Social Forwarding in Large Scale Networks: Insights Based on Real Trace Analysis

Abderrahmen Mtibaa
School of Computer Science
Carnegie Mellon University
Email: amtibaa@cmu.edu

Khaled A. Harras
School of Computer Science
Carnegie Mellon University
Email: kharras@cs.cmu.edu

***Abstract*—**
Social forwarding, recently a hot topic in mobile opportunistic networking, faces extreme challenges from potentially large numbers of mobile nodes, vast areas, and limited communication resources. Such conditions render forwarding more challenging in large-scale networks. We observe that forwarding techniques based on social popularity fail to efficiently forward messages in large scale networks. The social popularity of nodes might not scale with the network size in a way that necessarily correlates with the contact opportunities and mobility patterns of these nodes. In this paper, we demonstrate, based on real mobility traces, the weakness of existing social forwarding algorithms in large scale communities. We address this weakness by proposing strategies for partitioning these large scale communities into sub-communities based on geographic locality or social interests. We also examine exploiting particular nodes, named MultiHomed nodes, in order to disseminate messages across these sub-communities. We introduce *CAF*, a Community Aware Forwarding framework, which can easily be integrated with most of the state-of-the-art social forwarding algorithms, in order to improve their performance in large scale networks. We use real mobility traces to evaluate our proposed techniques. Our results empirically show a performance increase of around 40% and 5% to 30% better success delivery rates compared to state-of-the-art social forwarding algorithms, while incurring a marginal increase in cost.

***Index Terms*—CAF, Social Forwarding, DTN, PeopleRank, scalability, mobility traces.**

I. INTRODUCTION

The proliferation of a new generation of powerful mobile devices has led to the rise of new infrastructure-less based communication paradigms and applications. This new communication environment is characterized by a variety of new challenges such as mobility, disconnections, and energy constraints. While researchers in the area of Mobile Ad Hoc Networks [13] (MANETs) have traditionally addressed some of these challenges, their solutions fail in scenarios where end-to-end paths may not exist. Delay Tolerant Networks [12] (DTNs) in general, and opportunistic networks in particular, have recently attempted to address this failure through a variety of message *store-carry-and-forward* techniques. The most pressing concern in these types of opportunistic networking solutions, is the fundamental problem of deciding on when to forward a message and who this message should be forwarded to. How to optimally select the next hop towards a destination in a way that minimizes delay and maximizes success rate is so far unknown.

One approach to the message forwarding challenge that recently received increased attention, is to exploit social networking properties to opportunistically forward messages [17], [15], [16], [14]. Nowadays the number of social networking websites exploiting friendship links, such as Facebook, LinkedIn, and MySpace, is constantly growing. Furthermore, mobile phones provide constant Internet access and allow for a continuous maintenance of these online social networking websites. Social interaction between people can largely be used as a good predictor for human mobility. Social information can therefore be used to optimize opportunistic forwarding decisions in Mobile Opportunistic Networks.

Our work addresses scalability issues of existing opportunistic social forwarding algorithms. In large scale opportunistic networks, the transmission of messages through the most socially popular people, as proposed in the BubbleRap algorithm [11], will ultimately consume most of their device resources. Moreover, it is hard to defend the assumption that a subset of socially high ranked nodes will physically meet all other nodes in large scale networks. We validate this intuition based on real life traces; we show, in section III, that state-of-the-art social forwarding algorithms achieve satisfactory performance within small communities such as conferences, campuses, etc. but suffer in large scale networks.

This paper contributes to a better understanding of the weaknesses of existing social forwarding algorithms in large scale Mobile Opportunistic Networks, and proposes insights to deal with such issues. We propose partitioning large scale communities into multiple sub-communities based on various common social characteristics such as locality or social interests. We introduce (in section IV) *CAF* a Community Aware Framework which can easily be integrated with most of the existing social forwarding algorithms, in order to improve their performance in large scale networks. CAF uses particular nodes called *MultiHomed* nodes to disseminate messages across all the sub-communities in the network. The original social forwarding algorithm can then behave normally within a local sub-community. Besides the simplicity of CAF, it uses a relatively negligible overhead compared to the overhead induced by state-of-the-art algorithms, such as BubbleRap, to compute the global node ranking in the large scale network. CAF remains a distributed forwarding algorithm and relies on a local social/contact information to estimate future transfer opportunities.

A major contribution in our work is based on the fact that our insights, evaluation, and analysis of social forwarding algorithms as well as the CAF framework, are all based on real mobility traces. We utilize in this paper the largest data set (To the best of our knowledge) that captures human mobility contacts in large scale networks in addition to the corresponding social information. Our results, in section V, show that we obtain a performance increase of around 40% compared to the state-of-the-art social forwarding algorithms, while incurring a marginal increase in cost. We also show that CAF outperforms BubbleRap and achieves 5% to 30% delivery rate improvement.

