Routing Scheme Based on Community Correlation in Socially Aware Networking

QIONG XU1,2, MIN DENG1, FANG XU1,2, XIAOCHAO ZHAO1, ZHE LI1, ZENGGANG XIONG1,2, NAN XIAO1,2, GUOTAO ZHAO1, AND ZHEN LIU1

1School of Computer and Information Science, Hubei Engineering University, Xiaogan 432000, China
2School of Computer Science and Information Engineering, Hubei University, Wuhan 430062, China

Corresponding authors: Min Deng (dm2010@whu.edu.cn) and Fang Xu (xf2012@whu.edu.cn)

This work was supported in part by the Ministry of Education in China (MOE) Project of Humanities and Social Sciences under Grant 20YJAZH112, in part by the National Natural Science Foundation of China under Grant 61972136, and in part by the Hubei Provincial Department of Education Outstanding Youth Scientific Innovation Team Support Foundation under Grant T20200017 and Grant T201410.

ABSTRACT Since most existing routings in socially aware networking only perform well in a certain statistic of network performance (e.g., Epidemic has high delivery rate but high network overhead, Direct Delivery has low network overhead but poor performance in latency and delivery rate), a new routing scheme: Community Correlation Based Routing (CCBR) is proposed in this paper to balance the delivery rate and network overhead. In CCBR, message forwarding is divided into two phases: in-community and across-community. We first define three new indicators by analyzing the social attributes of nodes: the social relationship between nodes, the active of node and the community correlation between communities. Then we use the three new indicators to form two functions: in-community forwarding utility and across-community forwarding utility. When messages are in the stage of in-community forwarding, we select the node with higher in-community forwarding utility as the relay node to make the messages deliver to the destination node faster. And when messages are in the across-community forwarding stage, the relay node with higher across-community forwarding utility is selected to avoid the messages being confined to the local community and make sure that the messages can always be transmitted to the destination community. Extensive simulations show that the proposed CCBR routing scheme can effectively improve the message delivery rate and greatly reduce the network overhead. And it can achieve better performance compared with the existing routing schemes even under the limited cache.

INDEX TERMS Socially aware networking, routing scheme, social attributes, community correlation.

I. INTRODUCTION

With the rapid development of wireless network technology and improvement of the performance of mobile devices, people use mobile smart phones more frequently in their daily life than desktop computers or laptops. According to statistics, by March 2020, the number of Chinese netizens has reached 904 million, among which the number of mobile Internet users has reached 897 million, accounting for 99.3% of the total number of Internet users [1]. The increasing popularity of mobile devices and the growing number of mobile traffic, as well as the continuous integration of mobile communication network and social network theory, socially aware networking (SAN) came into being. We use the concept of SAN to represent a new model that takes the social connections or characteristics between mobile nodes as the main basis for network communication design [2]. It can sense and take advantage of the context information of network nodes. And the mobile network is becoming more and more intelligent through the use of context awareness computing [3].

SAN is an evolution of delay tolerant networks (DTNs). It works similarly to DTNs, they all lack end-to-end path between the source node and the destination node. They generally utilize meeting opportunities to implement transmission in multi-hop mode, in which messages are transmitted through a “store-carry-forward” strategy. Different from DTNs, the carriers of smart devices in SAN are mostly composed of humans who have their own social attributes (such as gender, age, education level, etc.). And there are also some relationships among humans (such as relatives, friends, classmates, colleagues, etc.). Correct analysis and use of this social relationship to assist routing design can greatly improve routing performance. On the contrary, in any kind of network, if the wrong node is identified as the influential node, the network may fail [4]. Therefore, analyzing the
social relations of nodes has become an important factor in
designing routing protocols in SAN.
Research shows that community structure can improve for-
warding efficiency [5]. In community-based routings, mes-
sages are delivered by sending them first to the destination
community and then to the destination node [6]. It is more
beneficial to forward the message to the community where
the destination node is located if the community with higher
connection with the community of the destination node is
chosen as the relay community. In addition, social relations
between nodes affect the encounter probability to a certain
extent. It indicates that nodes with higher social relations tend
to encounter each other frequently [7]. And the more active
the node, the more likely it is to establish connections with
other nodes. Based on this, three new metrics are defined in
this paper which we designate as social relationships between
nodes, the activity of node and community correlation. Those
three metrics are from two perspectives namely nodes and
communities, and the calculations are also given. Finally,
a routing algorithm is designed by combining the new metrics
mentioned above. The routing first divides the nodes in the
network into communities. If the destination node is in the
same community with the source node, the node with stronger
social relationship with the destination node and higher activ-
ity is selected as the relay node. If the message needs to
be forwarded across the community, the node that is more
closely related to the destination node and whose community
has a higher correlation with the community of destination
node is selected as the relay node to improve the performance
of the routing.

The structure of this paper is organized as follows: Section II introduces the related work. Section III introduces
the system model used in this paper and defines the new
metrics and gives the calculation formulas. Section IV intro-
duces the routing scheme based on community correlation.
Section V simulates and evaluates the routing scheme, and
section VI summarizes the whole paper.

