IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 3, 2019, accepted June 5, 2019, date of publication June 12, 2019, date of current version June 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2922350

A Transmission Prediction Mechanism Exploiting Comprehensive Node Forwarding Capability in Opportunistic Networks

PENG ZHENG[1], HONGXIAO FEI[1], JIA WU[1,2], (Member, IEEE), AND YEQING YAN[1,2]

[1]School of Computer Science and Engineering, Central South University, Changsha 410075, China
[2]"Mobile Health" Ministry of Education-China Mobile Joint Laboratory, Changsha 410075, China

Corresponding author: Hongxiao Fei (hxfei@csu.edu.cn)

This work was supported in part by the "Mobile Health" Ministry of Education−China Mobile Joint Laboratory and in part by the National Natural Science Foundation of China, under Grant 41571397, Grant 51778242, and Grant 41871364.

ABSTRACT Opportunistic network enables users to form an instant network for data sharing, which is a type of Ad-hoc network in nature, thus depends on cooperation between nodes to complete message transmission. Because of intermittent communication and frequent changes of topology structure in opportunistic networks, the duration of node encounters is limited, as well as the length of established connections. If the amount of interactive data is large and the communication bandwidth is small, the messages that need to be transmitted are not guaranteed to be delivered successfully every time. In this regard, this paper establishes a transmission prediction mechanism exploiting comprehensive node forwarding capability (TPMEC) in opportunistic networks. When quantifying the forwarding capability of nodes, the algorithm not only considers the cooperative tendency but also discusses the encounter strength between nodes. At the same time, in order to find out all key nodes during the transmission process, the algorithm adopts the theory of matrix decomposition to predict and supplement the missing forwarding capability value of nodes, thus improving the efficiency of message transmission. Simulation results show that compared with ITPCM algorithm, ETNS algorithm, Spray and Wait algorithm, and PRoPHET algorithm, the proposed scheme has the highest transmission success ratio and the lowest routing overhead.

INDEX TERMS Opportunistic networks, forwarding capability, cooperative tendency, encounter strength, matrix decomposition.

I. INTRODUCTION

With the advent of 5G network era and the explosive growth of intelligent terminals, more and more researchers have begun to consider how to organize these devices into opportunistic networks, thereby solving problems such as traffic overload and network congestion [1]. Opportunistic network is an intermittently connected network [2] that does not require a complete end-to-end communication path between the source node and the destination node. Different from traditional wireless networks, the opportunistic network has the characteristics of frequent node movement, intermittent communication, and limited resources [3]. Therefore, in opportunistic networks, the message transmission depends on the movement of nodes to generate the meeting opportunities for communication, in which the nodes use the "storage-carry-forward" mechanism for data delivery [4]. With the advantages of not relying on infrastructure, zero traffic cost and enhancing wireless network coverage, opportunistic networks are considered as one of the important development directions of mobile Ad-hoc networks, as well as a significant part of ubiquitous computing in future [5].

In the new round of technological change and industrial upgrading, the integration of the internet and various fields for development has become an irresistible trend. With the rapid development of internet of vehicles, internet of things and 5G network, people's lifestyle has been transformed into a social form based on the internet [6]. The popularization of mobile devices, as well as people's demand for data communication on the internet at any time and any where, has jointly

The associate editor coordinating the review of this manuscript and approving it for publication was Wen Chen.

spawned the era of data explosion. In order to reduce the load and congestion caused by huge amounts of data on cellular networks, people can use mobile devices to self-organize into opportunistic networks for communication [7]. For example, all kinds of handheld devices that support wireless communication can realize information sharing and data communication through self-organized networks. Vehicles shuttling on the road can be self-organized into vehicle networks to achieve vehicle diversion, traffic safety, and danger warning. The adoption of opportunistic networks has the advantages of not relying on infrastructure, enhancing wireless network coverage and zero traffic cost [8], making the opportunistic networks usher in unprecedented opportunities for development in the context of big data and 5G networks.

