International Journal of Security and Its Applications
Vol. 11, No. 3 (2017), pp.1-16
http://dx.doi.org/10.14257/ijsia.2017.11.3.01

A Social Popularity Based Probabilistic Routing in Delay Tolerant Network

Jianbo Li*[1], Youmei Song[2] and Shan Jiang[3]

[1,2,3] *College of Computer Science and Technology, Qingdao University, Ningxia Road 308, Qingdao, Shandong, China*
lijianbo@188.com[1], *songyoumei513@163.com*[2], *jiangshan66@163.com*[3]

Abstract

Due to uncertainty in network topology incurred by nodal mobility, designing a routing protocol in Social Delay Tolerant Networks (SDTNs) always faces with challenges. The current research achievements on routing algorithms tend to evaluate the available profit for each prospective relay node separately. In this paper, we not only consider characteristics of each node but also consider characteristics of node's social group. Then a Social Popularity Based Probabilistic routing algorithm (SPPR) is proposed, which defines a metric called social popularity to evaluate node influence in networks. In this paper, message copies are delivered to more popular nodes until they are brought to social group of destinations. Social group is some nodes who interact more frequently with the given node. And then the probability for delivering a message to its destination within time-to-live is computed. We further propose an optimal message schedule policy implemented in destination's social group, by modeling the buffer management problem as 0-1 knapsack. Through the backtracking search, the maximum sum of delivery probability of the problem can be achieved. Extensive simulation results show that SPPR significantly outperforms Epidemic, Prophet and NPSW in general. Especially the message delivery of SPPR is 56.7% higher than Epidemic in Infocom 5 Scenario.

Keywords*: Social Delay Tolerant Networks, routing, popularity, social group, probability*

1. Introduction

In the context of Delay Tolerant Networks (DTNs), the assumptions of traditional networks model, *e.g*. the existence of at least one end-to-end path [1] between source and destination pair, usually fail. DTNs are networks of self-organizing wireless nodes. And they are characterized by lack of contemporaneous and continuous peer-to-peer connections, but only opportunistically intermittent connections. In this case, each message occupies node buffer for a long time. To provide reliable multi-hop message transmission between participating mobile nodes in an unfavorable environment, DTNs adopt the "store-carry-forward" [2-3] message dissemination scheme. That means after being exchanged among two nodes, some messages can be stored and relayed until next opportunistic connection being available. This mechanism makes DTNs to be more promising networks for some applications, e.g., weather service, addressing the successful message transmission while being tolerant of latency. Thus, DTNs have achieved great success in various challenging networks, such as underwater sensor networks [4], vehicular ad hoc networks [5], pocket switched networks [6], and mobile social networks [7].

In recent years, the rapid development of mobile devices such as smartphones, laptops, and tablet PCs makes people contact or share messages with each other much easier. Researchers use the term ''Social Delay Tolerant Networks (SDTNs)''

ISSN: 1738-9976 IJSIA

to describe those consisting of nodes carried by people who move around and communicate with each other in a specific area (e.g., remote village, campus, and local community). Routing algorithm design is still the most challenging problem in SDTNs, even though intermittent connectivity is incurred by mobility of people. Many proposed routings focus on making message delivery more available by replicating and relaying messages round by round until message reaches target node in the SDTNs. However, unacceptable network resource consumption accompanied by high delivery ratio in multi-copy routing is also a tough problem. To solve it, we take social relations among users in networks into consideration, which influence the users' movement pattern and connection opportunity. Mobile device users can further propagate information through their social participation, when their mobile devices are within the transmission range of each other. In this paper, we focus on exploring how to extract and leverage social relationship to enhance routing performance.

Firstly, in real SDTNs [8-9], each mobile device carrier has its own social group. An individual social group is defined as some certain users that are tightly linked to the given user. Members of the group usually share interesting properties, such as common hobbies, social functions, and occupations, with the given user. For example, student *A* interacts more frequently with others in same class. Whereas, student *A* has little chance to connect others who major in different profession. The social relations in node group are more reliable and stable than those in strangers. Secondly, a message should be carried by popular nodes which can guide it closer to destination. Each node plays a different role in transmitting message. For example, stars' behaviors can reach more people than normal people's obviously. Then these star users are considered as the structural importance of roles because they have a stronger capability of connecting other network members. However, there is a misunderstanding that the more connections with others the more popular the node is. Where the connections lead to, which influences the given node's capability in distributing message widely indirectly, is a problem worthy of being explored. Finally, even a message has arrived at the destination social group, there are still chance that message carrier turns away from destination due to random node movement. Then in destination's social group, probability for message being delivered to destination within its time-to-live (TTL) is computed to improve message delivery rate. We further propose an optimal message schedule policy implemented among social group of destination, by modeling the buffer management problem as special 0-1 knapsack. The node with higher probability of meeting destinations implies that it has higher weight value.

In this paper, we propose a heuristic method to measure node influence by exploiting the characteristics of node group, instead of just taking one-hop topology information into consideration as previous routing algorithms proposed. Message is relayed by influential nodes until it arrives at the destination's social group or the most popular node, reducing redundant message copies and other network consumption. In destination's social group, we select next hop node based on the probability of node meeting destination before the message expires. Based on the above, we propose a social popularity based probabilistic routing in DTNs (SPPR). Our main contributions are summarized as follows:

(1) We propose a way to define social group based on inter-contact time between pair of nodes. The inter-contact time is time between the initiations of two successive contacts between pairs of nods. The more frequently members of social group contact the given node, the more reliable and stable the links between them are for distributing message.
(2) Under the definition of social group, we present a metric called social popularity to measure how powerful a node is in network. The value of

popularity is influenced by not only external connections of social group but also the closeness of interior social group. We select relay nodes by evaluating node popularity until message is delivered to destination' social group or the most popular node. In this case, message copies is controlled and propagated efficiently.

(3) In destination's social group, we calculate the probability of node meeting destination. Then we formulate the relay selection problem as a knapsack problem. Through resorting to the back track technique, the maximum delivery probability and making reasonable use of nodal buffer can be achieved.