II. RELATED WORK
So far, many experts and scholars have done a lot of
researches on routing performance. They have put forward
various routing algorithms to improve the message delivery
ratio, reduce delay and save the network overhead. Some
classical routing algorithms in the early such as Epidemic
[8] based on flood idea, Direct Delivery [9] and Spray
and Wait [10] based on copy control, Prophet [11] based
on historical information and SimBet [12], SimBetTs [13]
and Bubble Rap [14] based on simple social information,
etc., which are the research foundations of many scholars.
On this basis, later researchers paid more attention to the use
of social attributes of nodes. Studying the social attributes
and mobile characteristics of nodes has become the main-
stream trend of current routing research. At present, the
popular social attributes mainly include social graph, com-
munity, centrality, similarity, tie strength, human mobility
pattern, etc. [15]. And most routing schemes that based on
social attributions use centrality or similarity to facilitate
messageing [16].
For example, literature [17] introduces the concept of
home-aware community, defines the centrality in the com-
munity and the betweenness centrality between communities
to measure the importance of nodes in forwarding messages
in or across communities. Reference [18] measures the social
similarity through nodes’ different levels of local activity,
so that messages will not be forwarded to nodes with low
activity and thus improve the efficiency of message for-
warding. Reference [19] proposes a metric for detecting the
quality of the relationships between nodes by considering
the contact time, contact frequency and contact regularity.
The simulation results show that the routing achieves better
delivery rate than SimBet and Bubble Rap, the classic routing
protocols, without affecting the average delay. Reference [20]
fully utilizes the sociality and mobility of nodes. For the
message forwarding in the same community, the routing uses
the activity of the node as the evaluation index, and select the
node with the higher frequency of the destination node to for-
ward. For the message forwarding in different communities,
it first uses the community correlation as an evaluation index,
and selects the nodes that have a higher correlation with
the destination node’s community. Besides, [21] uses node’s
social properties to calculate the social similarity utility, and
uses the social connection of network nodes to calculate the
betweenness centrality utility of node. The forwarding metric
is combined with two utility functions to derive the social
strength among users and their importance, and it is used
to determine the best relay node. Moreover, [22] used BP
eural network to predict the encounter regularity of mobile
nodes in terms of time and space dimensions. Simulation
analysis and experimental results show that the proposed
routing algorithm can effectively improve the message deliv-
eration ratio and reduce the network overhead. Literature [23]
divides the nodes into communities according to the social
characteristic attributes of them. When a node forwards a
message, it preferentially selects a node that is in the same
community as the destination node. For nodes that are not in
the same community, the utility function of the node feature
value is calculated by comprehensively considering the node
intimacy, social correlation and social activity. The node with
a larger value is preferentially forwarded. In [24], through
studying the effect of variety of context information on the
mobility patterns in this mobile social networks, using three
dimension contexts, which are physical adjacency, social
similarity and social interactivity to make routing decisions
dynamically.

Although many experts and scholars have designed and
studied lots of routing algorithms in SAN, and have obtained
certain achievements. However, many of them have failed
to fully explore the different roles played by different social
attributes of nodes in different forwarding stages. And there
are more or less defects in the measurement of social
attribution-related indicators, thus cannot accurately reflect
the relationship between nodes. If we can make full use of
the social properties of nodes and communities, and measure reasonable indexes, routing performance will be improved greatly.

III. SYSTEM MODEL
The sociality of nodes in a SAN makes nodes with the same interest more likely to visit the same area. This social structure has been called community by Newman [25]. The community structure divides the location relationship between source node and destination node into two situations: in the same community and in different communities. Therefore, message forwarding is also divided into two stages: in-community forwarding and across-community forwarding. In the stage of in-community forwarding, the source node and the destination node are in the same community, the message only needs to be forwarded within the community. At this time, select the node with higher intimacy with the destination node and stronger forwarding ability within the community as the relay node, which can improve the message delivery rate. And in the stage of across-community forwarding, the source node and the destination node are in different communities, a node that has a closer relationship with the destination node and a greater correlation between its community and the target community will be chosen as a relay node. It can help to avoid the message being confined to the local community and enable the message to be delivered to the direction of the destination node. Nodes in the same community encounter each other more frequently. So forwarding the message to the community where the destination node is located is more conducive to the successful delivery of the message.

Therefore, we define three new metrics: “social relationships between nodes”, “activity of node” and “community correlation”. In the stage of in-community forwarding, the relay ability of node is evaluated based on the social relationship between node and destination node as well as the activity of the node. And in the stage of across-community forwarding, the social relationship and the correlation between the communities of node and destination node are taken as the basis for the selection of relay nodes.

A. MODEL AND ASSUMPTIONS
For ease of description, we use symbols to describe the information (as shown in table 1).

At the same time, the following assumptions were set:
1) The connections between nodes in a socially aware networking are abstracted as graph \( G(V, E) \), in which \( V \) is the node set, \( E \) is the connection set. The number of nodes in the network is denoted by \( N \). When nodes encounter each other, connections are established. All useful information of nodes’ connections in the network is recorded, such as node pairs, serial number of community that the node belongs to, time of the connection establishment and disconnection, etc. \( E(i, j) \) denotes the connection between node \( i \) and node \( j \), \( E(i, j) \) and \( E(j, i) \) are consistent, and \( i \neq j \). If two nodes are connected once, then \( E(i, j) \) add one.
2) The community is abstracted as graph \( G'(C, E') \), where \( C \) is the community set and \( E' \) is the connection between communities. Assumption that node \( i \) belongs to the community \( C_i \), node \( j \) belongs to the community \( C_j \), and \( C_i \neq C_j \). The connection between any node \( i \) and any node \( j \) is the connection between community \( C_i \) and community \( C_j \). We mark the connection as \( E'(C_i, C_j) \). Every time the nodes between the two communities are interconnected, \( E'(C_i, C_j) \) adds one.
3) Each node belongs to at least one community, and there may be overlapping areas between multiple communities. \( C(i) \) denotes the community set containing node \( i \), the community number set containing node \( i \) marks as \( l_i \), and \( C(i) = \{C_i(i)/ l_i\} \). Each node knows the community set it belongs to in advance. So the source node can detect if the destination node is in the same community or otherwise.
4) The nodes in the network are always valid and in a mobile state and unconditionally provide forwarding services to other nodes.