Traditional wireless Ad-hoc networks lack of measures to deal with network fragmentation and connection interruption [9]. When the nodes in the network are sparsely distributed, frequently moving and blocked by obstacles, the performance of the network will obviously decline, and even make the network unable to communicate. In the foreseeable challenge environment, people need to complete data communication at any time and any where. Opportunistic network does not need the complete communication path of the source node and the destination node [10], so it is more suitable for the actual requirements of wireless networks.

However, due to intermittent communication and frequent changes of topology structure in opportunistic networks, the duration of node encounters is limited, as well as the length of established connections. If the amount of interactive data is large and the communication bandwidth is small, the messages that need to be transmitted are not guaranteed to be delivered successfully every time [11]–[15]. At the same time, there may be anomalous nodes in the network, forging and spreading malicious Trojans and viruses. How to identify anomalous nodes is also a problem that needs to attract attention in opportunistic networks [16]. In addition, the key nodes that refer to non-anomalous nodes with strong forwarding capability are extremely important for the efficiency of data delivery. How to make full use of the existing information to find out all key nodes as relay nodes during message transmission process is also a problem to be considered in opportunistic networks.

To solve these problems, this work proposes a transmission prediction mechanism exploiting forwarding capability of nodes in opportunistic networks. The strategy is divided into two phases: the preparation phase and the transmission phase. In the preparation phase, each node establishes an archive file that stores network information, and then updates the list of state sequences by interacting with each other. In the transmission phase, the node comprehensively quantifies the forwarding capability of neighbor nodes through the collected information, including cooperation tendency and encounter strength. At the same time, in order to find out all key nodes during message transmission process, we adopt the method of matrix decomposition to predict and supplement the missing value of nodes. In addition, the proposed strategy is based on a multi-copy routing strategy. In order to control the number of message copies, we employ positive transmission method to deliver data, which can make the data spread in the direction of increasing forwarding capability of nodes. In a word, this paper improves the success rate of data delivery and reduces the average end-to-end delay by calculating the forwarding capability of nodes and using matrix decomposition strategy. Specifically, the contribution of this paper is mainly in the following aspects:

1) We have constructed an effective information collection strategy in which nodes can efficiently collect and update information about the network in a short period of time.
2) When measuring the cooperative tendency between nodes, we comprehensively evaluate the reputation value from sending trust and receiving trust, which adapts to the dynamic change and supports the automatic formation.
3) The concept of encounter strength is proposed by combining the probability and duration of encounters, as an important index to measure the forwarding capability of nodes.
4) The method of matrix decomposition in machine learning is adopted to predict and supplement the missing forwarding capability value of nodes.

The rest of the paper is structured as follows. In Section 2, we will briefly give a overview of the existing related work. The system model of the proposed algorithm will be presented and analyzed in Section 3. In Section 4, We verify the proposed algorithm through the experiments. In the last Section, a final summary will be presented.

II. RELATED WORK

Opportunistic networks have been extensively researched for the ability to accomplish communication tasks without relying on the inherent communication facilities. At present, the research hotspots of opportunistic networks mainly focus on routing mechanisms, communication security, user privacy and congestion control, among which the research on routing mechanisms is particularly prominent.

In opportunistic networks, some researchers have proposed community-based routing protocols. They believe that the movement of nodes has the characteristics of regularity and periodicity. Nodes in the same community have closer social relations, so they have a greater probability to transmit messages. Chen *et al.* [17] proposed a scheme of dynamic access to node information to divide communities, which can update the social characteristics and contact behavior of nodes in a timely manner. When the source node needs to transmit information, the node chooses the appropriate relay node according to the distribution structure of the community in the network. J. Tao *et al.* [18] proposed a community forwarding strategy based on contact perception, which is divided into internal forwarding stage and external forwarding stage. This strategy studies the activity of the node, evaluates the probability of the node meeting the destination node, and improves the success rate of message transmission by passing messages to the nodes in the same community as the destination node. Li.D *et al.* [19] studied the potential properties of node profiles and proposed a community detection strategy based on user profiles. In this method, the nodes in the network are divided into communities by analyzing their characteristics such as interest and behavior. Yan *et al.* [20] suggested removing inefficient nodes in the process of dividing the community, which could reduce the computational complexity and save energy in the process of message delivery. However, the algorithm ignores the fact that receiving too many copies of messages will cause the active node network to become congested.