The rest of this paper is organized as follows. Section 2 introduces previous related work. Section 3 produces the definition of social group and node popularity. Section 4 introduces the method to select the relay nodes according the probability. The routing details are given in Section 5. Section 6 analyses the simulation results compared to other proposed routing protocols. We give the conclusion in Section 7.

2. Related Work

Epidemic routing [10] exploits flooding mechanism, where each of pair of encountering nodes replicates all message it has not carried from the other node. This method enhances the message delivery efficiency at the price of high resource consumption. Thus Epidemic can't work well in resource-limited DTNs. In Prophet [11], a message is always delivered to an intermediate node with higher utility metrics based on encounter probability. The utility function considers both direct encountering probability and indirect relay through another node and is updated upon each encounter and aged over time. Prophet is essentially flooding algorithm. NPSW [12] is a routing protocol that limits the number of message copies. It is mainly evaluated by the number of encountered nodes in a period. In addition, the authors design a utility metric to select next hop nodes with the consideration of the combination of encountering count to destination node and the last contact time. Defining proper number of message copies is a difficult problem in different network environments.

In [13], Wu *et al*. exploit human contact features and propose a hypercube-based multi-path social routing algorithm. There are two processes in this routing: social feature extraction process and multipath routing process. In the first process, they transform the routing problem into hypercube-based feature matching problem before they capture the most informative features by using Shannon Entropy. Secondly, two kinds of multi-path forwarding methods have been proposed for message transmission.

Bubble Rap [14] combines the community with node centrality. Each node has a global rankness and a local rankness, reflecting the popularity of the node in the whole network and its local community, respectively. Messages are bubbled based on the global rankness of nodes until destination communities accept them. Then in local community, the local rankness is used to measure the delivery capability of nodes until meeting final destination. Message is guided to destination from source through node rankness. This avoids message redundancy efficiently.

Xia [25] *et al*. introduce how important to exploit social properties in replica allocation of nodes for ASNETs. They propose ComPAS based on partitioning of social community combined with social relationship and a user level replication. And the number of replicas requires for each node is fixed. This type of replica allocation method improves the availability of different message items in a partitioned social community.

3. Social Popularity

Social characteristics of nodes are of long-term regularities. To obtain reliable and stable connections for message transmission, we identify social group of node firstly. Social group is constructed by aggregating historical contact information and reflects surrounding social network of each node. Then we present a better metric called popularity to evaluate a node's distribution capability based on it. In this section, we select relay nodes with higher popularity to carry message. These popular nodes guide message copies to destination social group or the most popular node in short period, limiting the number of copies and economizing network resources.

3.1. Social Group Definition

A node is expected to have around the same social relations with other nodes. However, each node in real SDTNs can't have close relationships with the rest of all nodes. Due to some social reasons such as common interests, working in common companies or living in common residential communities *et al*., some people a person frequently encountering are usually stable [16]. Thus, we define that some nodes encountering the given node frequently are called regular neighbors represented by set *Nei*, while other nodes rarely encountering the given node are called random neighbors in SDTNs. In this case, the regular neighbors of node i ($i \in N$) and itself constitute its social group. We use G_i to represent it. Node i in its social group is called central node. The social group definition enables more accurate relay node selection through a broader view.

We assume that empirical pairwise inter-contact time distributions for a large portion of node pairs can be well fitted by exponential distributions but with different rates λ. This corresponds with most real traces. Many previous works [17-18] also make a similar assumption. Let us just consider a pair of nodes i and j. The expected delay $E(X_{i,j})$ of nodes i and j can be simply derived from the parameter $\lambda_{i,j}$. That is

$$E\left(X_{i,j}\right)=\frac{1}{\lambda_{i,j}} \tag{1}$$

The smaller expected delay of node i and j is, the more close the relation between them is. Therefore, we can use the parameter $\lambda_{i,j}$ to measure the relationship between node i and j directly. To filter the regular neighbors and prevent the transmission decision from degenerating to random neighbors, we define the social group G_i of node i as follows.

$$G_i = Nei_i \cup \{i\} \tag{2}$$

where

$$Nei_i = \{ j \mid \lambda_{i,j} \geq \delta;\ j \in N,\ j \neq i \} \tag{3}$$

And the link $l_{i,j}$ between node i and j is efficiently only when the relationship satisfies formula (4).

$$l_{i,j} = \begin{cases} 1, \lambda_{i,j} \geq \delta; \\ 0, \text{otherwise} \end{cases} \tag{4}$$

Social group can be developed easily. It only needs to estimate the parameter $\lambda_{i,j}$ based on recorded historical contact information. Setting the value of δ is important. If relationship between friends should be more stable and reliable, δ would be set big. At the same time, some node's social groups may be fall in that case. In order to set appreciate value of δ, we observe and record inter-contact time of each pair of samples and compute their expectation $E(X)$. All time will be given in seconds.

Then we take the reciprocal of $E(X)$ as the value of δ. In this case, the social group will be more stable and reliable. If a new node in network contacts the given node so frequently that satisfies the regular neighbor' selection standard, the local social group is maintained and updated timely, keeping routing decision accurate.

3.2. Popularity Metric

When mobile node fails to access destination through its own connections, it can try to transfer message to more powerful nodes. There are various ties to evaluate social position of a node in network. Centrality is a common term used to measure an importance of a node in networks. And it has many kinds. We will give the definition of degree centrality in the following section.

3.2.1. Degree Centrality

Degree centrality refers to the number of direct connections a node has to its neighbors in the traditional sense. Degree centrality is defined as:

$$D_i = \sum_{j=1,\ i \neq j}^{N} l'_{i,j} \tag{5}$$

where

$$l'_{i,j} = \begin{cases} 1, \text{node } i \text{ contacted node } j \\ 0, \text{otherwise} \end{cases} \tag{6}$$

A node with high degree centrality contacts with numerous other network nodes. Such nodes can be seen as popular nodes with large numbers of neighbors, which may occupy a structural position, acting as a conduit for information exchange. However, the more connections the better might be misleading. Where the direct connections lead to is also influential to the position of a node in the networks. Apparently, node i with popular neighbors is more powerful than node j with isolated neighbors, because node i can affect more other nodes indirectly through one-hop and influent nodes.

