Social Feature Enhanced Group-based Routing for Wireless Delay Tolerant Networks

Fan Li* Chao Zhang* Zhenmin Gao* Lunan Zhao* Yu Wang†
* School of Computer Science, Beijing Institute of Technology, Beijing, 100081, China.
† Department of Computer Science, University of North Carolina at Charlotte, Charlotte, NC 28223, USA.

Abstract—Mobile devices in delay tolerant networks (DTNs) are used and carried by people, whose behaviors could be described by social models. Understanding social behaviors and characteristics of mobile users can greatly help the routing decision in DTN routing protocols. However, to obtain the stable and accurate social characteristics in dynamic DTNs is very challenging. To achieve efficient delivery of messages at low costs, in this paper, we propose a novel enhanced social group-based routing protocol in which the relay node is selected based on multi-level cross-community social group information. We apply a simple group formation method with both historical encounters (social relationships in physical world) and social features of mobile users (social relationships in social world) and build multi-level cross-community social groups, which summarize the wide range of social relationships among all mobile participants. Our simulations over a real-life data set demonstrate the efficiency and effectiveness of the proposed method by comparing it with several existing DTN routing schemes.

Index Terms—routing, relay selection, social features, multi-level, delay tolerant networks

I. INTRODUCTION

In wireless Delay Tolerant Networks (DTNs) [1], mobile devices are moving around and contact each other opportunistically in cooperative way to deliver data in challenging environments. The possible applications of wireless DTNs include vehicular networks [2]–[4], space communications [5]–[7], pocket switched networks [8]–[10], and mobile social networks [11], [12]. Intermittent connectivity in these DTNs results in lack of instantaneous end-to-end paths, large transmission delays and unstable network topology. Therefore, the classical ad hoc routing protocols, which request the existence of end-to-end paths between the source and the destination during the route discovery and data delivery phases, are not applicable or perform poorly in DTNs.

Many opportunity-based routing protocols [13]–[19] have been proposed for delay tolerant networks to handle intermittent connectivity. Most of these DTN routing methods share the same principle, store and forward, which is described as follows. If there is no connection available at a particular time, the node with the packet can store and carry the packet until it encounters other nodes. When the node has such a forwarding opportunity, it needs to decide which encountered node to be the relay of this packet towards the final destination. Various DTN routing methods adopt different forwarding strategies to select such relay nodes based on certain routing metrics, such as estimated delivery probability, historical contact frequency, available network resources, or estimated delay.

In many delay tolerant networks, especially pocket switched networks or mobile social networks, a multitude of mobile devices are used and carried by people, whose behaviors could be described or modeled by social models. If social characteristics of the mobile users or devices can be obtained or learned, they can be used to improve the DTN forwarding decision. Notice that social relationships and behaviors among users are usually long term characteristics and less volatile than node mobility. This inspires a new type of DTN routing: social-based DTN routing [20]. These social-based approaches [8], [10], [21]–[23] exploit various social characteristics in DTNs (such as community and centrality) to assist the relay selections. For example, nodes with higher social centrality (more popular) are selected as relay nodes (such as in SimBet [21], Bubble Rap [8], and friendship based routing [22]); or nodes within the same community (or social group) with the destination are preferred as relay nodes (such as in Label routing [10], Bubble Rap [8], group based routing [23] and friendship based routing [22]).

Social-based DTN routing methods have been proved more promising than pure opportunity-based routing protocols for certain DTNs since they take advantages of relatively stable characteristics (social properties) to predict and deal with the dynamics of DTNs. However, to obtain the stable and accurate social characteristics in dynamic DTNs is a very challenging task. Most of the existing social-based approaches obtain social metrics (such as community and centrality) from the trajectory and/or the contact history of mobile users. This information is dynamic and usually has to be collected globally through a long-term process. Recently, both [24] and [25] found that certain social features (such as occupation, personal interest, age) can be used to improve the performance of DTN routing if they are available. These social features are relatively static and can be obtained before the deployment of the network. Routing methods based on these social features do not need to maintain any dynamic states except for these static social features. Thus, sometime, they are called stateless protocols.