B. SOCIAL RELATIONSHIP MODEL
The social relationship between users can be described in a variety of ways. For example, the frequency, duration or interval of encounters between nodes, or the extent of matching of social attributes. But a single dimension cannot describe it accurately. It is widely known that two nodes have better social relations if they are in frequent contact with each other and have more mutual friends. Therefore, this paper chooses to combine the frequency of encounters between nodes and the proportion of common neighbors as the evaluation index of social relations between nodes.

\[
\text{Definition 1: Encounter Frequency} \ f(i, j) \ \text{It refers to the proportion of the number of encounters between node } i \ \text{and node } j \ \text{in the total number of encounters between node } i \ \text{and any other node in the network, that is}
\]

\[
f(i, j) = \frac{E(i, j)}{\sum_{j=1}^{N} E(i, j)} \quad (1)
\]
Definition 2: Common Neighbor Ratio \( n(i, j) \). It refers to the proportion of common neighbors of two nodes in all their neighbors, that is

\[
n(i, j) = \frac{|n(i) \cap n(j)|}{|n(i) \cup n(j)|}
\]

(2)

where \( n(i) \) is the neighbor set of the node \( i \), \( n(j) \) is the neighbor set of the node \( j \).

Definition 3: Social Relationship \( SR(i, j) \). It's a linear combination of encounter frequency and common neighbor ratio of nodes, that is

\[
SR(i, j) = \alpha f(i, j) + \beta n(i, j)
\]

(3)

C. NODE ACTIVITY MODEL

The activity of nodes reflects the forwarding ability of nodes. It reflects a statistics of encounter probability in a node’s certain community. The more times a node encounters with other nodes in the community, the higher activity the node has in the community. When the node and the destination node are in the same community, choosing the node with higher activity is conducive to delivering messages to the destination node. Therefore, we only calculate the activity of the node when the node and the destination node are in the same community.

Definition 4 (Activity \( A(i) \)): We define node activity by the number of encounters between the node and other nodes in the local community. And the node \( i \)'s activity is marked as \( A(i) \), that is

\[
A(i) = \sum_{l \in l_i} \sum_{(i', j') \in C(i)} \frac{E(i, j')}{\sum_{(i', j') \in C(i)} E(i', j')}
\]

(4)

D. COMMUNITY CORRELATION MODEL

In SAN, the mobility range of nodes is limited. When nodes and destination nodes are in different communities, messages need to across one or more communities to deliver to the destination nodes. Avoid limiting messages to the local community, as this will result in messages not being transmitted to the intermediate community, not to mention the community where the destination node is located, for a limited lifetime. As a result, message transmission will fail and network resources will be wasted. Therefore, we define the community correlation to evaluate the overall correlation between the community where the node is located and the community where the destination node is located.

According to whether there is an intersection between different communities, the relationship between communities can be divided into overlapping and mutually independent relations. There are some common nodes between overlapping communities, that is, some nodes belong to two or more communities at the same time. Well, there is no common node between independent communities. (As shown in Figure 2(a), community A and community B are overlapping community relations, community A and community C are independent community relations). Compared with the independent community relationship, the communities with overlapping community relationship have more commonalities. The common nodes in the overlapping area generally have strong intermedation between the two communities. This kind of nodes can play a ferry role when messages need to be forwarded across communities. And the more common nodes the two communities have, the more similar the two communities are. In addition to considering the number of common nodes between communities, the historical interaction between two communities should also be evaluated in order to calculate the community correlation. If the nodes between the two communities encounter each other more frequently, the interaction between the two communities is stronger, and the message is more likely to be delivered to each other’s communities.

Therefore, the community correlation comprehensively considers the proportion of nodes in the overlapping area of overlapping communities and the historical interaction of the two communities. Choosing the community with greater correlation with the destination node’s community as the relay community can effectively prevent the transmission of messages being limited within the community that without the destination node. It makes the transmission of messages more directional.

Definition 5 Community Correlation \( CR(C_i, C_j) \). It refers to the degree of correlation between the communities where node \( i \) and node \( j \) are located. It is related to the degree of overlap and interaction between the two communities. If the node is in the overlapping area of the communities, the community correlation between the communities where the node is located and other communities is the sum of the community correlation of each community belonging to the node and other communities. While when the node is in an independent community, the community correlation between two communities is calculated only by their interactions. The specific calculation formula is as follows:

\[
CR(C_i, C_j) = \sum_{i \in l_i, j \in l_j} \left( \gamma \left| C_i \cap C_j \right| \right) + \delta \frac{E'(C_i, C_j)}{\sum E'(C_i, C_j')}
\]

(5)

where \( C' \) represents any other community. The community correlation is used only when the message needs to be forwarded across communities. Therefore, it is usually to calculate the community correlation between the communities of the node and the destination node. Assuming that the destination node is \( d \), that is, \( j=d \) in Formula (5).