3.2.2. Popularity Definition

Degree centrality only considers one-hop contact information. To overcome the drawbacks of degree centrality, we propose a metric called node popularity which extracts surrounding information of social group. We assume that social group can be regarded as a virtual super node. The super node's degree centrality is more reliable than single node's because it takes two-hop topology information into consideration. The super node with high degree centrality means the central node may occupy an importance place in message flow. But the internal relations among social group also have effects on external properties of virtual super node. In other words, a labyrinth of relationships in social group can also influence the popularity of the central node. For example, in Figure 1, node i and node j have own social group and the degree centrality of them is same, that is 3. However, if the link between node i and a fails, node i loses the opportunity to spread message more widely. But if the link of node j and b fails, message can still spread from j to b through c. From Figure 1, we can see that node j is superior to node i in meeting internal dynamic topology and random movement even though they have similar degree centrality. That is because node j has social group with more solidarity. The internal fragility degrades the power of super node i. Then we propose a metric called reliability represented by R_i to evaluate the inside stability of node i's social group.

$$R_i = \frac{2E_i}{n_i(n_i - 1)} \tag{7}$$

where

$$n_i = |Nei_i| \tag{8}$$

$$E_i = \sum_{\substack{k,j \in Nei_i; \\ k \neq j}} l_{k,j} \tag{9}$$

Then we calculate the social group reliability of node *i* and *j* in Figure 1, they are,

$$R_i = 0;\ R_i = 1/3$$

To evaluate the popularity of node *i*, we define node popularity P_i.

$$P_i = R_i \square DC_{G_i} \tag{10}$$

where

$$DC_{G_i} = \sum_{k=1;k \in G_i}^{|G_i|} \sum_{j=1;j \notin G_i}^{|N|} l_{k,j} \tag{11}$$

Then in Figure 1, we can know the popularity of node *i* and *j*, respectively:

$$P_i = 0;\ \ P_j = 1$$

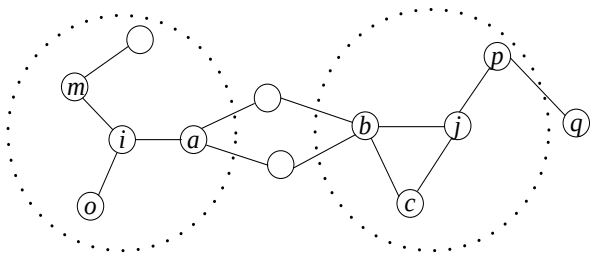


Figure 1. The Example of Nodes' Popularity

Through the calculation, we can know that node *j* is more popular than node *i*, as shown in Figure 1. We rank potential message carriers depending on the social popularity, and try to deliver message via surrounding nodes which are more popular than the given node. The delivery process outside destination's social group fits our intuition and is inspired by real life experiences. We usually deliver package or message via surrounding people more popular than ourselves like postmen. Using the smart popularity prediction scheme, message can be transferred closer to the most popular node or social group of destination quickly.

4. Probability Based Routing

In this section, we assume that the message has arrived at destination's social group by being carried by popular nodes. However, the chance that current message carrier getting further away destination still exists due to random movement of node. What's more, failure to message may be due to two mechanisms: the first is node buffer being in short supply due to unreasonable message distribution. The other is that the delay time outstrips message's TTL. And overmuch dropped messages consume unnecessary network resources, detrimental to improve routing performance. To increase message delivery rate and decrease the number of dropped message, we use probability to describe capability of connecting with destination node directly and indirectly, combined with the interval time of other nodes to destination and message TTL. We further propose an optimal message schedule

policy by modeling the buffer management problem as 0-1 knapsack. Through resorting to the back track technique, the maximum delivery probability sum of the problem can be achieved.

4.1. The Probability Definition

Message is delivered successfully when node inter-contact time is smaller than message TTL. Thus it is importance to select next candidate with high probability of connection with destination within message TTL. We calculate the probability from two perspectives.

Firstly, the direct probability that a message can be successfully delivered is equal to the probability that the next inter-meeting time between the node and the destination is not greater than the sum of message's remaining TTL and the time that has elapsed. Then the direct probability $P_{i,d}^{direct}$ between i and d at time $T^{current}$ is calculated by:

$$P_{i,d}^{direct}(X_{i,d} \le T) = 1 - e^{-\lambda_{i,d}T} \tag{12}$$

where

$$T = TTL_m + (T^{current} - T_{i,d}^{last}) \tag{13}$$

Secondly, if node i frequently encounters node j, and node j frequently encounters node d, then node i can also send message to destination d through node j, *i.e.*,$i \to j \to d$. In this case, node i is more probable to connect node d. Then the probability that i forward message to d via j is represented by $P_{i,d}^{indirect}$. We assume that the probability density function of inter-contact time $X_{i,j}$ between node i and j is $f_{i,j}$. $P_{id}^{indirect}$ can be calculated using the convolution of $f_{i,j}$ and $f_{j,d}$.

$$P_{i,d}^{indirect}(X_{i,j} + X_{j,d} \le T') = \int_0^{T'} f_{i,j}(t) \otimes f_{j,d}(t) dt \tag{14}$$

where

$$T' = TTL_m + (T^{current} - T_{i,j}^{last}) \tag{15}$$

$$f(t) = \begin{cases} \lambda e^{-\lambda t}, t > 0, \\ 0, \text{otherwise} \end{cases} \tag{16}$$

Assuming $Z_{i,d}=X_{i,j}+X_{j,d}$, then

$$\begin{aligned} P_{i,d}^{indirect}(Z_{i,d} \le T') &= \int_0^{T'} dz \int_0^{z} f_{x_{i,j}}(z - x_{j,d}) f_{x_{j,d}}(x_{j,d}) dx_{j,d} \\ &= \int_0^{T'} dz \int_0^{z} \lambda_{i,j} e^{-\lambda_{i,j}(z - x_{j,d})} \lambda_{j,d} e^{-\lambda_{j,d} x_{j,d}} dx_{j,d} \end{aligned} \tag{17}$$

Finally, the sum of direct probability and indirect probability can reflect the real ability of node i connecting destination d:

$$P_{i,d} = \begin{cases} P_{i,d}^{direct} + P_{i,d}^{indirect}, & (P_{i,d}^{direct} + P_{i,d}^{indirect} < 1) \\ 1 & , \text{ else} \end{cases} \tag{18}$$

Thus, even though the current node encounters destination with a low direct connection probability, the connection with a high indirect contact probability between them could still happen if they share common intimate regular neighbors, referred as bridge nodes to connect destination with other nodes.