To achieve efficient delivery of messages at low costs, in this...
paper, we propose an enhanced group-based routing protocol for delay tolerant networks, in which the relay node is selected based on social group information obtained from both historical encounters (social relationships in physical world) and social features of mobile users (social relationships in social world). We adopt a simple but efficient formation method [23] to build multi-level cross-community social groups, which summarizes the wide range of social relationships among all mobile participants. Our group-based routing method forwards the packet greedily toward the destination’s social groups. Simulation results on a real life tracing data [26] demonstrate the efficiency and effectiveness of our social feature enhanced group-based routing method, comparing with several existing methods.

The rest of the paper is organized as follows. Section II briefly reviews the group-based routing and its group formation method from our recent work [23]. Section III provides our detailed design of social feature enhanced group-based routing for wireless delay tolerant networks. Section IV presents our simulation results over a real-life mobile tracing data [26]. Finally, Section V concludes the paper.

II. GROUP-BASED ROUTING

Our social feature enhanced group-based routing is based on a multi-level social graph based routing from our previous work [23]. Social group (or community), an important concept from sociology [27], is usually defined as a group of interacting people living in a common location. It has been shown that a member of a given social group is more likely to interact with another member of the same group than with a randomly chosen member of the population [27], [28]. Therefore, social groups may reflect possible encounter patterns among mobile users in DTNs and our social group-based routing [23] uses social group information obtained from historical encountered data (a contact graph) to make its forwarding decision.

Many methods [29]–[32] can construct social groups or communities from the encountered data. However, most of them are relatively complex. Instead, we adopt a very simple social group formation method based on the number of past encounters among nodes. For any two nodes \( v_i \) and \( v_j \), if there are more than \( t \) encounters between them in the past, they will be placed into the same group. Here \( t \) is an adjustable threshold which defines how strong the social tie between two members is inside a group. Assume that \( G \) is the contact graph including all past encounter relationships among nodes; we only keep an edge between \( v_i \) and \( v_j \) when the number of their encounters is larger than or equal to \( t \). For the graph \( G_t \) formed by all remaining edges, we treat each connected component as one social group. If two nodes are within the same group, there must be a path connecting them in \( G_t \) with all “strong” contact history. Notice that by defining different values of \( t \), we can construct multi-level social groups. Larger \( t \) leads to smaller groups with stronger ties. See Figure 1 for illustration.

Since it is believed that devices within the same social group have higher chances to encounter with each other, our multi-level group-based forwarding method \texttt{mGroup} intends to choose the members of the destination’s social group as the preferred relay nodes. Algorithm 1 shows the details of \texttt{mGroup}. In \texttt{mGroup}, we consider \textit{m}-level social groups \( g_1(), g_2(), \ldots, g_m() \) formed by different thresholds. We assume that \( t_1 > t_2 > \cdots, t_m \), thus the first level group (top level) requires the strongest social tie among its members while the \textit{m} level group has the weakest social tie. The current node \( v_i \) with a message \( M \) destined to \( v_d \) encounters a set of nodes \( R \) which do not hold \( M \). Information of \textit{m}-level social groups \( g_1(), g_2(), \ldots, g_m() \) is available, where \( t_1 > t_2 > \cdots, t_m \).