IV. ROUTING ALGORITHM

The most important task of socially aware networking is to deliver messages. Traditional network has a stable and reliable end-to-end connection, but the network topology in SAN is unstable, and the message adopts the “store-carry-forward” transmission mode. In this mode, the node stores and carries messages while moving, and forwards the messages when it encounters the relay node, and the relay node forwards the messages to other relay nodes until it meets the destination node and then deliver the messages.
Therefore, the most important task in the design of routing strategy is how to select the most appropriate relay node. In this section, we synthesize the social relations between nodes, node activity and community correlation that mentioned above, and give the evaluation of functions about in-community and across-community message forwarding utility. On this basis, we design a routing algorithm. By comparing the forwarding utility of nodes, we choose the appropriate relay node, so as to improve the forwarding efficiency.

A. FORWARDING UTILITY FUNCTIONS
1) IN-COMMUNITY FORWARDING UTILITY
We use $ICF(i)$ for the function of in-community forwarding utility. It is the forwarding utility of node $i$ in the community, which is a linear combination of the social relationship $SR(i, d)$ between the node $i$ and the destination node $d$ and the activity $A(i)$ of the node $i$, that is,

$$ICF(i) = \lambda SR(i, d) + \eta A(i)$$  \hspace{1cm} (6)

2) ACROSS-COMMUNITY FORWARDING UTILITY
We use $OCF(i)$ for across-community forwarding utility. This function is node $i$’s across community forwarding utility, which is a linear combination of the social relationship $SR(i, d)$ between the node $i$ and the destination node $d$ and the community correlation $CR(C_i, C_d)$ of node $i$’s community $C_i$ and destination node $d$’s community $C_d$, that is,

$$OCF(i) = \phi SR(i, d) + \omega CR(C_i, C_d)$$  \hspace{1cm} (7)

B. ALGORITHM DESCRIPTION
Based on the above-mentioned functions of in-community forwarding utility and across-community forwarding utility, the basic idea of CCBR proposed in this section. Firstly, the message forwarding is divided into in-community forwarding and across-community forwarding according to the community location relationship (in the same community or different communities) between the node and the destination node. Secondly, select appropriate social attributes to evaluate node’s forwarding utility. Finally, appropriate relay node can be chosen according to the forwarding utility. In this way, the message can be quickly delivered to the destination node when it only needs to be forwarded within the community. And when the message needs to be forwarded across the community, it can avoid the resource cost caused by the message being limited to the current community or in other communities with little correlation with the community where the destination node is located. Therefore, the message can be transmitted in a more favorable direction.

Figure 1 shows how the CCBR algorithm makes the decision of the messages forwarding process, which proceeds according to the following steps:

Input: Network topology $G(V, E)$, $G'(C, E')$

1) When $i$ encounters other node $j$ with messages, if $j$ is the destination node of the message, it will directly deliver the message to $j$. Otherwise, it will detect whether node $i$ and the destination node are in the same community.

If true, it will go to step 2), otherwise it will go to step 3).

2) Detect whether the meeting node $j$ and the destination node are in the same community. If true, calculate separately $ICF(i)$ and $ICF(j)$. If $ICF(j) \geq ICF(i)$, $i$ will forward the message to $j$. Otherwise, $i$ will continue to carry the message to wait for meeting with the next node.

3) Detect whether the meeting node $j$ and the destination node are in the same community. If so, $i$ will forward the message to $j$. Otherwise, it will be calculated respectively $OCF(i)$ and $OCF(j)$.

If $OCF(j) > OCF(i)$, $i$ will forward the message to $j$. Otherwise, $i$ will continue to carry the message to wait for meeting with the next node.

The following examples illustrate the calculation method of forwarding utility of nodes. And the selection methods of relay nodes in different situations are explained with the examples. For convenience of calculation, weight coefficients in the formula are not considered for the moment. Figure 2 is a simplified network topology diagram (a) and a community connection graph (b). The number on the edge of the connection indicates the number of times the connections have been established between two nodes. Among them, node $i$ is the messages carrying node, node $j$ is the encountering node, and node $d_1$, $d_2$ and $d_3$ are the destination nodes of messages in different situations.

Situation 1: $d_1$ is the destination node. At this time, node $i$, $j$ and $d_1$ are in the same community, and messages only need to be forwarded within the community. By comparing
Therefore, 

\[ f(j, d_1) = 0 \]

\[ n(i, d_3) = 1/7 = 0.143 \]

\[ SR(j, d_3) = 0.143 \]

\[ CR(C_j, C_d) = 0 + 3/(3 + 9) + 2/7 + 6/(6 + 9) = 0.936 \]

and

\[ OCF(j) = 0.143 + 0.936 = 1.079 \]

Compare two nodes’ across-community forwarding utility, the result shows

\[ OCF(i) < OCF(j) \]

so node \( i \) forwards messages to node \( j \).