4.2. The Basic Approach

In social group of destination, relay selection can be done only based on the local knowledge of destination's contacted regular neighbors. We can tell the capability of nodes connecting destination through calculating the probability proposed in last section. Although message carriers can only select nodes with higher probability to

destination as the next hop, it would cause highly bandwidth and delay time in the context. The reason is that there is no explicit optimization objective, which might take away the relay opportunity of those messages that have a little larger size, but great improvement on the delivery probability. For example, the residual buffer size of node i and j are 10MB and 5MB, respectively. There are two messages m_1 whose size is 3MB and m_2 whose size is 4MB cached by node i. The probability set of the node connecting destinations are $P_{i,d}=\{p|p_{i,d1}{}^{m1}=0.6,\ p_{i,d2}{}^{m2}=0.5\}$ and $P_{j,d}=\{p|p_{j,d1}{}^{m1}=0.7,\ p_{j,d2}{}^{m2}=0.9\}$. We assume node i carrying messages m_1 and m_2 encounters node j, it's not rational to deliver all messages to node j even though the sum probability of node j connecting the destination is the most because node j has insufficient buffer. Node i should keep m_1 and forward m_2 to node j for optimization objective. Then the probability of at least one message being delivered is 0.96 with limited buffer size. Thus we formulate relay selection in social group as a special 0-1 knapsack problem, which is well known combinatorial optimization problem. We view each message as the item in the knapsack problem, the probability of node encountering destination before message expires as the value of each item, the size of each message as the weight of each item, the free buffer size as the maximum weight that knapsack can carry. Then the routing problem is equivalent to the special 0-1 problem, and that is formula (19).

$$\begin{cases} \max \ \sum_{i=1}^{M} p_{a,d_1} x_{a,m_i} + p_{b,d_1} x_{b,m_i}, 1 \le i \le M \\ s.t. \quad x_{a,m_i} + x_{b,m_i} = 1 \\ \qquad x_{a,m_i}, x_{b,m_i} \in \{0,1\} \\ \qquad \sum_{i=1}^{M} Size(m_i) x_{a,m_i} \le freeBuffer(a) \\ \qquad \sum_{i=1}^{M} Size(m_i) x_{b,m_i} \le freeBuffer(b) \end{cases} \tag{19}$$

It is of much concern to resolve this special 0-1 knapsack problem. The common way to solve knapsack problem is dynamic programming algorithm. But considering it has the shortcoming of incurring excessive memory consumption in this problem, we use the back track technique to solve the issue. The basic approach of backtracking is depth-first search. For the special 0-1 knapsack problem, it is first transformed into a state space tree. Each full path of the tree represents a possible solution. And backtracking method can greatly enhance the efficiency of the program and avoid unnecessary repeated search by constraint functions. The scheme to solve the optional decision-making problem is stated in Algorithms 1-3. In Algorithm 1, we record and initialize the data information used in Algorithm 2-3 and call the Algorithm 2 to search. Backtrack (i) means search the sub tree at level i. Algorithm 2 generates the space tree dynamically. In algorithm 3, we introduce the bound function to cut off the sub tree which is not the optional, improving operational efficiency. That means we cut off the subtree if its upper bound is less than the current best value.

Table 1. The Initialization of Root Node Algorithm

Algorithm 1 The initialization of root node
Input: $msgList=\{m|m_1,m_2,\ldots,m_n;\ destination(m)\in(G_a\cup G_b)\}$, node a
for each m_i **do**
$w[i]\leftarrow m_i.size$
$p[i]\leftarrow$the delivery probability of m_i
end for
current size: $cw\leftarrow 0$; current probability: $cp\leftarrow 0$; buffer size: $c\leftarrow 0$; the rest buffer size: $r\leftarrow 0$; current optimal value: $bestp\leftarrow 0$;
current path: x[1,2,…,n+1] ←[0,0,…,0]; best path: $bestx[1,2,\ldots,n+1]\leftarrow[0,0,\ldots,0]$;
call Backtrack(1)

Table 2. Backtrack Process

Algorithm 2 Backtrack process
Input: Starting index: i; node a
If $i>n$ **do**
$bestP\leftarrow cp$
for $j\leftarrow 1$ to n **do**
$bestx[j]\leftarrow x[i]$;
end for
end if
$r\leftarrow r-w[i]$;
if $cw+w[i]<=c$ **do**
$cw\leftarrow cw+w[i]$;
$cp\leftarrow cp+p[i]$;
$r\leftarrow r-w[i]$;
$x[i]\leftarrow 1$;
Backtrack(i+1);
$cw\leftarrow cw-w[i]$;
$cp\leftarrow cp-p[i]$;
$x[i]\leftarrow 0$;
$r\leftarrow r+w[i]$;
end if
if Bound(i+1)>$bestp$ **do**
$x[i]\leftarrow 0$;
Backtrack(i+1);
end if

Table 3. Bound Process

Algorithm 3 Bound process
Input: Starting index: i; node a
the rest buffer size: $cleft\leftarrow c-cw$
$b\leftarrow cp$;
while $i<=n\&\&w[i]<=cleft$ **do**
$cleft\leftarrow cleft-w[i]$;
$b\leftarrow b+p[i]$;
$i\leftarrow i+1$;
end while
if $i<=n$ **do**
$b\leftarrow b+p[i]\times cleft/w[i]$
return;
end if;

5. Routing

Generally, more copies generated for a message could achieve higher delivery performance if there were enough network resources, although along with higher delivery overhead. In this paper, we proposed SPPR providing an acceptable balance on network resources cost and delivery rate.