Algorithm 1 Multi-level-Group-based Forwarding (mGroup)

1. if \( v_d \in R \) then
2. Forward \( M \) to \( v_d \)
3. else
4. \( k = 1 \)
5. while \( k < m \) do
6. if there exists \( v_j \in R \) within the same group at level \( k \) with \( v_d \), i.e., \( \exists v_j \in R \cap g_k(v_d) \) then
7. Let \( v_k \) be the node in \( R \cap g_k(v_d) \) which has contacted \( v_d \) most recently
8. Forward \( M \) to \( v_k \)
9. return
10. else
11. \( k = k + 1 \)
12. end if
13. end while
14. Hold \( M \)

15. end if
destination at this level, mGroup will check with the second level social group. This procedure continues until either it finds a relay node within the same group of the destination or m-level groups are all explored. In the latter case, \(v_i\) continues holding \(M\). In mGroup, multi-level social group information is used to increase the chance of meeting the destination while the FRESH pick tries to deliver the message to destination as soon as possible. If only one-level social group is used, we denote the routing method as Group. In [23], simulation results show that mGroup can achieve better performance than Group, by taking the full advantages of wide ranges of social relationships at all levels.

### III. Social Feature Enhanced Group-based Routing for DTNs

In [23], we build the multi-level social groups purely from the contact graph obtained based on historical encounters, which exploit possible physical contacts between pairs of devices in the physical world. However, the multi-level social group techniques mGroup can be applied to other types of social graphs, if those information are available. In this paper, by introducing social groups defined by social features, we propose an enhanced group-based routing for DTNs. We begin with discussions of possible social features.

#### A. Social Features

Social features of a mobile user could include nationality, affiliation, speaking language, and so on. These social features can represent either physical features (such as gender and height) or logical ones (such as membership in an organization). These information could reflect certain level of social relationships among users in a virtual social world. Using these social features of each individual, it is possible to measure the social similarities between individuals and generate social groups with common social features.

Social features are available and have been used in DTNs for routing guidance [24], [25]. For example, the data set of Infocom 2006 [26] includes answers from each participant to a questionnaire with a number of social information about this person, such as nationality, affiliation and speaking language. In [25], Wu and Wang showed that in this data set the total contact times and contact durations between two individuals reduce when the social feature differences between them increase. The individuals with only one different social feature have about 36.5% more contact times and 32.6% longer contact durations than the individuals with two different features. Similarly, in [24], Mei et al. also found that individuals with similar social features tend to contact more often in DTNs. Therefore, it is also possible to consider the social features or relationships among mobile users to improve the performance of DTN routing. Both [24] and [25] directly use social features as routing metrics, however, this paper considers how to use social features to discover social groups so that social group-based routing can be applied.

One advantage of using social features for routing guidance is that social-feature based routing does not need to collect and maintain routing state information. Social features are static internal features of each mobile node and usually can be obtained before the deployment of the network or during user registration phase.

In this paper, we also use the Infocom 2006 trace data [26], which includes Bluetooth sightings by groups of users (i.e., 79 participants) carrying iMotes for four days during Infocom 2006 conference in Barcelona, Spain. In addition to the Bluetooth contact information among participants, it also has social features of each participant, which are the statistics of participants’ information returned from a questionnaire form. Since some social features in participants’ questionnaire forms are blank, we extract eight social features from the original data set: nationality, graduated school, languages, current affiliation, current position, city of residence, country of residence, and interested topics.

#### TABLE I

**Entropy of social features for Infocom 2006 Dataset**

<table>
<thead>
<tr>
<th>Social Feature</th>
<th>Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graduated School</td>
<td>5.223</td>
</tr>
<tr>
<td>Topics</td>
<td>4.804</td>
</tr>
<tr>
<td>Affiliation</td>
<td>4.015</td>
</tr>
<tr>
<td>City</td>
<td>4.379</td>
</tr>
<tr>
<td>Nationality</td>
<td>4.333</td>
</tr>
<tr>
<td>Country</td>
<td>3.928</td>
</tr>
<tr>
<td>Languages</td>
<td>3.524</td>
</tr>
<tr>
<td>Position</td>
<td>1.390</td>
</tr>
</tbody>
</table>