**V. PERFORMANCE EVALUATION**

In order to evaluate the performance of routing algorithms, this paper uses the widely used simulation tool ONE (Opportunistic Network Environment) version 1.5.1 to do some simulation experiments. ONE is an application of simulation based on Java programming language and is special for networks that use opportunistic sending of the packets. And we assume that the metrics involved in the formulas are of equal importance. So, the weight coefficients in the formulas are not considered (We set them all to be 1).

In this section, we compare our CCBR algorithm against the following three classical routing algorithms: Epidemic, Prophet and Direct Delivery. Among them, Epidemic is a simple encounter-based algorithm. It forwards the message to all the meeting nodes that means each node in the network can be a relay node. Prophet uses the history of node contacts to predict the future encounter probability. The heuristic method aim at finding appropriate relay node that has high probability to meet the destination nodes. Direct Delivery only forwards the message to the destination node without any relay process.

In order to verify the performance of CCBR proposed in this paper, the message delivery ratio, network overhead ratio and average delay of each routing algorithm are compared under different simulation duration and different Node buffer size.

**A. PERFORMANCE EVALUATION METRICS**

The optimization objectives of routing algorithms in SAN are to improve message delivery ratio and reduce transmission delay and network overhead. Therefore, the simulation experiment uses the following three metrics to evaluate the performance of the routing algorithms:

- Delivery Ratio: the ratio of the number of successfully delivered messages to the total number of created messages.
- Overhead Ratio: the proportion of the difference between the numbers of relayed messages and successfully delivered messages out of the successfully delivered messages.
- Average Latency: the average messages delay for all the successful sessions.

**B. SIMULATION ENVIRONMENT AND PARAMETER SETTING**

The relevant parameters in the simulation processes are shown in table 2.

**C. EXPERIMENT RESULTS AND ANALYSIS**

We evaluate the performance of the algorithm by changing the simulation duration and the buffer size of node.

1) THE EFFECT OF THE SIMULATION DURATION ON THE ALGORITHM PERFORMANCE

At this time, the buffer size of each node is set to 50M, and the simulation duration is set respectively to 2h, 4h, 

\[ ICF(i) \] and \( ICF(j) \), we can decide whether to forward the message to node \( j \). According to Formula (1), we can calculate

\[ f(i, d_1) = 3/(3 + 2 + 1) = 0.5 \]

And get

\[ n(i, d_1) = 1/5 = 0.2 \]

according to Formula (2). So we can get node \( i \) and destination node \( d_1 \)’s social relationship

\[ SR(i, d_1) = 0.5 + 0.2 = 0.7 \]

according to Formula (3). Then calculate node \( i \)’s activity

\[ A(i) = (3 + 2)/(3 + 2 + 2 + 1) = 0.625 \]

According to Formula (4). Thus node \( i \)’s in-community forwarding utility value

\[ ICF(i) = 0.7 + 0.625 = 1.325 \]

by using (6). Similarly,

\[ f(j, d_1) = 2/(2 + 2 + 1 + 1 + 1 + 1) = 0.222 \]

\[ n(j, d_1) = 2/8 = 0.25 \]

so

\[ SR(j, d_1) = 0.222 + 0.25 = 0.472 \]

\[ A(j) = (2 + 2 + 1)/(3 + 2 + 2 + 1) = 0.625 \]

and

\[ ICF(j) = 0.472 + 0.625 = 1.097 \]

Compare two nodes’ in-community forwarding utility, the result shows

\[ ICF(i) > ICF(j) \]

so node \( i \) continues to carry messages.

Situation 2: \( d_2 \) is the destination node. In this case, node \( i \) and the destination node \( d_2 \) are in different communities, node \( j \) and the destination node \( d_2 \) are in the same community. The node \( i \) will forward the message to node \( j \).

Situation 3: \( d_3 \) is the destination node. At this time, \( i, j \) and \( d_3 \) are respectively in different communities, and messages need to be forwarded across communities. In this case, it will decide whether to choose \( j \) as the relay node by comparing

\[ OCF(i) \] and \( OCF(j) \). According to Formula (1), calculate

\[ f(i, d_3) = 0 \]

According to Formula (2), calculate

\[ n(i, d_3) = 0 \]

So, node \( i \) and the destination node \( d_3 \)’s social relationship

\[ SR(i, d_3) = 0 \]

according to Formula (3). Thus we can get node \( i \) and the destination node \( d_3 \)’s community correlation

\[ CR(i, d_3) = 0 + 3/(3 + 9) = 0.25 \]

according to Formula (5). Therefore,

\[ OCF(i) = 0.25 \]

In the same way, we can calculate

\[ f(j, d_3) = 0 \]

\[ n(i, d_3) = 1/7 = 0.143 \]

\[ CR(j, d_3) = 0 + 3/(3 + 9) + 2/7 + 6/(6 + 9) = 0.936 \]

and

\[ OCF(j) = 0.143 + 0.936 = 1.079 \]

FIGURE 2. Simplified network topology and community structure.
TABLE 2. Simulation parameter setting.

| Parameter                | Value                          |
|--------------------------|--------------------------------|
| Topological area         | 4500m*3400m                    |
| Moving model             | Shortest Path Map Based Movement |
| Number of nodes          | 126                            |
| Number of communities    | 6                              |
| Node communication mode  | Bluetooth                      |
| Signal transmission radius | 10m                           |
| Node buffer size         | 50MB                           |
| Message creation interval | 25 to 35 second               |
| Message sizes            | 500 KB to 1000 KB             |
| Simulation duration      | 8 hours                        |
| Message TTL              | 300 minutes                    |

FIGURE 3. Comparisons of routing algorithms on delivery ratio under different simulation duration.