Routing protocols based on node context information have also attracted researchers' attention. When selecting the relay node, the source node will compare the social attributes of neighbor nodes to calculate the similarity between nodes. The more similar the relay node is to the target node, the more likely it is to complete the message transmission. Such routing algorithms often need to collect a large number of node information and perform repeated and complex calculations. Kiyoung *et al.* [21] uses the method of fuzzy reasoning to combine the location similarity and attribute similarity of nodes, and chooses the appropriate relay nodes by comparing the similarity of neighbor nodes. This method can also accomplish the task of information communication under the condition of limited resources. Kumar *et al.* [22] proposed a routing algorithm that is aware of the node context information. In the algorithm, each node stores the environment information and adopts fitness quasi side optimization algorithm parameters. The source node uses the genetic search method to predict the transmission path strength of message and selects the optimal relay node by evaluating the transmission efficiency of neighbor nodes.

Liu *et al.* [23] proposed a routing algorithm to comprehensively evaluate the similarity of nodes from multiple dimensions. The algorithm evaluates the transmission preference of nodes by calculating their mobility characteristics and social attributes. By comparing the similarity between the neighbor nodes and the destination node, the source node can get the recommendation of the best relay node. This algorithm can effectively improve the efficiency of data transmission, but frequent computation consumes a lot of node resources. Lin.Y *et al.* [24] evaluated the similarity by analyzing the packet communication distance between nodes. By combining the characteristics of opportunistic network and social network, this strategy can effectively avoid the problem of low transmission efficiency caused by the movement of nodes and the dynamic change of networks.

Many algorithms use models in mathematics and machine learning to optimize the performance of optimization algorithms, such as graph theory, Markov chain, decision tree, etc. In order to obtain crowd density in group outsourcing tasks, Nguyen *et al.* [25] connected it with vertex coverage in graph theory, and proposed a context-aware computing algorithm to find vertex coverage in crowd perception. In this algorithm, people's interests and habits are taken as the starting point, and people are initially allocated and planned. Then, people-centered sensor tasks are handed over to a more active device to find a appropriate vertex coverage. Dhurandher *et al.* [26] proposed an opportunistic network routing strategy based on markov chain. The algorithm takes into account that the geographical location of the node changes constantly with the passage of time, so markov chain is used to predict the future location region of the node based on the current location of the destination node. At the same time, the method evaluates the probability of nodes to the destination region to select the appropriate relay nodes so as to complete the task of communication. Li N *et al.* [**?**] suggested an opportunistic network routing protocol based on fuzzy logic and topology control. In this algorithm, nodes are divided into different types according to their degrees. The higher the degree, the higher the priority. At the same time, the algorithm considers the influence of node moving speed and moving square on the transmission efficiency. When selecting the relay node, the source node evaluates the practicability of the neighbor node through the algorithm, and selects the node with strong practicability as the relay node.

The algorithms based on the history interaction information in opportunistic networks have also been deeply studied by many scholars. This kind of algorithm usually records and collects the information of nodes in the network. According to historical information, relevant indicators are calculated and inferred to provide theoretical support for data delivery. Therefore, effective data collection method is very important for these algorithms. Lee *et al.* [27] uses the history of message transmission between nodes to evaluate the probability that a node can deliver data to the destination node after receiving it. At the same time, when the previous predictability value of the neighbor node is higher than that of the source node, the source node will pass the message to the node.

Wang *et al.* [28] proposed a routing algorithm based on identity perception. According to the importance of social attributes of nodes to information transmission, the algorithm takes social identity and social influence into account. This strategy only considers the social relationship of nodes, but lacks the analysis of transmission preference and mobile behavior of nodes, resulting in a single application scenario. Borah *et al.* [**?**] proposed an energy-saving routing strategy based on the history of node encounter and data delivery. The author believes that nodes need to consume a lot of energy in the process of information transmission, so how to maximize the energy saving is an important problem faced by routing algorithms. When designing the routing algorithm, Zhou *et al.* [29] proposed a routing algorithm based on time compactness and centrality, taking into account the mean separation time and separation time variance between nodes. This strategy analyzes the contact correlation between nodes and predicts the future contact patterns of nodes to optimize the algorithm of message transmission.