Among multiple social features, some are more important than others for either routing or group formation purposes. One way to measure the importance of a social feature is to calculate its entropy value over the data set, as did in [25]. In information theory, Shannon entropy is a measure of the uncertainty associated with a random variable. We use entropy to quantify the value of the information contained in the social features. Specifically, there are \(n\) mobile users in the network, and each user has \(N\) social features, denoted as \(f_1, f_2, \ldots, f_N\). For a social feature \(f_j\) with \(M_j\) possible outcomes, \(\{x_i : i = 1, \ldots, M_j\}\), its Shannon entropy, denoted by \(E(f_j)\) is defined as

\[
E(f_j) = -\sum_{i=1}^{M_j} p(x_i) \log_2 p(x_i), \quad j = 1, \ldots, N
\]

where \(p(x_i)\) is the probability of outcome \(x_i\) in the data set. Table I shows the entropy of each social feature listed in the descending order for Infocom 2006 data set. It is clear that a social feature with larger entropy means better distinction among mobile users. But if we use common values of social features to form social groups (i.e. two users are in the same group if they share the same values of certain social features), using only social features with largest entropy may lead to small or isolate groups. This is not good for group-based routing. On the other hand, using social features with lowest entropy may lead to a huge social group, since everyone have the same values. This is also useless for routing purpose. In our
simulations (Section IV), we will study the trade off among different social features.

B. Group-based Routing with Social Features

Motivated from the fact that people come in contact with each other more frequently if they have more social features in common, we apply social features to divide mobile users into different social groups. Recall that in [23] we use different values of contact strength threshold \( t \) to form multi-level social groups based on encounter frequencies in contact graph (as shown in Figure 1). The same approach can be easily adopted to form social groups using the social features.

As illustrated in Figure 2, we can define the social feature strength between any two nodes as the number of common identical social features. For example, if nodes \( v_i \) and \( v_j \) only share the same nationality and affiliation, we give their social strength weight of 2. Large social strength implicitly implies strong social tie. By defining social strength among nodes, we can have a weighted social feature graph \( G' \). Using different value of threshold \( t' \), we can then define multi-level social groups among mobile users. In addition, based on different sets of social features where the common social features are considered, we can have different ways to define various multi-level social groups.

Our multi-level group-based forwarding method \texttt{mGroup} still works on this multi-level social groups. Since people come in contact with each other more frequently if they have more social features in common, we can prefer the nodes in the same social group with the destination as possible relay nodes. If multiple social groups co-exist, the one with strongest social strength (highest in the multi-level structure) can be picked. To distinguish this method from the method using historical encounters, we use \texttt{mGroup-CG} and \texttt{mGroup-SF} to represent group-based routing purely using encounter frequencies from contact graph and the one purely using common social features, respectively. Once again, if only one level social group is used, we denote the method as \texttt{Group-CG} or \texttt{Group-SF}.

C. Social Feature Enhanced Hybrid Group-based Routing

Notice that the social groups based on social features are distinct from the social groups based on historical encounters. The first ones are in the virtual world irrespective of physical distance among mobile users while the second ones are in the physical world which depend on the physical proximity of users. However, both types of social groups are complementary. One of the advantages using social features is that they can be obtained before the deployment of DTNs and there is no need to maintain any states during the routing except the static social graph. On the other hand, the encounter-based social groups reflect the dynamic way people exchange information through direct, face-to-face contacts. Therefore, both types of social groups can be combined and used as a hybrid cross-community multi-level social graph, as shown in Figure 3. We denote the group-based routing over these hybrid social graphs as \texttt{mGroup-H}. This approach actually makes more sense for DTNs, since modern people are living in a cross-space and multi-community co-existed world. Our simulation results in Section IV will verify this conclusion.

