Benefiting From the Community Structure in Opportunistic Forwarding

Bing Bai\textsuperscript{1}, Zhenqian Feng\textsuperscript{1}, Baokang Zhao\textsuperscript{2}, Jinshu Su\textsuperscript{2}

Department of Computer,
National University of Defense Technology,
Changsha, China
\textsuperscript{1}{nudt.bb, fengzhenqian1983}@gmail.com, \textsuperscript{2}{zbk, sjs}@nudt.edu.cn

Abstract. In Delay Tolerant Networks (DTNs), an end-to-end connectivity cannot be assumed for node mobility and lack of infrastructure. Due to the uncertainty in nodal mobility, routing in DTNs becomes a challenging problem. To cope with this, many researchers proposed opportunistic routing algorithms based on some utilities. However, these simple metrics may only capture one facet of the single node's mobility process, which cannot reflect the inherent structure of the networks well. Recently, some researchers introduce the Complex network analysis (CNA) to formulate and predict the future contact in DTNs. The community structure is one of the most important properties of CNA. And it reveals the inherent structure of the complex network. In this paper, we present a community-based single-copy forwarding protocol for DTNs routing, which efficiently utilizes the community structure to improve the forwarding efficiency. Simulation results are presented to support the effectiveness of our scheme.

Keywords: Social Network, Forwarding, Delay Tolerant Network, Community

1. Introduction

In Delay Tolerant Networks (DTNs) [1], an end-to-end connectivity cannot be assumed for node mobility and lack of infrastructure. In such environment, two nodes can transmit messages between each other only when they are in contact (i.e., move within transmission range). Due to the uncertainty in nodal mobility, routing in DTNs becomes a challenging problem. To cope with this, many researchers proposed opportunistic routing algorithms[2][3], in which messages are forwarded between mobile nodes opportunistically upon contacts; and a relay selection is determined separately in each hop, aiming to get higher delivery probability.

To cope with the inherent unpredictability of future contact opportunities, many protocols[4][5] forward multiple copies of the same messages to achieve short latency and high delivery probability. However, many studies[6]
Bing Bai et al.

have shown that node mobility in DTNs is not entirely random, and instead, some patterns can be found more or less. Based on this observation, some researchers proposed utility-based routing protocols[7][33][34], in which messages are forwarded to the nodes with higher probability to deliver it to the destination. In the utility-based routing protocols, many methods of the utility computing have been proposed [8][9][10]. Among them, a number of schemes implicitly utilize the social properties of nodes. For example [11] uses time of last encounter and [12] uses contact frequency to make the prediction of the future of the network, both of which are the social properties called similarity in fact. However, these simple metrics may only capture one facet of the single node’s mobility process, which cannot reflect the inherent structure of the networks well.

Recently, some researchers introduce the Complex network analysis[14] (CNA) to formulate and predict the future contact in DTNs. For example, SimBet[16] uses the combination of nodes’ centrality and similarity as the utility to conduct the forwarding of messages. And BubbleRap[15], defines the node’s social properties as their rankings to determine the forwarding priority of nodes. The community structure is one of the most important properties of CAN. And it reveals the inherent structure of the complex network. In this paper, we present a community-based single-copy forwarding protocol for DTNs routing, which efficiently utilizes the community structure to improve the forwarding efficiency.

To utilize the inherent community structure of DTNs, we decompose the problem into four steps:

1) **Mapping of contacts to social graphs;** In DTNs, nodes usually have the knowledge of their contacts, also called encounter history, which can be depicted by a series of time-node couple. Our work is utilizing these encounter history to generate a social graph, in which the vertexes denote the nodes and the weighted edges represent the encounter history of node pairs.

2) **Detecting the communities on social graphs;** Many studies have been done for the community detecting in social network[28][29][30][31][38]. And the new community detection algorithm in DTNs has also been proposed[35]. It is our future research direction. So it is not discussed in this paper. We use Newman’s community detection algorithm[30] in our simulation. The algorithm is offline and the community information is distributed by the profiles.