II. Related Work

Mobile communication opportunities between nodes in Opportunistic Networks are intermittent in nature and end-to-end paths between a source and a destination may never exist. Since node contacts are mostly unpredictable, scheduled relay approaches such as Message Ferrying [24] could not be effective. Replication is the most common technique to maximize the number of successful message delivered. Naive forwarding protocols based on flooding are extremely inefficient [25], [22]; because flooding is very costly in terms of resource and energy consumption. Most of the work done is on designing controlled flooding algorithms to reduce the number of replica copies in the network, and achieving a satisfactory delivery rates. Proposed algorithms use a "utility" metric to make forwarding decisions based on contact information [4], [5], [2], [18], [8], probabilistic schemes [18], or social properties [10], [11], [7], [19], [21], [9], [20].

While contact and learning based forwarding schemes have been quite popular in literature, there has been much less work on social based forwarding. We classify existing social forwarding algorithms as follows:

Degree-Based Forwarding consists of forwarding messages to socially well connected nodes. Paths are constructed according to a non-decreasing social node's degree rule [19].

Centrality-Based Forwarding builds on the idea that central nodes in social graphs are more likely to socialize with other people and therefore suitable to forward messages to the destination [7], [19]. Simbet [7] is a well known Centrality-Based Forwarding algorithm which uses potential nodes to forward messages to destination based on their centrality characteristics.

PeopleRank is a fully distributed algorithm that ranks nodes in a social graph similar to what PageRank [3] does for web pages - *i.e.,* it measures the relative "importance" of a node in a social graph. Message forwarding decisions can then follow a non-decreasing rank rule [21].

Most of these contributions generally highlight the superior performance of social forwarding algorithms within specific communities such as conferences, campuses, etc. Our goal in this paper is to: (*i*) *emphasize the weakness of such algorithms in large scale networks* and (*ii*) *propose a technique, which can easily be integrated with most social forwarding algorithms, in order to improve the success rate in large scale networks.*

Studying the scalability of forwarding algorithms in large ad hoc networks is not a new research topic [10], [11]. Previous work that study opportunistic forwarding in large scale networks have focused on the "*mobility properties*" [24], [26], [27], [10]. However, connecting "*social characteristics*" of individuals and their mobility to classify them into communities remains largely unexplored. In order to interconnect isolated regions (*i.e.,* communities) in a large scale network, pre-scheduled relay approaches such as Message Ferrying [26] have been proposed; special mobile nodes called "ferries" aid connectivity between the nodes in the network. Since mobility is, in general, unpredictable in opportunistic networks, scheduled approaches could not be effective.

Most relevant to this work, BubbleRap [10], [11] is a forwarding algorithm that uses *contact properties* of node to estimate nodes' popularity and classify nodes in communities. Besides the fact that the computation of nodes' centrality and communities are deduced from contact properties (and evaluated with the same contact trace), BubbleRap assumes that each node has a global ranking across the whole network; we believe that such assumption is surrealistic in a large scale environment. In our work, we do not rely on such simplistic assumptions, and use *explicit* and *local* social interactions between individual to form communities and disseminate messages across these communities. We show in our evaluation how our Community Aware Forwarding framework (CAF) outperforms BubbleRap in most scenarios.

III. Social Forwarding Weakness in Large Scale Networks

In this section, we first present a brief overview on the idea behind social forwarding algorithms. We then describe the drawbacks of the current state-of-the-art algorithms with large scale networks using a motivating scenario. Finally, we confirm our observations and quantify these drawbacks using experimental validation and analysis.

A. Social Forwarding Overview

In social-based opportunistic networks, we are generally 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 social relationships between mobile nodes using a non-time varying graph, which we denote as $G_s = (V_s, E_s)$. Social graphs reflect the interaction or interrelation between people/nodes. 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).

A *social forwarding algorithm* is a store-carry-forward algorithm which spreads a message M among nodes (*relays*)