IV. SIMULATIONS

In this section we conduct extensive simulations of our proposed method with realistic contact traces [26], \textit{i.e.}, Infocom 2006 trace data, which is publicly available at Crawdad [33]. This data set includes Bluetooth sightings by groups of users (\textit{i.e.}, 79 participants) carrying iMotes for four days during Infocom 2006 conference in Barcelona, Spain. We use all 79 users in our simulations. There are 74,981 contacts between 79 participants over a period of 337,418 seconds. We divide the period using time slot with length of one hour and test the routing performance over randomly chosen 120 time slots. We use the first 40 hours data as historical data to obtain social group information and then evaluate the performance of routing tasks over the remaining 80 hours. For each simulation, we try all possible routing pairs in the network, \textit{i.e.}, each mobile user tries to send messages to other 78 users. The number of message replicas allowed is set to 10 except for epidemic routing. As described in Section III-A, we use eight social features from the original data set, which are summarized in Table I, to generate the social groups used by our group-based routing.

In all experiments, we compare each routing method using the following routing metrics.

![Fig. 2. Multi-level social groups based on common social features.](image1)

![Fig. 3. Multi-level social groups based on different social graphs from heterogeneous sources for the same network.](image2)
A. Single-Level Group-based Routing with Social Features: Grouping with Two Common Features

In the first set of simulations, we test a single-level group-based routing based on social features Group-SF. Here, we set $t' = 2$, i.e., two nodes belong to the same social group if they have identical values of two or more social features. We implement three different versions:

- **Group-SF_1**: All eight social features listed in Table I are considered.
- **Group-SF_2**: Only the top four social features listed in Table I (graduated school, interested topics, current affiliation and city of residence) are considered.
- **Group-SF_3**: Only the middle four social features listed in Table I (current affiliation, city of residence, nationality, and country of residence) are considered.

Figure 4 shows the simulation results. It is clear that Group-SF_1 can achieve best delivery ratio. Thus considering more social features is helpful. Compared with Group-SF_2, Group-SF_3 has better performances though they both consider four social features. This shows that it does not necessarily lead to better performance using social features with higher entropy. Notice that if the entropy of a social feature is very large, it is hard to find common values of that social feature among users.

B. Multi-Level Group-based Routing with Social Features: Grouping with Two Common Features

In the second set of simulations, we test a multi-level group-based routing with social features mGroup-SF. Here, we again set the social strength threshold $t' = 2$. We implement two 3-level mGroup-SF methods:

- **mGroup-SF_1**: The first level considers the top four social features listed in Table I; the second level considers the top six social features; while the third level considers all eight features.
- **mGroup-SF_2**: The first level considers the middle four social features listed in Table I; the second level considers the middle six social features; while the third level considers all eight features.

Figure 5 shows the simulation results. In order to compare the performance of multi-level group-based routing with single-level group-based routing, we also plot Group-SF_1 in Figure 5. Though there is no much difference between mGroup-SF and single-level Group-SF_1 in term of their delivery ratios, it is clear that using multiple levels of social groups can improve the performances (smaller hop counts, delay and number of forwarding). Since mGroup-SF already uses multiple levels of social groups generated from different sets of social features, how to choose among social features becomes not such important.

C. Single/Multi-Level Group-based Routing with Social Features: Grouping with Multiple Common Features

In the third set of simulations, we test our Group-SF and mGroup-SF by changing the values of social strength threshold $t'$. In other words, we now consider different strengths over multiple social features to form the social graphs. We implement the following new methods:

- **Group-SF_4**: We set $t' = 4$ and consider all eight social features. In other words, it requests that two group members share at least 4 common social features.
- **Group-SF_5**: We set $t' = 6$ and consider all eight social features.
- **mGroup-SF_3**: A 3-level social groups where each level has different $t'$. The first level uses social groups from Group-SF_5 ($t' = 6$); the second level considers social
groups from \text{Group-SF\_4} (t' = 4); while the third level considers social groups from \text{Group-SF\_1} (t' = 2).

Figure 6 shows the results. For \text{Group-SF} with single-level social groups, larger threshold $t'$ leads to higher delivery ratio since it provides the information of stronger social ties. However, overall \text{mGroup-SF\_3} with 3-level social groups has the highest delivery ratio. This confirms our original conjecture of better performance with more information.