3) **Virtualizing social graphs to community graphs;** To make the community structure simple from the node view, we define community graph, in which the vertexes denote the communities and the weighted edges represent the encounter history of community pairs.

4) **Routing with the community information;** Our routing protocol is based on the community graph. So it includes inter- and intra- community routing.

Our approach is based on the weighted network model for DTNs. Although the unweighted network model has been discussed in DTNs[17], we believe that the edges with different numeric can reflect the nodes’ relation better than the ones with only 0 or 1.
The rest of this paper is as follows. Section 2 introduces the related work of current DTNs routing protocols and community detection in weighted networks. Section 3 proposes our community-based routing protocol. Section 4 discusses our simulation method and results. Finally, Section 5 presents conclusions and future research directions.

2. Related Work

In recent years, a lot of routing protocols have been proposed to cope with the challenge environment in DTNs. According to the number of copies of each message that can coexist in the network, current DTNs routing protocols can be classified into two categories: single-copy routing and multi-copy routing.

The single-copy routing schemes keep only one copy of a message in the network. The simplest case of such schemes is that the source node holds the message and forwards it only to its destination. This scheme obviously has minimal overhead, but the delivery delay of a message could be unbounded [19]. Researchers proposed many schemes [2][7][26][27] to forward the message to its destination by intermediate nodes. Usually, the forwarding decisions are made according to the estimation of the node candidates. In [2], four knowledge oracles are defines to represent the amount of knowledge about the network topology. And for different oracles available, the authors present corresponding routing. The lack of this approach is that each node must know the accurate oracle. To overcome this weakness, [7] proposes minimal estimated expected delay (MEED) routing, which computes the expected delay only using the observed contact history instead of the knowledge oracles. Also utilizing the expected delay to make the local forwarding decisions, T. Spyropoulos et al. [20] give analysis of the random walk model and the estimation function of expected delay based on the distance between nodes.

In contrast, the multi-copy routing protocols may generate multiple copies of each message that can be forwarded to increase the message delivery rate. As mentioned before, epidemic routing is the straightforward idea of this case but with huge resource consumptions. Some researchers use history or predication-based approaches to reduce the number of copies spreading in the networks [8][13][21][22]. PROPHET [21] is a probabilistic routing protocol which defines a delivery predictability metric, reflecting the history of node encounters and transitive and time dependent properties of that relation. Another method of multi-copy routing, also called quota-based routing [10], is limiting the number of copies of each message that can be spread in the networks when message is created. Different algorithms have been introduced to efficiently forward the limited copies. Spray and Wait routing [5] is one of the most famous quota-based protocols in which the message holder forwards half copies to the encountered node per contact until one copy left, and delivers the left copy only to the destination. [10] uses the
average encounter times to distribute the copies between nodes. [9] models the message forwarding as an optimal stopping rule problem and proposes a variation of quota-based routing, hop-count-limited forwarding, to maximize the expected delivery rate while satisfying the hop count limited condition. Erasure coding techniques have also been proposed for DTNs routing[23][24]. The basic idea of erasure coding is to encode an original message into a large number of coding blocks. Once sufficiently large subset of the generated code blocks are received, the original message can be successfully decoded.

Recently, researchers have found that in many applications communication devices are taken by human beings, and so conforming to the characteristics of social networks. So several social network metrics, which are measured based on nodes’ direct or indirect observed encounters, are used to guide the packet forwarding in [15][16][25]. SimBet Routing [16] introduces the ego-centric centrality and similarity in social network to guide the DTNs routing. However, these metrics may only capture the single node’s mobility process, which cannot reflect the inherent structure of the whole networks well. LocalCom [37] is a social-based epidemic routing algorithm, in which messages are flooding inter-communities but forwarding intra-communities. In this paper, we propose a community based single-copy routing algorithm in DTNs, which utilizes the community structure to improve the forwarding efficiency.

3. Community-Based Routing (CBR)

As described in section 1, the CBR protocol, which utilizes the inherent social community structure to facilitate packet forwarding in DTNs, has four main steps: mapping of contacts to social graphs, detecting the communities on social graphs, virtualizing social graphs to community graphs, and routing with the community information.