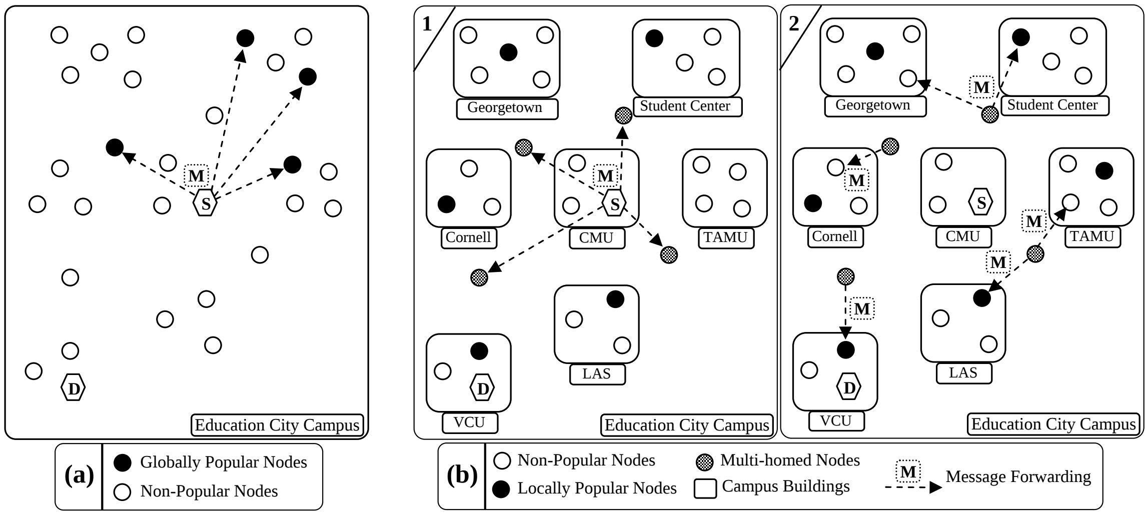


Fig. 1: Example showing (a) the weaknesses of existing social forwarding algorithms in a large scale community, and (b) identifying sub-communities within the same large scale one, and using MultiHomed nodes to disseminate messages to these sub-communities.

who have specific social properties. In this paper, our analysis is based on three existing social forwarding algorithms; Degree-Based Forwarding [19], Centrality-Based Forwarding (*Simbet*) [7], and PeopleRank [21]. Various social forwarding mechanisms differ by the method used to predict device mobility and future contact opportunities between devices based on simple social properties. These approaches, which use social properties to forward messages, *implicitly* assumes that opportunistic contacts correlate with the social property upon which the algorithms are designed. In large scale networks, where distances have larger impact such contact opportunities, it is hard to defend such an assumption.

B. Motivating Scenario

Fig. 1(a) illustrates a scenario where social forwarding attempts to disseminate a message M generated by a source S to the destination D. Without any notion of communities, M is forwarded in the wrong direction relying on "globally popular nodes" in the network (*i.e.,* the BubbleRap technique). Theses particular nodes, although popular, may not be able to deliver the message to all the nodes in the network. In the education city (EC) (a campus that includes 6 US university branches in Doha, Qatar), the founder of EC could be a very popular person in the whole campus (Globally Popular Node), but not likely suitable to relay the message M to a student in a particular university on campus. However, other nodes may be locally popular (*e.g.,* within a university of the campus as shown in Fig. 1(b)) and more suitable to deliver this message to its destination in a specific sub-community. Therefore, particular nodes which we call MultiHomed nodes such as postmen or campus bus drivers are more suitable to disseminate the message M across all sub-communities. This approach is opposed to the BubbleRap algorithm that uses globally popular nodes to disseminate the message to all communities. Therefore, the main idea (shown in Fig. 1(b)) is to first breakdown the original large scale community into multiple sub-communities, then disseminate the message to these sub-communities. Afterwards, locally popular nodes can then deliver the message M within its sub-community.

C. Experimental Validation

In this section, we highlight the weaknesses of existing social forwarding algorithms in large scale networks relying on two experimental data sets. We run analysis on the following experimental data sets (Table I summarizes the characteristics of the used data sets. More details can be found in [21], and CRAWDAD[1].):

Dartmouth: We use the WiFi access network of Dartmouth campus [1]. This data set spans roughly 1300x1300 square meters and over 160 buildings, and about 550 802.11b access points throughout. Dartmouth college covers student residences, sport infrastructures, administrative buildings, and academic buildings. The data set contains logs of client MAC addresses, and SSIDs of access points as well as their positions. We assume that two nodes are able to communicate if they are connected at the same time to the same access point. We use this trace to generate contacts between 100 nodes in order to simulate the message propagation in a pure ad-hoc manner. Note that the ping-pong effect in the Dartmouth trace [1] will not affect such assumption.

San Francisco: To the best of our knowledge, this is the largest data set that captures human mobility contacts as well as human social properties in large scale networks. We use the San Francisco taxi trace [23], and combine it with three other existing data sets to represent three different communities. The San Francisco taxi trace contains mobility traces of taxi cabs in San Francisco. It contains GPS coordinates of approximately

[1] crawdad.cs.dartmouth.edu/

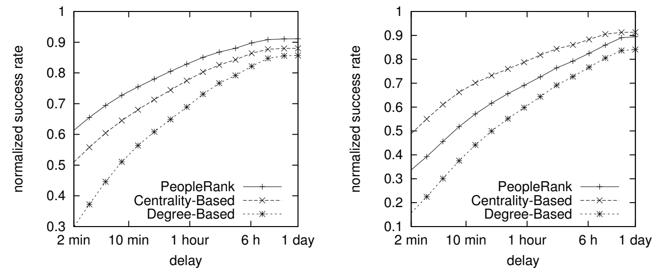
	Dartmouth	San Francisco		
		MobiClique	Infocom06	Conext08
duration	3 days	3 days	3 days	3 days
mobility pattern	WiFi	Bluetooth	Bluetooth	Bluetooth
# nodes	100	27	47	22
median inter-contact	6 mins	10 mins	15 mins	12 mins
median contact time	160 sec	240 sec	150 sec	120 sec

TABLE I: Dataset properties

500 taxis collected over 30 days. Each trace contains the reported time and location for each taxi. We incorporate traces for the duration of 3 days, interpolate the movement of the cabs, and then generate the contacts between these taxis. We assume a contact has occurred when a taxi comes within a proximity of 100 meters from another taxi. In this data set, taxis are connecting different zones of the San Francisco area such as the airport, downtown, and the sunset area. We artificially add real human mobility patterns in each of this areas based on real traces (see Table. I for more detail). To summarize, the resulting data set contains 3 sub-communities in three different areas of San Francisco. Taxis are moving between these areas. Contacts between taxis and nodes within an area are added based on the same contact time and inter-contact time distribution [6] of the corresponding area.