III. SYSTEM MODEL DESIGN

The characteristics of intermittent communication and frequent node movement determine how to improve the efficiency of communication is a fundamental problem in the researches of opportunistic networks. This paper comprehensively measures the forwarding capability of nodes from the two dimensions of cooperation tendency and encounter strength. Then the optimal relay node can be recommended by forwarding capability matrix. At the same time, we apply the method of matrix decomposition in machine learning to the routing algorithm, which can effectively improve the efficiency of message transmission. Based on this model, we will introduce the algorithm in detail from the following three stages.

A. COLLECT INFORMATION ABOUT NODES IN THE NETWORK

Information collection is a key research issue in wireless sensor networks, and it is also the theoretical basis of algorithm implementation. In this paper, we propose a special preparation phase in which each node obtains and updates accurate information in the network. Information is collected only in the preparation phase. In this scheme, each node has an archive file to storage the information related to the users. This information includes details about the device holder, the status of devices, record of node encounters, etc.

In the preparation phase, if the node collects the information in the network for a long time, too much statistical data may lead to insufficient computing resources and occupy the excessive cache of nodes, making the network delay increase significantly. On the contrary, a short-term preparation phase may mean that the node cannot obtain enough information about the network, and thus cannot provide sufficient data support for the message transmission decision. Therefore, proper preparation time is crucial to the whole algorithm. Depending on the social network application scenario, the length of the preparation phase ΔT is typically set according to the node's activity cycle [30]. Humans have a fixed physiological cycle of 24 hours a day, and it is better to engage in exercise or physical activity according to this rule. This physiological law is called Circadian Rhythm. Users using mobile devices are more likely to perform repeated mobile roads every day and have a greater probability of encountering the same person or arriving at the same location [25], [31]. So, in real social scenarios, the length of time for the preparation phase should be set to 24 hours. Of course, the optimal preparation time also varies in different application scenarios. For the sake of accurately quantifying the information acquisition and updating process in the preparation phase, the state sequence of node v is defined as:

$$List(v) = < Pos_v, Cre_v, \kappa_v * List_s > \tag{1}$$

$$*List_s = < List(a), List(b), List(c), \ldots, List(n) > \tag{2}$$

where Cre_v represents the cooperative tendency value of node v, including the sending trust and the receiving trust.

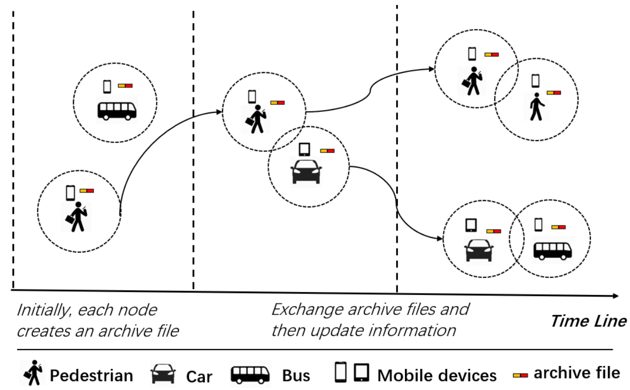

FIGURE 1. The process of collecting information.

Pos_v represents the mobile distribution of node v and κ_v is a feature vector of node v which has multiple attributes. $*List_s$ is the set of state sequences of other nodes encountered by node v during a given preparation period ΔT. At the same time, if a node v and another node u encounter in the preparation phase, the formula (3) is used to structure a table item $List(1)$ for recording the encounter information between nodes v and u. In addition, the encounter state matrix $E(n)$ of all the list items can be constructed by means of the equation (4).