For the community detection is not our contribution, we just introduce three steps of our scheme without the second step (detecting the communities on social graphs).

3.1. Mapping of contacts to Social Graphs

To map the node contact history to a social graph, we first need to determine the meaning of the weight of each edge in the graph. Similar as [9], we use the meeting probability of two nodes as the weight of the edge between them in the graph.

To calculate the meeting probability[9], we use a discrete residual time-to-live Tr for each message, with time-slot size U. Tr is a measurement in clock time. Let T_max be the maximum possible time-to-live of any message, the range of Tr is between 0 and T_max/U. Our delivery probability metric is a
function of \( T_r \), and it is calculated using an inductive method. The amount of computation for our delivery probability metric is inversely proportional to the length of \( U \), but its accuracy decreases as \( U \) increases. In each time-slot \( T_r \), a node can either meet or not meet another node. A node has the probability to meet several other nodes during the same time-slot, and we simply assume that all meetings start at the beginning of some time-slot. This assumption holds when \( U \) is smaller than any meeting duration, and we truncate all meeting durations so that the starting time of them are aligned in the beginning of their respective time-slots. The meeting probability of two nodes in any time-slot of length \( U \) is estimated under the assumption of exponential inter-meeting time by

\[
p_{i,j} = 1 - \exp\left(-\frac{U}{I_{i,j}}\right)
\]

(1)

Where \( U \) is the length of time-slot, and \( I_{i,j} \) is the mean inter-meeting time between node \( i \) and node \( j \). The Newman’s community detection algorithm require the weight of edges is integer, so what we use as the weight of edges is

\[
w_{i,j} = \left\lfloor C \times \ln(p_{i,j}) \right\rfloor.
\]

(2)

where \( C \) is the integralization factor.

Note that the calculation of \( p_{i,j} \) itself does not rely on the assumption of exponential inter-meeting times. Using a particular estimation that is more realistic for a network in question should result in better routing performance.

### 3.2. Virtualizing social graphs to community graphs

For we skip the second step of scheme introduce, so we assume that the community detection has been finished now.

Since the community information has been known by each node, we can simplify the social graphs. Based on the social graphs, whose communities have been detected, we further define the community graph, in which the vertexes denote the communities and the weighted edges represent the encounter history of community pairs. By this way, each node can get very simple view of the whole network, in which only the node in the same community with it and the other communities. Here, the other communities are treated as “big” nodes.
Fig. 1. Virtualizing social graphs to community graphs

Figure 1 gives an example. The left graph is the social graphs, whose communities have been detected. And the right one is the community graph for each node in community 1 (C1).

A. Aggregate the edges between community pairs

In figure1, we can see that in the social graph, there are two edges between C2 and C4, while in the community one, it reduces to one edge. To relief the depression of precision caused by the edge aggregation, we recalculate the weight of the aggregated edges between communities by

$$w_{C_m,C_n} = \left[ C \times \ln(P_{C_m,C_n}) \right].$$

where $C_m, C_n$ are adjacent communities, $C$ is still the integralization factor, and $P_{C_m,C_n}$ is

$$P_{C_m,C_n} = 1 - \prod_{i \in C_m, j \in C_n} (1 - p_{i,j}).$$

where $P_{C_m,C_n}$ represents the probability of any node in $C_m$ meet any node in $C_n$ based on the contact history.

Thus, we have finished the aggregation of the edges between community pairs. But treat communities $C_m$ and $C_n$ as normal node is still not proper. This issue will be disused in next part.

B. Evaluate the intra-closeness of each community

As described above, treat communities $C_m$ and $C_n$ as normal node is not proper, for the intra-community cost has been ignored.