Our evaluation methodology: We evaluate the performance of three state-of-the-art social forwarding algorithms in large scale networks relying on the two previously described data sets; Dartmouth and San Francisco. In our evaluation, we compute the sequence of optimal paths found between any source and destination in the data set. From the sequence of delay-optimal paths we deduce the delay obtained by the optimal path at all time. We uniformly combine all the observations of a trace among all sources, destinations, and for every starting time (the time in seconds when the message M was generated by the source node S). We present this aggregated sample of observations via its empirical CDF. We plot the success rate of the three social forwarding algorithms normalized by the success rate of flooding as a function of the message delivery delay. The detailed computation process could be found in [6]. Compared with previous generalized Dijkstra's algorithm, this algorithm directly computes representation of paths for all starting times. In our experimental evaluation, we utilize the following metrics to evaluate a given forwarding algorithm f: (*i*) the *normalized success rate within time* t: the probability of f to successfully deliver the message to its destination within time t normalized by the same probability given by epidemic forwarding algorithm (optimal success rate within the same time t), and (*ii*) the *normalized cost*: the fraction of contacts (*i.e.,* number of replica copies) used by f normalized by the fraction of contacts used by epidemic forwarding algorithm (the most expensive).

Fig. 2(a) plots the normalized success rate of three social forwarding algorithms (PeopleRank, Degree-Based, and Simbet) with respect to the Dartmouth data set. Note that the value of the CDF (Cumulative Distribution Function) for a given time t is equal to the probability to successfully find a path within time t, when sources, destinations and message generation time are chosen at random. If no path exists, we include an infinite value in the distribution. We then normalize by the CDF given by an epidemic algorithm (flooding).



(a) Dartmouth data set (b) San Francisco data set

Fig. 2: Scalability issues of social-based algorithm relying on experimental data sets.

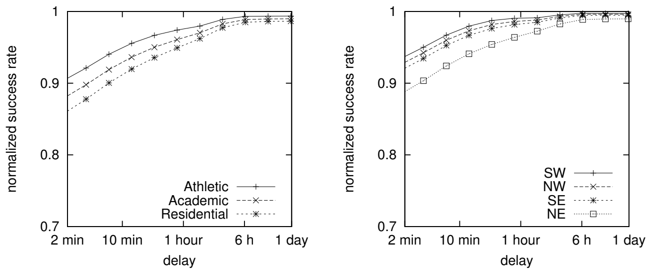
Despite the fact that social inputs match the contact properties of nodes, there are 25% to 55% of losses compared to Epidemic forwarding, within a 10-minutes timescales. In fact, in large scale networks, social forwarding algorithms loose many opportunities to reach destinations in optimal delays. Similar results are observed using the San Francisco data set. Fig. 2(b) shows that the three considered social forwarding algorithms achieve only 35% to 65% compared to the success rate given by epidemic forwarding algorithm. At this point, we can clearly see how social forwarding suffers in large scale networks. These results match the intuition presented in the EC example and strongly motivate the need for agile solutions that take such situations into account.

IV. Social Forwarding across multiple sub-communities

So far, we have shown that using social inputs in large-scale areas have serious drawbacks. Our main hypothesis is that in large-scale networks where multiple sub-communities may exist, social prediction has its limitations and two people socially connected may not frequently meet because they could be far away from each other. In this section, we first introduce and compare the impact of different large-scale community classification techniques, and ensure that the state-of-the-art social forwarding algorithms perform well within the resulting sub-communities. We then propose CAF, a community aware framework, that can be easily integrated with these algorithms to improve their performance in large scale networks.

A. Classification and Forwarding in Sub-Communities

A common property of social networks is cliques or communities; circles of friends or acquaintances in which every member knows every other member. In large-scale networks, people can be regrouped into sub-communities. Our experimental data set can be classified in multiple communities according to different classification techniques. The San Francisco data set is by default classified into three communities (airport, downtown, and sunset areas) relying on a geographic classification. However in the Dartmouth data set, users can be regrouped according to these two community classifications:



(a) Activity-Based Classification (b) Geographic Classification

Fig. 3: Normalized success rate distribution of PeopleRank relying on different community classification

Geographic Classification: Since the Dartmouth campus area is roughly 1300x1300 square meters, people going to campus every day are mostly visiting the same places. Usually, these places are selected in a way that minimizes the walking distance. To capture this classification, we split the Dartmouth campus into regions (Northwest *NW*, Northeast *NE*, Southwest *SW*, and Southeast *SE*)[2]. A node i belongs to a region R if it has been connected to more access points belonging to this corresponding region as compared to the other regions.