Our routing strategies contain two parts. Firstly, we select next relay node based on node popularity before message arrives at social group of destination. This method can control unnecessary message copies redundancy and spread message widely. Secondly, in destinations' social group, we formulate the relay node selection as knapsack problem based on the probability of meeting destination within message TTL. The given node queries message from peers in their proximity, who not only have higher probability to destination, but also have sufficient buffer size, then forwards the message to them for optimization objection.

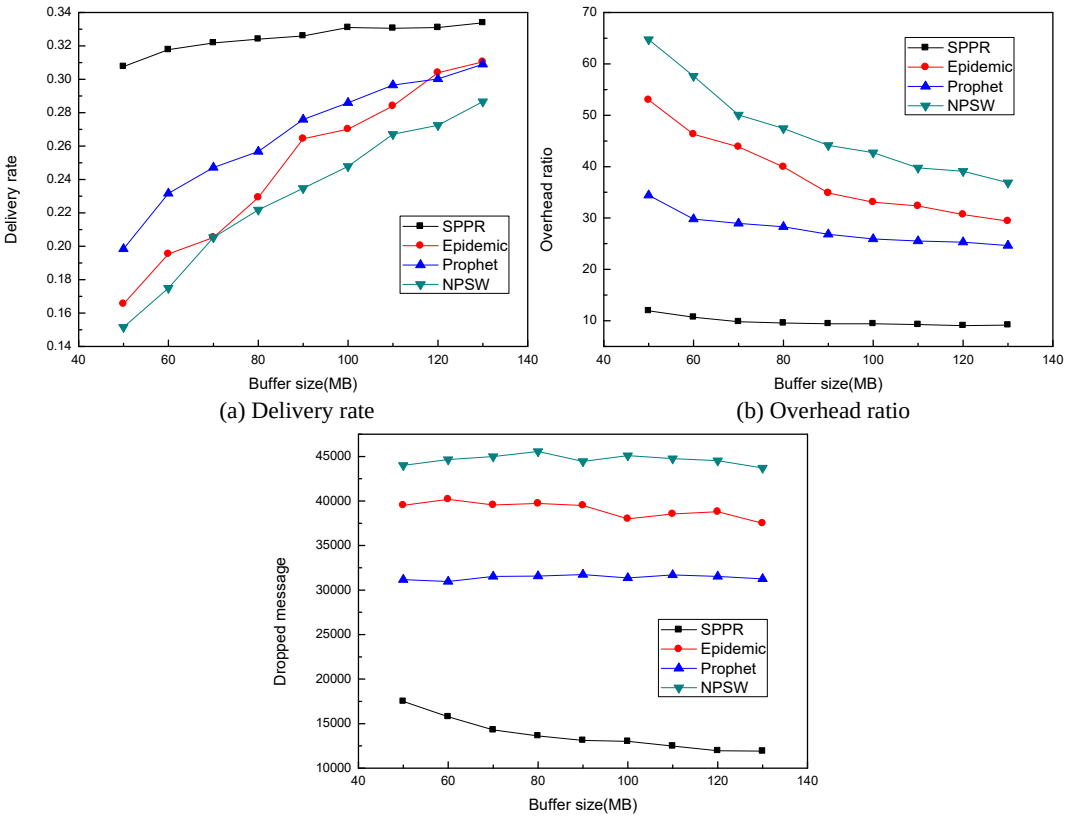
We give the routing protocol details in algorithm 4. This distributed algorithm runs on each node in the network. Each node has own social group G and each message has a list *List* recording the popularity of historical relay nodes, P_1, P_2,…,P_n. The order of objections of *List* statistics by $P_{1'} \leq P_{2'} \leq, \ldots, \leq P_{n'}$.When node i encounters node j, they exchange the neighbor node information firstly. Then they exploit the historical contact message to update own social group G and the popularity P according the regular neighbor selection measure and popularity definition, respectively. At the same time, we detect that whether there are messages in node i's buffer aim to j's social group. If all messages destine for another social groups, we copy messages to node according to hierarchical social popularity. And to elaborate a little further on it, we select node with higher rank compared to $P_{n'}$, until message reaches the most popular node or destination's social group in Algorithm 4 (5)-(6). Then in destination's social group, we require calculation of the probability of node meeting destination, combined with the inter-contact time between pair of nodes and message TTL. The relay selection in destination's social group is formulated as the knapsack problem and we further push message using a backtrack approach to solve it.

Table 4. The Routing Process

Algorithm 4 The routing process
Input: node i, message m, destination d
(1) **for** node $j \in N$
(2) **update** social group G_i;
(3) **update** popularity P_i;
(3) **if** j *is* d
(4) forward m to d;
(5) **if** $j \notin G_i$ && $P_j > P_{n'}$
(6) copy message to j;
(7) **if** $j \in G_d$ && $i \notin G_i$
(8) copy message to j;
(9) **if** $i,j \in G_d$
(10) update the probability $P_{i,d}$, $P_{j,d}$
(11) **Algorithm 2**

6. Evaluation

In order to evaluate the effectiveness of our algorithm, we conduct intensive simulations on the widely-used ONE (Opportunistic Network Environment) simulator [19].We evaluate the Epidemic, Prophet, and NPSW for algorithm performance comparison in terms of message delivery rate, network overhead ratio and number of dropped message. The evaluation is divided into the following categories: (1) varying buffer size in Infocom 5 Scenario; (2) varying interval time in Infocom 5 Scenario; (3)