Although many studies has shown that the nodes in same community is “closer” than the outsiders, its cost is still cannot be ignored especially in DTNs. So we have to evaluate the intra-closeness of each community. We use the expectation of the closeness between each node-pair in a community, which is the average longest path length
Benefiting From the Community Structure in Opportunistic Forwarding

$$w_m = \frac{1}{2} \frac{1}{N_m(N_m - 1)} \sum_{i,j \in C_m} d_{i,j}.$$  \hspace{1cm} (5)

where \(N_m\) is the node number in the community \(C_m\), \(d_{i,j}\) is the longest path length between node \(i\) and \(j\) in \(C_m\).

Thus, we have finished the virtualization of social graphs to community graphs, both the inter- and intra- community edges. Based on this, we can design our CBR protocol.

3.3. Routing with the community information

Based on the above work, we propose our community based routing protocol. The key problem of routing is how to select the next relay. Here, we define a Closeness function to help make decision.

As shown in Algorithm 1, each message carries the destination node address \(D_m\) and its community address \(C_m\). When node \(i\) meet node \(j\), for each message \(M_m\) held by node \(i\), calculate the Closeness value \(u\) of \(i\) and \(j\) according to their community address. If \(u_i < u_j\), send \(M_m\) to node \(j\).

For node \(j\), the progress is the same.

Algorithm 1 community-based routing

Let \(N_1, \ldots, N_N\) be nodes
Let \(M_1, \ldots, M_M\) be messages
each message carries the destination node address \(D_m\) and its community address \(C_m\)
On contact between \(N_i\) and node \(N_j\):
for every \(M_m\) held by \(N_i\) do
if \(C_i = C_m\) do \(u_i \leftarrow\) Closeness \((N_i, D_m)\)
else \(u_i \leftarrow\) Closeness \((N_i, C_m)\)
end if
if \(C_j = C_m\) do \(u_j \leftarrow\) Closeness \((N_j, D_m)\)
else \(u_j \leftarrow\) Closeness \((N_j, C_m)\)
end if
if \(u_i < u_j\) do
Send \(M_m\) to \(N_i\)
end if
end for
4. Evaluation

To evaluate our protocol, CBR, we use the Opportunistic Network Environment simulator (ONE) [19], which is a specifically designed simulation tool for delay tolerant networks. Simulation results show that CBR improves delivery ratio by utilizing the inherent community structure in DTNs.

4.1. Simulation setup

We compare CBR against other routing protocols using the dataset, Haggle project [36]. In Haggle project, about fifty devices were distributed to students attending Infocom 2005 student workshop. And the contacts were logged and provided. We divide the original trace files into discrete sequential contact events as the inputs of the simulator. Each contact record includes the start time, end time, and ID of the nodes in contact.

Before the simulation starts, we first map the contacts to a social graph, using the formula 1 and formula 2 mentioned in section 3.1. Here the integralization factor $C$ is 100. And then we use the Newman’s community detection algorithm[30] to divide the communities. At the third step, we use the formula 3,4,5 to virtualize the social graph to community graph. After these work, the information calculated (for example, the intra-closeness of each community) is distributed to each node. And all of this is worked offline.

In our simulations, we primarily focused on two parameters: 1) Delivery ratio: the proportion of packets that arrived at the destination within the delay requirement; 2) Average Delay: although it is not considered so important in DTNs; and 3) Goodput[10]: the number of messages delivered divided by the total number of messages transferred (including those transfers that did not result in a delivery).

For each round of simulation, 1000 messages are created, uniformly sourced between all node pairs. The packet size constant at 25KB, and the buffer space constant at 1MB. Besides, we run each simulation 10 times with different random seeds of events creator for statistical confidence.

4.2. Performance Results

To evaluate our CBR protocol, we compared it with three other famous protocols: Epidemic, Prophet, and LocalCom [37]. Epidemic is the extreme case of multi-copy routing protocols, which is just used as a bound. The Prophet is one of the most famous utility-based routing, representing the traditional routing protocols in DTNs. And LocalCom is also a community-based forwarding algorithm, which is however a flooding one.

In this evaluation, we are trying to answer two questions: 1) Is CBR better than the traditional utility-based routing algorithm in DTNs? 2) Can CBR get
better performance than the other existing community-based routing protocols in DTNs.