Activity-Based Classification: The Dartmouth College campus has over 160 buildings. Usually people visiting the campus are interested in a few buildings. People can therefore be classified based on their activity or interests. For example, the campus contains more than a dozen athletic facilities and fields. Most of them are located in the southeast corner of the campus. Athletic people are more likely to meet each others and be classified in an athletic community. We consider people more connected to the athletic building's access points as part of the athletic community. Similarly we define academic and residential communities.

After classifying the large-scale community, we now ensure that PeopleRank performs well within a single community; similar results have been obtained with other forwarding algorithms. We plot the normalized success rate of PeopleRank according to the two community classifications described above: (*i*) Activity-Based Classification (in Fig. 3(a)), and (*ii*) Geographic Classification (in Fig. 3(b)).

We observe that the geographic classification leads to better PeopleRank performance. PeopleRank achieves 92% to 97% of normalized success rate within 10-minutes timescales according to the geographic classification (in Fig. 3(b)), and 90% to 94% within the same timescale according to the activity based classification in Fig. 3(a). These results confirm that short distances (*e.g.,* people living in the same neighborhood or region) typically leads to strong social ties, and relevant social classification.

Moreover, we notice that, in Fig. 3(a), PeopleRank achieves higher success rate among athletic users than others according to the activity based classification. Relying on the athletic community, PeopleRank outperforms its own success rate performance by roughly 3% and 5% within 10-minutes timescale compared to respectively the academic and the residential communities. As described above, most of the athletic buildings are located in the southeast corner of the campus which leads to a combination of geographic and activity based classification.

Overall, the results confirm that social forwarding algorithms achieve satisfactory performance within sub-communities. However, it was shown in the previous section that they suffer in large scale networks where multiple sub-communities may exist. We therefore propose a strategy to help existing social forwarding algorithms deal with this issue and successfully forward messages across multiple sub-communities.

B. The Community Aware Framework (CAF)

Motivated by the satisfactory performance of social forwarding within single sub-communities, we propose a *community aware framework* (CAF) that can easily be integrated with most social forwarding algorithms in order to deal with the weaknesses described above in large scale networks. This framework extension relies on the fact that social forwarding algorithms operate normally within the same sub-community. Indeed, messages will be forwarded socially toward nodes which belong to the same sub-community.

On the other hand, particular nodes will operate as an inter-communities backbone and circulate the message to the other sub-communities, and within these sub-communities messages will then be socially forwarded. We call these nodes *Multi-Homed* nodes (MH). MultiHomed nodes are characterized by their higher mobility and belong to multiple sub-communities (*i.e.,* MultiHomed nodes are moving from one sub-community to another). These node can be postmen, buses, cabs, etc. depending on the large community. We then rank the MultiHomed nodes according to the number of sub-communities (MH_{rank}) they belong to. For example, if we consider the geographic classification in the Dartmouth campus data set, MH nodes belonging to four sub-communities are high ranked compared to MH nodes belonging to only two or three sub-communities. Therefore, MH nodes carrying a message forward it to other MH nodes according to a non-decreasing MH_{rank}.

Algorithm 1 summarizes the additional operations (described above) on top of the current state-of-the-art social forwarding algorithm (which will refer to by SFA in Algorithm 1). Besides the simplicity of our proposed algorithm, we would like to emphasize that the overhead is relatively negligible compared to the overhead induced by BubbleRap to compute the global node ranking in the whole system (in a large scale network). Our proposed algorithm remains a distributed forwarding algorithm and relies on local social/contact information to estimate future transfer opportunities. Moreover, this framework extension can easily be integrated with most of the social forwarding algorithms. Next, we evaluate the CAF-extended version of three state-of-the-art social forwarding algorithms that integrate this proposed

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

Algorithm 1 CAF-SFA(node i)

{Node i is running a social forwarding algorithm SFA}
{$SFA(i)$ denotes the rank of node i according to SFA}

Ensure: $MH^{i}_{rank} \leftarrow$ #sub-communities
  while (1) **do**
    while (i is in contact with j) **do**
      $update(SFA(i), SFA(j))$
      while ($\exists\ m \in buffer(i)$) **do**
        if $[(community(i) == community(j))$ AND $(SFA(j) \geq SFA(i))]$ OR $[j = destination(m)]$ OR $[MH^{j}_{rank} \geq MH^{i}_{rank}]$ **then**
          $Forward(m, j)$
        end if
      end while
    end while
  end while

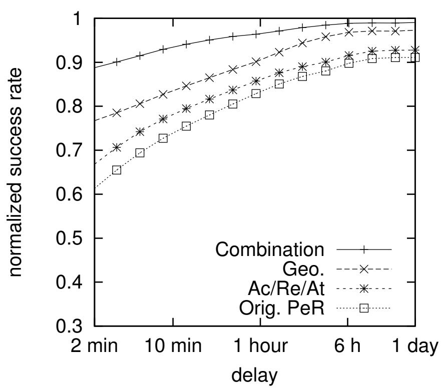


Fig. 4: Impact of community classification on CAF-PeopleRank success rate.

framework.

