in order to solve the scaling issues. It was shown that such a technique helps social algorithms to spread messages among the communities and uses the social properties within communities to predict future contacts. It helps social forwarding algorithms to outperforms their original performances (considering a single community social graph) by roughly 30%.

References

[1] E. Anderson, K. Eustice, S. Markstrum, M. Hansen, and P. Reiher. Mobile contagion: Simulation of infection and defense. In *PADS '05: Proceedings of the 19th Workshop on Principles of Advanced and Distributed Simulation*, pages 80–87, Washington, DC, USA, 2005. IEEE Computer Society.

[2] H. R. Bernard, P. D. Killworth, and L. Sailer. Informant accuracy in social-network data v. an experimental attempt to predict actual communication from recall data. *Social Science Research*, 11(1):30–66, March 1982.

[3] G. Bigwood, D. Rehunathan, M. Bateman, T. Henderson, and S. Bhatti. Exploiting self-reported social networks for routing in ubiquitous computing environments. In *WIMOB '08: Proceedings of the 2008 IEEE International Conference on Wireless & Mobile Computing, Networking & Communication*, pages 484–489, Washington, DC, USA, 2008. IEEE Computer Society.

[4] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.*, 30(1-7):107–117, 1998.

[5] B. Burns, O. Brock, and B. Levine. MV routing and capacity building in disruption tolerant networks. 2005.

[6] A. Chaintreau, A. Mtibaa, L. Massoulié, and C. Diot. The diameter of opportunistic mobile networks. In *Proceedings of ACM CoNext*, 2007.

[7] E. M. Daly and M. Haahr. Social network analysis for routing in disconnected delay-tolerant manets. In *MobiHoc '07: Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing*, pages 32–40, New York, NY, USA, 2007. ACM.

[8] T. Hossmann, T. Spyropoulos, and F. Legendre. Know thy neighbor: Towards optimal mapping of contacts to social graphs for dtn routing. In *Proceedings of IEEE INFOCOM*. IEEE, 2010.

[9] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot. Pocket switched networks and the consequences of human mobility in conference environments. In *WDTN*, 2005.

[10] P. Hui, J. Crowcroft, and E. Yoneki. Bubble rap: Social-based forwarding in delay tolerant networks. In *Proceedings of ACM MobiHoc '08*, New York, NY, USA, 2008. ACM. (also appeared as Cambridge University TR: UCAM-CL-TR-684).

[11] D. Kempe, J. Kleinberg, and A. Kumar. Connectivity and inference problems for temporal networks. In *Proceedings of ACM STOC*, 2000.

[12] S. Milgram. The individual in a social world : essays and experiments. In *Reading, Mass.: Addison-Wesley Pub. Co.*, 1977.

[13] A. Mtibaa, A. Chaintreau, J. LeBrun, E. Oliver, A.-K. Pietilainen, and C. Diot. Are you moved by your social network application? In *WOSN'08: Proceedings of the first workshop on Online social networks*, pages 67–72, New York, NY, USA, 2008. ACM.

[14] A. Mtibaa, M. May, M. Ammar, and C. Diot. Peoplerank: Social opportunistic forwarding. In *Proceedings of IEEE INFOCOM (mini conference)*. IEEE, 2010.

Appendix

We have chosen to evaluate our forwarding algorithm using analysis on real traces. In the following, we present the datasets used to evaluate the performances of opportunistic forwarding algorithms.

MobiClique data [13] Contains mobility information and information about the social relation between the participants. Visitors of a the CoNext conference were asked to carry a Smartphone device during three consecutive days with the MobiClique application installed. Prior to the experiment start, each participant was asked to indicate the participants of all CoNext participants he knew or had a connection to. During the experiment, the social networking application indicated when a friend, or a friend-of-friend, was in Bluetooth range. The MobiClique dataset was collected on 28 Windows Mobile devices that were given to a preselected set of participants the first day of the conference. Each device was used for an average of 2.2 days since people arrived and left at different times.

Infocom06 dataset [6] was collected with 78 participants during the IEEE Infocom 2006 conference. People were asked to carry an experimental device (*i.e.,* an iMote) with them at all time. These devices were logging all contacts between participating devices (*i.e.,* called here internal contacts) using a periodic scanning every 2 seconds. In addition, they logged connections established with other external Bluetooth-enabled devices (*e.g.,* cell phones, PDAs). For this study, we are using results for internal contacts only. Questionnaires were given to participants to fill theirs nationalities, languages, countries, cities, academic affiliations and topic of interests. Based on theses information, we consider three different social graphs for this experiment; based on (1) their common topics of interest when two users are sharing k common interest, (2) their Facebook connectivity (obtained offline), and (3) their social profile (union of nationality, language, affiliation, and city).