D. Social Feature Enhanced Hybrid Group-based Routing: Combining Social Features with Contact Graphs

So far we only test our group-based routing with social groups generated based on social features. Via simulations, we can find that the delivery ratio of these methods is still not high compared with some existing DTN routing methods. Therefore, in the last set of simulations, we test our proposed social feature enhanced hybrid group-based routing \text{mGroup-H} which combines both social groups from contact graphs and social groups from social features. We implement the following \text{mGroup-H} and \text{mGroup-CG}.

- \text{mGroup-CG} [23]: It is the group-based routing with a 3-level social group structure built from historical encounters. The threshold values of encounters $t$ are 100, 50, 10 from the upper level to the lower level, respectively. We use this as a baseline method for all \text{mGroup-H} methods.
- \text{mGroup-H\_1}: We use a hybrid 4-level cross-community social groups, where the top three levels are from \text{mGroup-CG} and the last level is from \text{Group-SF\_1} with 2 common social features.
- \text{mGroup-H\_2}: We use a hybrid 4-level cross-community social groups, where the top three levels are from \text{mGroup-CG} and the last level is from \text{Group-SF\_4} with 4 common social features.
- \text{mGroup-H\_3}: We use a hybrid 4-level cross-community social groups, where the top three levels are from \text{mGroup-CG} and the last level is from \text{Group-SF\_5} with 6 common social features.

- \text{mGroup-H\_4}: We use a hybrid 6-level cross-community social groups, where the top three levels are from \text{mGroup-CG} and the last three levels are from \text{mGroup-SF\_3} with $t' = 2, 4, 6$.

We compare our \text{mGroup-H} methods with five other existing opportunity-based routing methods which are listed below.

- \text{Epidemic} [14]: During any encounter, the message is forwarded to all encountered nodes.
- \text{Spray and Wait} [15]: When node $v_i$ has $k > 1$ message replicas and meets node $v_j$, it gives $v_j$ half of its replicas and keeps the other half. Initially, the source have 10 copies of the message. If there are multiple nodes during the encounters, $v_i$ randomly picks one to share its copies.
- \text{FRESH} [17]: The message is only forwarded from node $v_i$ to node $v_j$ if $v_j$ has met the destination more recently than $v_i$ does. If there are multiple nodes satisfying such a condition, $v_i$ forwards the message to the one who has met the destination most recently.
- \text{Destination Frequency} [18]: The message is only forwarded from $v_i$ to $v_j$ if $v_j$ has met the destination more often than $v_i$ does. If there are multiple nodes satisfying such a condition, $v_i$ forwards the message to the one who has met the destination most often.
- \text{Greedy-Total} [19]: The message is only forwarded from $v_i$ to $v_j$ if $v_j$ has more total contacts with all other nodes than $v_i$ does. If there are multiple nodes satisfying such a condition, $v_i$ forwards the message to the one who has most contacts.

Table II shows all simulation results. First, it is clear that our social group based routing method can achieve better performances than most of existing opportunity-based DTN routing methods except for epidemic routing, but the number of message forwarding of \text{mGroup} is much less than epidemic routing and \text{mGroup} is at the similar level in terms of hop count, delay and number of forwarding with those of other opportunity-based methods. Notice that even though epidemic routing has the best delivery ratio, it costs extremely large amount of forwarding. It is also clear that with new additional information from social features our hybrid multi-level social group method \text{mGroup-H} can further improve the delivery ratio while maintaining similar level of other metrics. This confirms the complementary properties of two types of social groups obtained from both physical and virtual worlds and the benefits of combining useful social information cross-space/community.

V. CONCLUSION

Social features of mobile users could reflect certain level of social relationships among users in a virtual social world. By using these social features, we propose a new social-group based DTN routing protocol which forwards message based on multi-level cross-community social group information obtained both from social features and historical encounters. Our simulation results demonstrate the great performance of the proposed method and the advantages of considering diverse
social relationships among nodes during relay selection. We leave exploring more complex social group analysis to achieve further performance improvement as one of our future works.

REFERENCES