![Simulation results](image)

**Fig. 2.** Simulation results

As shown in Figure 2, CBR reaches higher performance than Prophet in both metrics. Also CBR can reach nearly the same delivery ratio as LocalCom, but better goodput than it. The delivery ratio and cost of the Epidemic scheme represent the upper bound in all three cases. Since the simple flooding scheme utilizes all the possible paths over time to forward the packet, if a path that can satisfy the delay requirement exists, it will be included.

In Fig.2(a), the delivery ratio of CBR is very close to the LocalCom during the whole simulation. And CBR can reach totally higher performance than Prophet. The reason for this is that the CBR utilize the community structure which reflects the inherent property of the whole network, while the Prophet only knows the local information of the nodal movement history.

In Fig.2(b), we can see that compared to LocalCom and Prophet, the average delay of CBR is very close. And it is very amazing that when the Expiration TTL of each message is increased from 1 day to 3 days, the average delay of all the four protocols did not change. The reason for this is that some of the messages sent from source nodes are inherently unreachable. For these messages, the increasing of TTL is useless.

In Fig.2(c), the Goodput of CBR is higher than Prophet and LocalCom. The reason of this is that LocalCom is based on community level broadcast and Prophet is multi-copy scheme, while CBR is a single-copy routing scheme. So in the same time, there is only one copy of each messages in networks in CBR. But there may be multiple copies in the other two schemes.

In summation, CBR outperforms the Prophet in terms of delivery ratio, and goodput, while the average delay is close. Although the delivery ratio of CBR is sometimes not better than LocalCom, the gap is acceptable. And the goodput of CBR is clearly outperforms the LocalCom. So CBR can reach better performance than both the traditional utility-based routing algorithms and the other existing community-based routing protocols.
5. Conclusions and future work

Social network properties are observed in many DTNs and tend to be stable over time. In this paper, we seek to utilize the community structure, which is based on social network properties, to improve routing performance. We define the social graph and community graph based on nodes’ encounter history to depict the neighboring relationship between nodes. The simulation based on real mobility traces shows that CBR can get good performance. In the future, we plan to study the multi-copy routing based on the community structure and the community detection in DTNs.

References

Bing Bai et al.

**Bing Bai** received the B.S degree in Computer Science from National University of Defense Technology, Changsha, China, in 2005. He is currently a PhD student in School of Computer at the National University of Defense Technology, Changsha, China. His research interests include Routing in Delay Tolerant Networks and WSN.

**Zhenqian Feng** is an Assistant Professor in the School of Computer Science, National University of Defense Technology. He received his Ph.D. degree in Computer Science from National University of Defense Technology, in 2012. His current research interests include Cloud Computing and Data Centre Networks.

**Baokang Zhao** is an Assistant Professor in the School of Computer Science, National University of Defense Technology. He received his Ph.D. degree in Computer Science from National University of Defense Technology, in 2009. He served as a program committee member for several international conferences and a reviewer for several international journals. He serves on the editor board of Journal of Internet Services and Information Security (JISIS). His current research interests include security and privacy in wireless networks, algorithms and protocols in computer networks, design and optimization in embedded systems. He is a member of the ACM, IEEE and CCF.

**Jinshu Su** received the B.S degree in Mathematics from Nankai University, Tianjin, China, in 1983, the M.S. degree in Computer Science, National University of Defense Technology, Changsha, China, in 1989, and the PhD degree in Computer Science, National University of Defense Technology, Changsha, China, in 1999. He is a full professor at the School of Computer Science, National University of Defense Technology, and serves as head of the Institute of network and information security, NUDT. He is the academic leader of the State Innovative Research Team in University (“Network Technology” Innovative Team) awarded by the Ministry of Education, CHINA. He has lead several national key projects of CHINA, including national 973 projects, 863 projects and NSFC Key projects. His research interests include high performance routers, internet routing, high performance computing, wireless networks and information security. He is a member of the ACM and IEEE.

Received: September 21, 2012; Accepted: March 18, 2013