6h, 8h and 10h. The simulation results show that with the increase of simulation time, CCBR algorithm achieves better performance than Epidemic, Prophet and Direct Delivery. It can improve the delivery ratio and effectively reduce the network overhead ratio. Next, we compare the performance of the four routing algorithms from three aspects: message delivery ratio, network overhead ratio and average latency.

Figure 3 shows the performance comparison of routing algorithms in delivery ratio with the change of simulation duration. As we can see from the figure, the delivery ratio of Direct Delivery algorithm is always the lowest. Because the source node of the algorithm always carries messages before encounters the destination node. And the message is not delivered directly until the source node encounters the destination node. In Epidemic routing, each encounter node is the relay node of messages. Messages can be spread to the whole network quickly, so it can have a high delivery ratio. The CCBR algorithm has a low delivery ratio at the beginning. But with the increase of simulation duration, the delivery ratio rises rapidly. Because there are more and more social information can be used to analyze the social relationship between nodes. So the selection of relay nodes is more and more accurate. And the delivery ratio can be rapidly improved, surpassing Prophet routing, just next to Epidemic routing.

FIGURE 4. Comparisons of routing algorithms on overhead ratio under different simulation duration.

Figure 4 compares the network overhead ratio of each routing algorithm under different simulation duration. Among them, the cost of Direct Delivery is zero because there is no message replication. Epidemic and Prophet will produce a large number of copies in the network. So the network overhead will be very high and the network performance will be unstable. The CCBR algorithm chooses a more suitable relay node when forwarding messages. It can reduce the number of times the messages need to be forwarded and effectively control the number of copies in the network. Therefore, the network overhead of CCBR is always lower than that of Epidemic and Prophet, and its performance is stable.

As it can be seen from Figure 5, with the increase of simulation duration, the average latency of each algorithm is on the rise. When the simulation duration is less than 4 hours, the average latency of the four algorithms is basically the same. But when the simulation duration is more than 4 hours, due to the message of Direct Delivery algorithm will be kept in the source node, waiting to meet the destination node, so its average latency is the largest. The average latency of Epidemic rises rapidly. That is because more messages are transferred due to the increase of simulation duration, thus increasing the time for successful reception of messages. At first, the average latency of CCBR algorithm
increases quickly. When the simulation duration exceeds 6 hours, the rising speed slows down. When the simulation time exceeds 8 hours, the average latency of CCBR is lower than that of Epidemic algorithm. The reason is that the longer the simulation time is, the more social information CCBR can obtain and the more suitable relay nodes can be selected to make messages reach the destination node through the best path. So there is a slowdown in the growth of average latency of CCBR. Prophet algorithm uses historical information to calculate the encounter probability between nodes. On the one hand, it can screen relay nodes. On the other hand, it has less calculation. So it is better than the other three algorithms in average latency.

2) THE IMPACT OF BUFFER SIZE ON ALGORITHM PERFORMANCE

In this group of experiments, we set the simulation duration as 8 hours. The buffer size of each node is set to 5MB, 10MB, 15MB, 20MB and 25MB respectively.

Due to the limited cache space of mobile devices, the amount of information that nodes in SAN can carry will be affected by the cache space. Therefore, the performance of routing algorithm in SAN is related to the buffer size of node. We analyze the impact of buffer size on the performance of the four routing algorithms through simulation experiments. The experimental results show that with the change of cache space, CCBR algorithm has the highest delivery ratio compared with the other three algorithms. And it has the lowest network overhead ratio except Direct Delivery. In Direct Delivery routing, the message is delivered directly to the destination node when it is connected with the destination node. There is no need for message replication, so the buffer size has little impact on the performance. The following is a detailed analysis of the experimental results of each routing in terms of message delivery ratio, network overhead and average latency.

Figure 6 compares the delivery ratio of each algorithm. Except for the Direct Delivery algorithm, the delivery ratios of the other three algorithms are greatly improved with the increase of buffer size. Direct Delivery routing does not copy messages, so it does not require high demand for buffer size. Therefore, with the increase of buffer size, the delivery ratio of the Direct Delivery routing is stable. Epidemic and Prophet will produce a large number of copies. When the buffer size is small, it is easy to cause congestion, resulting in the loss of message copies. Therefore, the delivery ratio is low. With the increase of buffer size, the constraint situation is improved and the delivery ratio is improved rapidly. Compared with the other three routings, CCBR algorithm has the best delivery ratio, and has a considerable delivery ratio even when the buffer size is small.

Figure 7 describes the comparison of network overhead ratio. With the increase of buffer size, the network overhead of Epidemic routing decreases rapidly. The overhead ratio of CCBR algorithm is significantly lower than that of Epidemic and Prophet. Its overhead ratio is 84.4% lower than Epidemic algorithm and 79.6% lower than Prophet algorithm. The Direct Delivery algorithm has no network overhead because there is no relay process.

With the increase of buffer size, the network overhead of Epidemic routing decreases rapidly. The overhead ratio of CCBR algorithm is significantly lower than that of Epidemic and Prophet. Its overhead ratio is 84.4% lower than Epidemic algorithm and 79.6% lower than Prophet algorithm. The Direct Delivery algorithm has no network overhead because there is no relay process.