V. CAF Evaluation

In this section, we apply CAF to PeopleRank, Simbet, and Degree-Based forwarding. We conduct our evaluation using real trace analysis where we compare CAF-extended social forwarding algorithms to BubbleRap algorithm [10], [11].

A. The Impact of Community Classification on CAF-enabled Social Forwarding Algorithms

We investigate the impact of different community classification (described in the previous section) on the performance of CAF. We show a representative set of results for the CAF-PeopleRank algorithm. Similar results have been observed for CAF-Simbet and CAF-Degree Based but are not shown due to space limitation.

Fig. 4 plots the normalized success rate of the extended PeopleRank algorithm, with different classification techniques, in the Dartmouth data set. We notice that extended PeopleRank outperforms the original PeopleRank for all timescales (5% to 30% of success rate improvement). Furthermore, such improvement differs from one community classification to another. Geographic classification gives better performance than activity-based classification; indeed, the activity-based classification in the Dartmouth campus groups people belonging to specific buildings. However, theses buildings are not always geographically close to each others, and therefore messages sent from a specific building can take a long time to reach other members of the same sub-community. We finally note that, combining two community classification approaches leads to better success rate performance, and more than a 30% improvement compared to the Activity-Based classification.

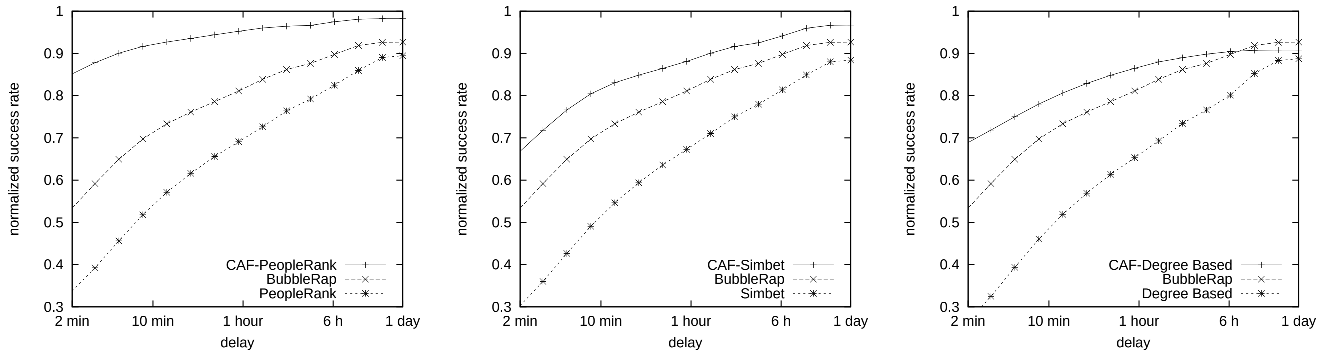
B. CAF vs. BubbleRap

Initially, we consider the San Francisco data set using only 5% of the total cabs to represent the MultiHomed nodes. We evaluate the performance of our proposed framework (CAF) and compare against: (*i*) the original social forwarding scheme (without the framework extension), and (*ii*) the BubbleRap algorithm. Later, we analyze the impact of different fraction of cabs on the performance in order to justify our 5% choice in the evaluation.

We apply CAF to three state-of-the-art forwarding algorithms; PeopleRank, Simbet, and Degree-Based forwarding. Fig. 5 compares the performance of the extended versions of these three algorithms against the original versions (without CAF extension) and the BubbleRap algorithm using the San Francisco data set. We first show that in the three plots, the CAF extended algorithms outperform the corresponding original algorithm for all timescales; for a 10 minutes timescale CAF-PeopleRank outperforms PeopleRank by roughly 40% more delivery success rate, while CAF-Simbet and CAF-Degree-Based achieve respectively 30% and 25% better success rate compared to their original algorithms. For larger timescales CAF performance remain better, however the improvement is less significant since the proposed framework is designed for a better dissemination of the message in order to reach the destination in shorter delays.

Moreover, we show that CAF outperforms BubbleRap especially for PeopleRank and Simbet; the probability to successfully deliver the message using CAF-PeopleRank or CAF-Simbet is 5% to 30% larger than the success rate probability achieved by BubbleRap for all timescales. The reason behind this result is that BubbleRap uses node degree to estimate the social centrality. However, it was shown in [19] that node degree could not be considered to estimate efficiently future contacts. Therefore, BubbleRap performs poorly compared to Centrality-Based forwarding algorithms (CAF-Simbet and PeopleRank) and performs better compared to CAF-Degree-Based especially for large timescales.

We note that BubbleRap uses contact properties to compute communities and uses the same data set to evaluate the algorithm (the same data set is used for both design and evaluation). This may explain the better performance given by this algorithm in large timescales compared to CAF-Degree-Based algorithm. However, in shorter time scales CAF-Degree-Based achieves 7% more success delivery rate than BubbleRap. This can be explained by the use of explicit MultiHomed nodes instead of globally popular nodes as shown in the motivating example (Fig. 1).