(a) Delivery rate (b) Overhead ratio

(c) Dropped message

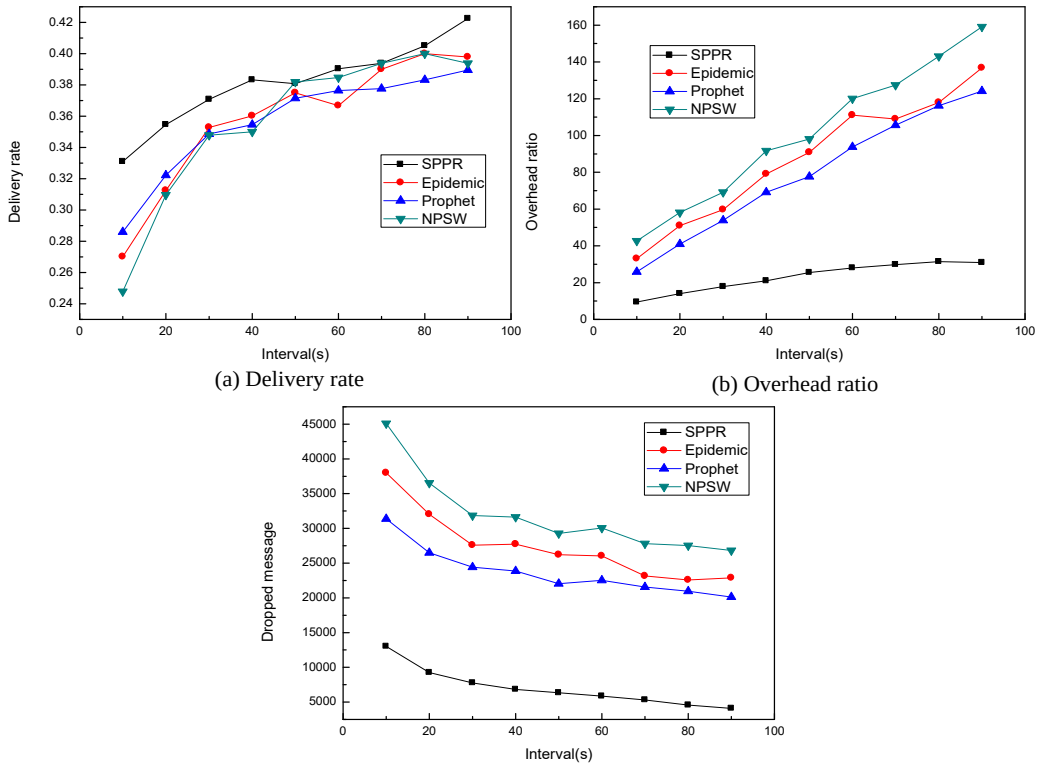
Figure 2. [Infocom 5] Buffer Size versus Delivery Rate, Overhead Ratio, and Dropped Message

varying buffer size in Infocom 6 Scenario; (2) varying interval time in Infocom 6 Scenario. We can see the different performance in different node density situations. The Infocom dataset is one of the most extensive and widely exploited data traces. This trace includes Bluetooth sightings by groups of users carrying small devices in various locations that we expected many people to visit.

6.1. Infocom 5 Scenario

Table 5. Simulation Settings of Infocom 5 Scenario

Parameter name	Range (Default value)
Number of nodes	41
Entire simulation time(h)	12
Word size(m×m)	4500×3400
Message size(KB)	500-1024
Message TTL(h)	10
Buffer size(MB)	50-130(100)
Message interval(s)	10-90(10)
Transmit radius(m)	10

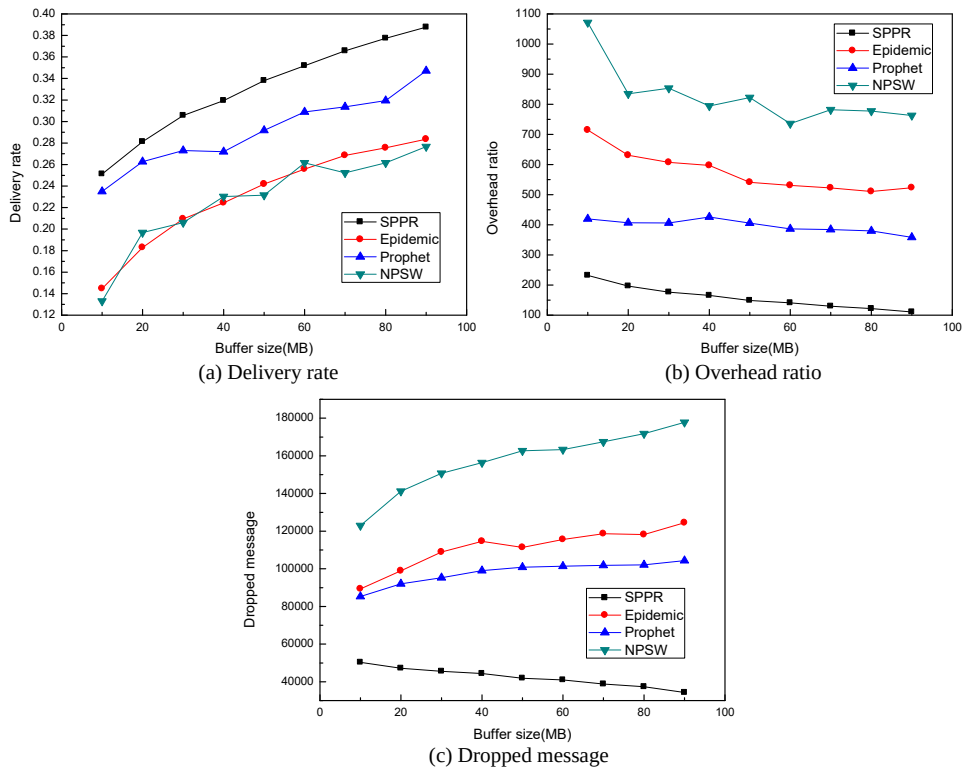


(a) Delivery rate (b) Overhead ratio

(c) Dropped message

Figure 3. [Infocom 5] Message Interval versus Delivery Rate, Overhead Ratio and Dropped Message