Figure 8 shows the effect of buffer size on transmission delay. Because the Direct Delivery algorithm only delivers messages when the connection is established with the destination node, there is a long waiting time, so the average latency is always the largest and has little change. The average latencies of other three algorithms are on the rise, and Epidemic and Prophet are rising faster, while CCBR is relatively slow. The performance of CCBR algorithm on average latency is not as good as that of Epidemic and Prophet. That is because each encounter node participates in message transmission in flooding mechanism of Epidemic, which greatly reduces the transmission delay of messages. Prophet routing only uses limited historical information to calculate the probability of meeting between nodes, which saves a lot of time. CCBR algorithm considers a variety of factors comprehensively, that makes the selection of relay nodes more targeted but the calculation more complicated. That is why the CCBR
algorithm has a higher average latency than Epidemic and Prophet.

Through the above two groups of simulation experiment results and analysis, it can be seen that the comprehensive performance of the CCBR proposed in this paper is better than the other three routing algorithms in terms of message delivery ratio and overhead ratio. The CCBR algorithm can effectively save network overhead while maintaining a high message delivery ratio. That is because the CCBR algorithm not only considers the influence of the social relationship between nodes on the mobility of the node, but also considers the impact of the node’s own activity on the forwarding ability of the node. Furthermore, the CCBR algorithm considers the bridge role of the community correlation when the message needs to be forwarded across the community. While other three algorithms only consider a single factor, so they only have good performance in a certain statistic of routing performance.

VI. CONCLUSION

By studying the social relationship and community correlation of nodes in SAN, this paper redefines the relationship between nodes, node’s activity and community correlation, and proposes a routing scheme based on community correlation in socially aware networking. The routing scheme can use the social attributes of nodes to evaluate the relay ability of nodes through the social relationship between nodes and the activity of nodes in the in-community forwarding stage. Whereas in the across-community forwarding stage, the social relationship between the pending relay node and the destination node, and the community correlation between the community where the pending relay node located and the community that the destination node belongs are used as the basis for relay node selection. Finally, through the simulation experiments, the routing algorithm proposed in this paper is compared and analyzed with the existing famous algorithms. This protocol effectively improves the message delivery rate and greatly reduces the network overhead. And the scheme we proposed can achieve better performance compared with the existing routing schemes even under the limited cache.

While ensuring a high delivery rate, it can also maintain a low overhead. Next, we will conduct research on the incentives and routing security of selfish nodes.