(a) CAF-PeopleRank vs BubbleRap (b) CAF-Simbet vs BubbleRap (c) CAF-Degree-Based vs BubbleRap

Fig. 5: Comparison of CAF-PeopleRank, CAF-Degree-based, and CAF-Simbet with BubbleRap (San Francisco data set using only 5% cabs).

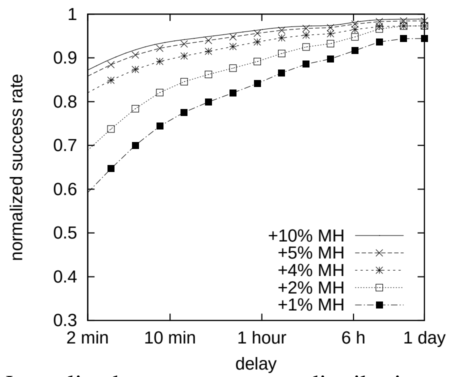


Fig. 6: Normalized success rate distribution of CAF-PeopleRank across multiple communities (San Fransisco).

C. The Impact of MultiHomed Nodes

The number of MultiHomed nodes (MH) needed to make CAF efficient may depend on different factors such as the number of communities, distance between communities, MH mobility, etc. Obviously, the more MultiHomed (MH) available for the system, the more successful it will be. Our goal in this section is to understand the impact of the number of MH nodes on the system's performance. We use the San Fransisco data set and vary the fraction of cabs used in the trace (we randomly pick $x\%$ of the total number of cabs); cabs in this data set are connecting the three disconnected areas of the San Fransisco city, and may operate as MH nodes.

In Fig. 6, we plot the normalized success rate of the CAF-PeopleRank algorithm with different fractions of MultiHomed nodes in the San Francisco data set. Obviously the more MH nodes used the better performance CAF can achieve. We show that the improvement is significant for the first MH added; the improvement from 1% to 2% of MH is roughly 10% of success rate however it is only 0.7% from 5% to 10% of MH nodes. We show that with only 5% of MultiHomed nodes (only 10 cabs in the data set) CAF-PeopleRank algorithm achieves a near to optimal performance; it performs more than 90% of success rate compared to epidemic forwarding (within 10-minutes timescale) which represent only 0.7% less than the optimal given by 10% of MH nodes (20% of MH gives no significant improvement compared to 10% of MH). These results are very promising since they do not require a large number of participants to be involved in order for CAF to be successful, and hence, minimizing the deployment barrier for such solutions.

An important observation from the figure to share is that adding MultiHomed nodes (*e.g.,* taxis in the San Francisco data set) is also beneficial for shorter delays. This might appear strange since one might expect that taxis may need non negligible time to drive from one community to another, and so, only large time delays would be affected. However, this effect can simply be explained by the fact that taxis are also used to improve the performance within a single sub-community; *e.g.,* within downtown, taxis could be considered as a better relay to efficiently disseminate the message within such sub-community.

D. The Cost of CAF

One may claim that CAF can be costly. The cost of a forwarding algorithm, defined as the fraction of contacts involved in the forwarding process, is very important in opportunistic networks. It is obvious that CAF uses more contacts than the original version since it disseminates the message first to all other sub-communities, and then proceeds normally within each sub-community. We therefore quantitatively compare the cost of each social forwarding algorithm and study more options in order to reduce the cost of the extended version of social forwarding algorithms.

In addition to the basic CAF we have presented, we also evaluate the cost of another community destination aware framework (CDAF). Such framework assumes a priori knowledge of the destination's sub-community. Such simplistic assumption, while surrealistic, is widely used in the literature. We use this framework as a benchmark and compare the overhead of CDAFs, CAFs, to the overhead of the original social forwarding algorithms.

We measure the overhead (*i.e.,* cost) of each algorithm as the fraction of contacts (*i.e.,* number of replica copies of the message) used by each forwarding algorithm normalized by the fraction of contacts used by the epidemic forwarding algorithm. Fig. 7 compares the normalized cost of the different schemes. Obviously, CDAF is outperforming all others schemes, and the cost is reduced considerably by roughly 20% to 30%. We also show that CAF uses fewer contacts

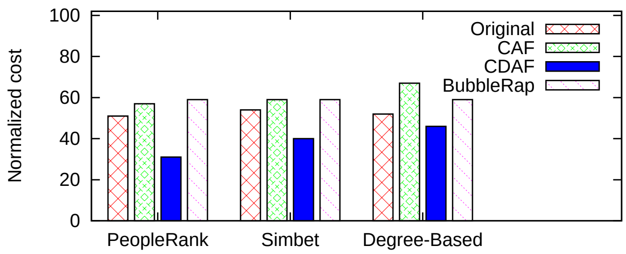


Fig. 7: Normalized cost of different social forwarding schemes using the San Francisco data set (+5% MH).

than BubbleRap in PeopleRank and Simbet cases (by 0.5% to 2%). However, BubbleRap uses 4% less contacts compared to CAF-Degree-Based. This is partly explained by the use of contact properties in both the design and the evaluation of the algorithm which optimizes the number of messages replicated in the system compared to CAF-Degree-Based algorithm. The overall message is that the cost incurred in CAF is negligible when compared to the gains of this framework.