The detailed simulation settings are shown in Table 6. Figure 2 illustrates the impacts of changeable buffer size on protocols mentioned previously. To guarantee the reliability of the experimental results, the buffer size is only variable. According to the results in Figure 2, the message delivery rate of Epidemic, Prophet and NPSW takes on ascend trend and the overhead ratio of them declines because the tension between node buffer and message is abated. And the delivery rate of SPPR keeps high and changes smoothly with the buffer size increasing. When node buffer is limited, SPPR still achieves lower overhead ratio and number of dropped message. When buffer size increases to 70MB, the delivery rate of our proposed algorithm is 56.7.0% higher and the overhead ratio is 77.5% lower approximately than Epidemic in Figure 2(a), (b). SPPR depends on node popularity to spread message until the message is carried to destination's social group. This routing strategy controls the number of message copies and unnecessary network resource consumption. At the same time, SPPR heuristically estimates the probability to destination before message expires directly and indirectly. This strategy makes message have more chance to be delivered successfully. Thus SPPR not only improves message delivery, but also consumes lower network resources by limited buffer size in Figure 2.



(a) Delivery rate (b) Overhead ratio

(c) Dropped message

Figure 4. [Infocom 6] Buffer Size versus Delivery Rate, Overhead Ratio and Dropped Message

In the simulation of varying message interval time, we set the node buffer size to be 100 MB. The shorter interval time of creating each new message is, the more messages there are in the network. Thus reducing message redundancy and avoiding unnecessary network consumption is of great importance for algorithm due to the restrictions of node buffer. From the results in Figure 3, we can know that even though buffer resource is highly constrained and message generates fast, SPPR is still a considerable choice for routing. NPSW, the limited message copies transmit in parallel, is still inferior to our proposed algorithm in delivery rate even though the interval increases. What's more, compared to Epidemic, Prophet and NPSW, SPPR has lower overhead rate.

6.2. Infocom 6 Scenario

Besides the Infocom 5 Scenario, we also operate along with other type of network, Infocom 6. The settings of this simulation are listed in Table 7. Figure 4(a) shows that the message delivery rate of our proposed algorithm is higher than other three algorithms even the node buffer size is restricted. At the same time, SPPR still keeps lower than them in the overhead ratio and the number of dropped message. Though flooding strategy is assurance of high delivery rate, Prophet and Epidemic also bring high network expenditure as shown in Figure 4(b), (c). NPSW limits the number of message copies, but it doesn't work well in overhead ratio and dropped message. The method to select relay nodes in SPPR is more efficient than message distribution based proposed unity function in NPSW.

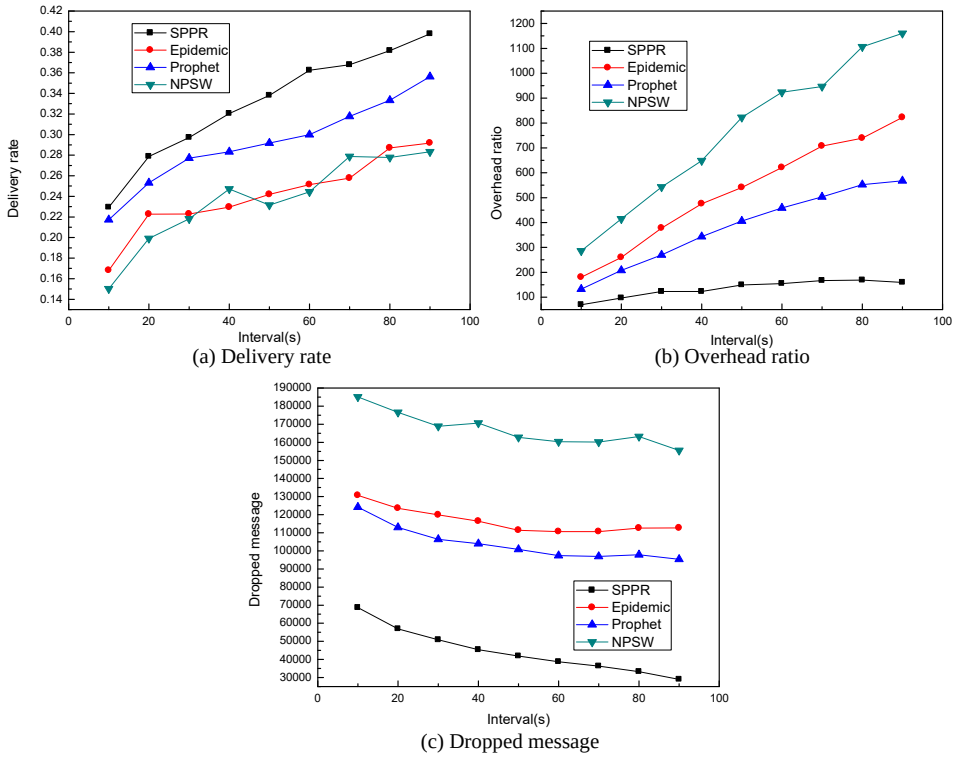


Figure 5. [Infocom 6] Message Interval versus Delivery Rate, Overhead Ratio and Dropped Message

In the simulation results shown by Figure 5, we set the buffer size to be 50MB. With the increase of pre-assigned message interval value, the delivery rate of algorithms is almost in rising trend in Figure 5(a). In Figure 5(b), (c), the overhead ratio and dropped message of SPPR is much lower than other algorithms. Furthermore, the overhead ratio of Epidemic, Prophet and NPSW grows rapidly with the increment of message interval, while our proposed routing changes smoothly in Figure 5(b).

In summary, our proposed SPPR has better performances in different node density situations compared to other routing algorithms. SPPR is a proposal that strikes a balance between message delivery rate and network resource consumption. And the problem caused by extensive message dissemination and limited node buffer can be solved by selecting relay node based on proposed node popularity. We can conclude that our adaptive strategies can make transmission operation more efficiently and useful.