REFERENCES

[1] China Internet Network Information Center. The 45th Statistical Report on Internet Development in China. [Online]. Available: http://www.cac.gov.cn/2020-04/27/c_1585535470378587.htm
[2] Z. Ning, L. Liu, F. Xia, B. Jedari, I. Lee, and W. Zhang, “CAIS: A copy adjustable incentive scheme in community-based socially aware networking,” IEEE Trans. Veh. Technol., vol. 66, no. 4, pp. 3406–3419, Apr. 2017.
[3] I. Bisso, F. Lavagetto, A. Sciaronne, T. Penner, and M. Gugjuis, “Context-awareness over transient cloud in D2D networks: Energy performance analysis and evaluation.” Trans. Emerg. Telecommun. Technol., vol. 28, no. 2, Feb. 2017, Art. no. e3002.
[4] M. M. Tulu, M. E. Mikramweni, R. Hou, S. Feisso, and T. Yousan, “Influential nodes selection to enhance data dissemination in mobile social networks: A survey.” J. Netw. Comput. Appl., vol. 169, Nov. 2020, Art. no. 102768.
[5] P. Hui and J. Crowcroft, “How small labels create big improvements.” in Proc. ACM CoNEXT Conf., Whiteplains, NY, USA, 2006, pp. 65–70.
[6] J. Shi, X. Wang, and M. Huang, “Community Aware MSN Routing Scheme Based on ICN Architecture,” J. Softw., vol. 31, no. 6, pp. 1786–1801, 2020.
[7] E. Bulut and B. K. Szymanski, “Exploiting friendship relations for efficient routing in mobile social networks,” IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 12, pp. 2254–2265, Dec. 2012.
[8] A. Vahdat and D. Becker, “Epidemic routing for partially connected ad hoc networks,” in Handbook of Systemic Autoimmune Diseases. 2000.
[9] M. Abdelmoumen, E. Dhib, M. Frihka, and T. Chahed, “How to improve the performance in delay tolerant networks under manhattan mobility model,” in Proc. 21st Annu. IEEE Int. Symp. Pers., Indoor Mobile Radio Commun., Instanbul, Turkey, Sep. 2010, pp. 2008–2013.
[10] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, “Spray and wait: An efficient routing scheme for intermittently connected mobile networks,” in Proc. ACM SIGCOMM workshop Delay-tolerant Netw., 2005, pp. 252–259.
[11] A. Lindgren, A. Doria, and O. Schelén, “Probabilistic routing in intermittently connected networks,” ACM SIGMOBILE Mobile Comput. Commun. Rev., vol. 7, no. 3, pp. 19–20, Jul. 2003.
[12] E. M. Daly and M. Haahr, “Social network analysis for routing in disconnected delay-tolerant MANETs,” in Proc. 8th ACM Int. Symp. Mobile Ad Hoc Netw. Comput., Montreal, QC, Canada, 2007, pp. 32–40.
[13] E. M. Daly and M. Haahr, “Social network analysis for information flow in disconnected delay-tolerant MANETs,” IEEE Trans. Mobile Comput., vol. 8, no. 5, pp. 606–621, May 2009.
[14] P. Hui, J. Crowcroft, and E. Yoneki, “BBBLE rap: Social-based forwarding in delay-tolerant networks,” IEEE Trans. Mobile Comput., vol. 10, no. 11, pp. 1576–1589, Nov. 2011.
[15] S. Tsugawa, “A survey of social network analysis techniques and their applications to socially aware networking,” IEICE Trans. Commun., vol. E102.B, no. 1, pp. 17–39, Jan. 2019.
[16] J.-W. Chang and C. Chen, “CROP: Community-relevance-based opportunistic routing in delay tolerant networks,” in Proc. IEEE Wireless Commun. Netw. Conf. (WCNC), Shanghai, China, Apr. 2013, pp. 1938–1943.
[17] J. Niu, “A message transmission scheme for community-based opportunistic network,” J. Comput. Res. Develop., vol. 46, no. 12, pp. 2068–2075, 2009.
[18] Z. Li, C. Wang, S. Yang, C. Jiang, and X. Li, “LASS: Local-activity and social-similarity based forwarding in mobile social networks,” IEEE Trans. Parallel Distrib. Syst., vol. 26, no. 1, pp. 174–184, Jan. 2015.
[19] X. Meng, “Message delivery strategy research based on community division in mobile social networks.” Ph.D. dissertation, Dept. CSTI, Jinli Univ., Changchun, China, 2017.
[20] W. Wang, “Research on message transmission strategy and caching mechanism in mobile social networks,” M.S. thesis, Dept. School Comput., Central China Normal Univ., Wuhan, China, 2018.
[21] M. Deng, F. Xu, Z. Xiong, C. Ye, and H. Xie, “Social-aware data forwarding algorithm in mobile social networks,” Comput. Eng. Des., vol. 30, no. 7, pp. 1835–1840, 2018.
[22] F. Xu, Q. Xu, Z. Xiong, N. Xiao, Y. Xie, M. Deng, and H. Hao, “Intelligent distributed routing scheme based on social similarity for mobile social networks,” Future Gener. Comput. Syst., vol. 96, pp. 472–480, Jul. 2019.
[23] Y. Xie, “Research on high performance message transmission algorithm and cache management mechanism in mobile social networks,” M.S. thesis, Dept. School Comput., Central China Normal Univ., Wuhan, China, 2019.
[24] F. Xu, M. Deng, Z. Xiong, C. Ye, and N. Xu, “Data forwarding algorithm based on multidimensional context matching in mobile social networks,” Comput. Sci., vol. 46, no. 2, pp. 81–87, 2019.
[25] M. Girvan and M. Newman, “Community structure in social and biological networks,” in Proc. NAS, New York, NY, USA, 2001, pp. 7821–7826.

QIONG XU was born in 1991. She received the B.S. degree in computer science and technology from the Hubei University of China in 2013, where she is currently pursuing the M.S. degree in computer technology. She is involved in research on routing algorithms for mobile social networks.

MIN DENG received the M.S. degree from Wuhan University, Wuhan, China, in 2010. She is currently a Lecturer with the School of Computer and Information Science, Hubei Engineering University, Xiaogan, Hubei, China. Her research interests include wireless mobile networks and context aware computing.

FANG XU received the M.S. and Ph.D. degrees from Wuhan University, Wuhan, China, in 2009 and 2016, respectively. He is currently a Professor with the School of Computer and Information Science, Hubei Engineering University, Xiaogan, Hubei, China. His research interests include social computing, wireless mobile networks, and context aware computing.

XIAOCHAO ZHAO received the B.S. and Ph.D. degrees in software engineering from Hunan University, Changsha, China, in 2012 and 2018, respectively. He has been a Lecturer with Hubei Engineering University since 2019. His research interests include image processing and computer vision.

ZHE LI received the Ph.D. degree from Northwestern Polytechnical University. He is currently an Associate Professor with Hubei Engineering University. His research interests include the consistency control of multiple agents in complex environments, multi-rates with multi-agent systems, and distributed control systems.

ZENGGANG XIONG received the M.S. degree in computer application from Hubei University in 2005 and the Ph.D. degree in computer application from the University of Science and Technology Beijing in 2009. He is currently a Professor in computer science with Hubei Engineering University. His research interest includes cloud computing and big data.

NAN XIAO was born in 1995. She received the B.S. degree in computer science and technology from Hubei Engineering University, China in 2018. She is currently pursuing the M.S. degree in computer technology with the Hubei University of China. She is involved in research on incentives for selfish nodes in mobile social network.

GUOTAO ZHAO is currently an Associate Professor with Hubei Engineering University, Xiaogan, Hubei, China. His research interests include computer network and image identification.

ZHEN LIU received the B.S. degree in computer science and technology from Hubei Engineering University, China, in 2008, and the M.S. degree in computer application technology from the Kunming University of Science and Technology, China, in 2011. He is currently pursuing the Ph.D. degree in control science and engineering with Jiangnan University, China. His research interests include pattern recognition and machine learning.