VI. Conclusion and Future Work

The proliferation of online social network platforms and applications such as Facebook, Orkut, or MySpace, makes information about the social interaction of users "easily" accessible. In opportunistic networks, such information can then be used to predict future encounters of participating devices. In this paper, we have studied the weakness of state-of-the-art social forwarding algorithms in large scale networks; in such a network when multiple sub-communities may exist, social prediction has its limitations. We have proposed a community aware framework CAF, which can easily. To summarize our findings, social information can be considered to guide and improve forwarding decisions within a sub-community. Within multiple sub-communities, CAF helps existing social forwarding algorithms to improve their performance by roughly 40%, and achieves 5% to 30% better success delivery rate than BubbleRap, with negligible incurred cost.

There are several venues we plan to purpose in our future work. First, in this paper sub-communities were chosen offline. An important research direction to pursue is to indicate whether these sub-communities can be efficiently identified using a distributed algorithm running with local information at the nodes. Second, we have shown that the cost is considerably reduced if we assuming an a priori knowledge of the destination's sub-community. Better distributed approaches to estimate this knowledge would be highly beneficial.

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] A. Balasubramanian, B. Levine, and A. Venkataramani. Dtn routing as a resource allocation problem. *SIGCOMM Comput. Commun. Rev.*, 37(4):373–384, 2007.

[3] 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.

[4] J. Burgess, B. Gallagher, D. Jensen, and B. Levine. Maxprop: Routing for vehicle-based disruption-tolerant networking. In *Proceedings of IEEE Infocom*, 2006.

[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] V. Erramilli, M. Crovella, A. Chaintreau, and C. Diot. Delegation forwarding. In *MobiHoc '08: Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing*, pages 251–260, New York, NY, USA, 2008. ACM.

[9] 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.

[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] P. Hui, J. Crowcroft, and E. Yoneki. Bubble rap: Social-based forwarding in delay tolerant networks. *IEEE Transactions on Mobile Computing*, 99(PrePrints), 2010.

[12] S. Jain, K. Fall, and R. Patra. Routing in a delay tolerant network. In *Proceedings of ACM SIGCOMM*, 2004.

[13] D. Johnson and A. Lysko. Comparison of manet routing protocols using a scaled indoor wireless grid. *Mob. Netw. Appl.*, 13(1-2):82–96, 2008.

[14] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146, New York, NY, USA, 2003. ACM.

[15] J. Kleinberg. The small-world phenomenon: An algorithmic perspective. In *Proceedings of the ACM International Symposium on the Theory of Computing*, 2000.

[16] J. Kleinberg. Complex networks and decentralized search algorithms. In *Proceedings of the International Congress of Mathematicians (ICM)*, 2006. http://www.cs.cornell.edu/home/kleinber/icm06-swn.pdf.

[17] J. M. Kleinberg. Navigation in a small world. pages 845–845, 2000.

[18] A. Lindgren, A. Doria, and O. Schelén. Probabilistic routing in intermittently connected networks. *SIGMOBILE Mob. Comput. Commun. Rev.*, 7(3):19–20, 2003.

[19] 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.

[20] A. Mtibaa, M. May, and M. Ammar. On the relevance of social information to opportunistic forwarding. *Modeling, Analysis, and Simulation of Computer Systems, International Symposium on*, 0:141–150, 2010.

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

[22] D. Nain, N. Petigara, and H. Balakrishnan. Integrated Routing and Storage for Messaging Applications in Mobile Ad Hoc Networks. In *WiOpt '03: Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*, March 2003.

[23] M. Piorkowski, N. Sarafijanovic-Djukic, and M. Grossglauser. CRAWDAD data set epfl/mobility (v. 2009-02-24). Downloaded from http://crawdad.cs.dartmouth.edu/epfl/mobility, Feb. 2009.

[24] M. M. B. Tariq, M. Ammar, and E. Zegura. Message ferry route design for sparse ad hoc networks with mobile nodes. In *MobiHoc '06: Proceedings of the seventh ACM international symposium on Mobile ad hoc networking and computing*, pages 37–48, New York, NY, USA, 2006. ACM Press.

[25] A. Vahdat and D. Becker. Epidemic routing for partially-connected ad hoc networks. Technical Report CS-2000-06, UCSD, 2000.

[26] W. Zhao, M. Ammar, and E. Zegura. A message ferrying approach for data delivery in sparse mobile ad hoc networks. In *MobiHoc '04: Proceedings of the 5th ACM international symposium on Mobile ad hoc networking and computing*, pages 187–198, New York, NY, USA, 2004. ACM.

[27] W. Zhao, M. Ammar, and E. Zegura. Controlling the mobility of multiple data transport ferries in a delay-tolerant network. In *in IEEE INFOCOM*, 2005.