Table 6. Simulation Settings of Infocom 6 Scenario

Parameter name	Range(Default value)
Number of nodes	77
Entire simulation time(h)	12
Word size(m×m)	4500×3400
Message size(KB)	500-1024
Message TTL(h)	10
Buffer size(MB)	10-90(50)
Message interval(s)	10-90(50)
Transmit radius(m)	10

7. Conclusion

We have proposed SPPR based on social popularity and probability to improve effectiveness of data dissemination. By collecting connection information, we update the social group of each node dynamically, which only contains a node's major relationships. Then we have defined a heuristic function to estimate the node popularity which utilizes social group properties. Besides, we have formulated the relay nodes selection as a special knapsack problem. Node with higher probability to destination means that it has higher value. Extensive trace-driven simulations have shown that our approach SPPR outperforms Epidemic, Prophet and NPSW in the overall performance. And SPPR has better performance than other three algorithms. We will further study social characteristics of social group to deal with the changeable network topology.

Acknowledgements

This research is supported in part by National Natural Science Foundation of China under Grant No.61502261, 61572457, 61379132, Natural Science Foundation of Shandong Province under Grant No.ZR2013FQ022 and Science and Technology Plan Project for Colleges and Universities of Shandong Province under Grant No.J14LN85.

References

[1] Y. Cao and Z. Sun, "Routing in delay/disruption tolerant networks: a taxonomy, survey and challenges", IEEE Communications Society, vol. 15, no. 2, **(2013)**.
[2] K. Fall and S. Farrell, "DTN: an architectural retrospective", IEEE Journal on Selected Areas in Communications, vol. 26, no. 5, **(2008)**.
[3] X. Jia, X. Feng and R. C. Wang, "Adaptive spray routing for opportunistic networks", International Journal on Smart Sensing and Intelligent Systems, vol. 6, no. 1, **(2013)**.
[4] P. Raghu Vamsi and K. Kant, "Trust aware data aggregation and intrusion detection system for wireless sensor networks", International Journal on Smart Sensing and Intelligent Systems, vol. 9, no. 2, **(2016)**.
[5] Z. Yu, F. Liu and H. Jie, "AMS-PSO method for jamming unmanned aerial vehicle network", International Journal on Smart Sensing and Intelligent Systems, vol. 8, no. 4, **(2015)**.
[6] V. Erramilli, A. Chaintreau, M. Crovella and C. Diot, "Diversity of forwarding paths in pocket switched networks", Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement(IMC), **(2007)**; California, USA.
[7] N. Vastardis and K. Yang, "Mobile social networks: architectures, social properties, and key research challenges", IEEE Communications Surveys and Tutorials, vol. 15, no. 3, **(2013)**.
[8] T. Spyropoulos, K. Psounis and C. S. Raghavendra, "Spray and wait: an efficient routing scheme for intermittently connected mobile networks", Proceedings of the ACM SIGCOMM Workshop on Delay-Tolerant Networking (WDTN), **(2005)**; Pennsylvania, USA.
[9] Y. Wang, WS. Yang and J Wang, "Analysis of a Hypercube-Based Social Feature Multipath Routing in Delay Tolerant Networks", IEEE Transactions on Parallel and Distributions System, vol. 24, no. 9, **(2013)**.
[10] A. Vahdat and D. Becker, "Epidemic Routing for Partially Connected Ad Hoc Networks", Duke University Technical Report CS-2000-06, **(2000)**.
[11] A. Lindgren, A. Doria and O. Schelen, "Probabilistic routing in intermittently connected networks", Mobile Computing and Communications Review, vol. 7, no. 3 **(2003)**.
[12] C. Q. Dai, J. B. Li and L. You, "An improved spray and wait routing algorithm based on node performance in delay tolerant networks", Information Technology Journal, vol. 12, no. 19, **(2013)**.
[13] J. Wu and Y. Wang, "Hypercube-based multi-path social feature routing in human contact networks", IEEE Transactions on Computers, vol. 63, no. 2, **(2014)**.
[14] P. Hui, J. Crowcroft and E. Yoneki, "BUBBLE Rap: social-based forwarding in delay-tolerant networks", IEEE Transactions on Mobile Computing, vol. 10, no. 11, **(2011)**.
[15] X. Feng, AM. Ahmed and L. T. Yang, "Exploiting social relationship to enable efficient replica allocation in Ad-hoc social networks", IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 12, **(2014)**
[16] X. Feng, L. Li and J. Li, "Social Aware Networking: A Survey", IEEE System Journal, vol. 9, no. 3, **(2013)**.

[17] M. J. Xiao, J. Wu and L. S. Huang, "Community-aware opportunistic routing in mobile social networks", IEEE Transactions on Computers, vol. 63, no. 7, **(2014)**.
[18] W. Gao, Q. Li and B. Zhao, "Multicasting in delay tolerant networks: A social network perspective", Proceedings of ACM MobiHoc, **(2009)**; LA, USA.
[19] A. Keränen, J. Ott and T. Kärkkäinen, "The ONE simulator for DTN protocol evaluation", Proceedings of the 2nd International ICST Conference on Simulation Tools and Techniques, **(2009)**; Rome, Italy.

Authors

Jianbo Li, he received the BS and MS degrees in computer science from Qingdao University, China in 2002 and 2005, respectively, and the PhD degree in computer software and theory from University of Science and Technology of China in 2009. He is currently a professor in Computer Science&Technology College of Qingdao University. His research interests include distributed and parallel computing, delay tolerant networks and algorithm analysis.

Youmei Song, she received the BS degree in software engineering from Shenyanglidong University in 2014, and is currently working toward the MS degree in College of Computer Science and Technology, Qingdao University. Her research interests include delay/disruption tolerant networks, with an emphasis on routing protocols, algorithms and social networks.

Shan Jiang, she received the B.S. degree in department of information management and information systems engineering and the M.E. degree in information management from Shandong Institute of Business and Technology and Nankai University in 2004 and 2007, respectively. She is a Lecturer in Computer Science&Technolgy College of Qingdao University. Her research interests include delay/disruption tolerant networks, with an emphasis on routing protocols and data offloading.