$$\begin{aligned} List(1) &= List(v) \cup List(u) \\ &= < Pos_v, Cre_v, \kappa_v, Pos_u, Cre_u, \kappa_u, *List_m > \end{aligned} \tag{3}$$

$$E(n) = [List(1), List(2), List(3), \cdots, List(n)] \tag{4}$$

As shown in Figure 1 and Figure 2, in the preparation phase, each node builds up a sequence list of states based on its own device information and historical interaction information. Because of the mobility of nodes, when a node encounters another node, the two nodes will update their lists by exchanging archive files from each other. In addition, multiple nodes in the network can also cooperate to complete the update process of lists. For each encounter between nodes, the strategy expands table items and records the meeting information, which contains all interaction data between nodes. At the same time, in order to set up a unified data set about nodes in the network and provide data support for the message transmission decision, the encounter state matrix $E(n)$ will be established in the storage space of nodes. Different from the inherent attributes of nodes in the network, the mobile status of nodes changes with time. Therefore, updating the nodes' encounter state matrix in time can improve the timeliness and accuracy of the statistical data.

B. FORWARDING CAPABILITY OF NODES

In opportunistic networks, nodes rely on the "storage-carry-forward" method to deliver data. When there is no complete communication path between the source node and the destination node, the choice of relay nodes becomes crucial. According to the principle of opportunistic network, the more

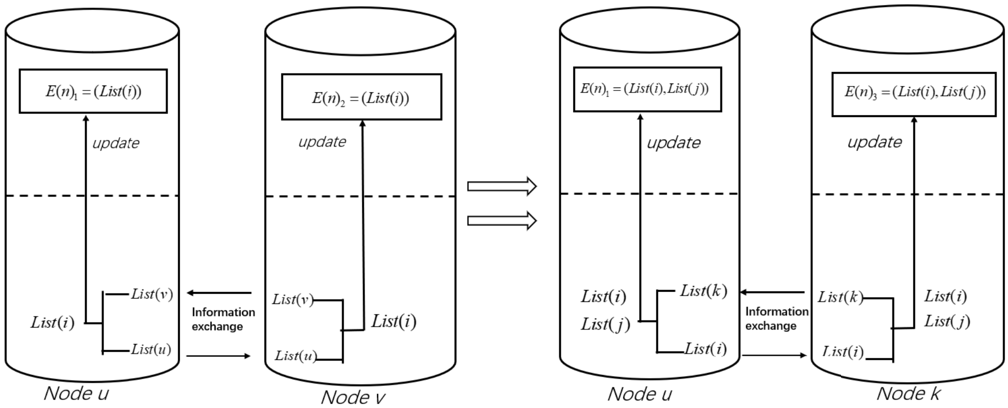

FIGURE 2. **Details of information collection in the preparation phase.**

copies of data in the transmission process, the greater of probability that the data will be successfully transmitted to the destination node. However, due to the limited cache and energy of nodes, blindly increasing the number of data copies will lead to excessive consumption of network resources. Therefore, reasonable selection of relay nodes and control the number of message replicas are very important to improve network performance.

In view of the phenomenon of refusal to receive messages and maliciously spread viruses in the opportunistic network, this paper uses the cooperative tendency indicator to identify anomalous nodes. At the same time, most routing algorithms do not take into account the impact of the duration of encounter between nodes on the propagation of messages. In this regard, we propose the concept of the encounter strength. Therefore, in this paper, the forwarding capability of nodes is measured by cooperation tendency and probability of encounter.

1) CALCULATION OF COOPERATION TENDENCY

This paper evaluates the cooperative tendency between nodes based on the historical interactive information, including local cooperative tendency and global cooperative tendency. Local cooperation tendency refer to the reputation assessment between two nodes. Global cooperative tendency refers to the reputation evaluation of each node from the perspective of the whole network. Historical interaction information is obtained by sending trial packets by nodes during the preparation phase.

In order to effectively quantify the cooperative tendency of nodes, we set up a feedback mechanism: the nodes will give satisfactory or unsatisfactory evaluation after the data delivery. The more satisfactory evaluation, the stronger cooperative nature of the node will be. The local cooperative tendency calculation model of node j to node i is as follows:

$$R_{ij} = \frac{N_{rh}(i,j)}{N_{rh}(i,j) + N_{rm}(i,j) \times N_{ps}} \tag{5}$$

$$S_{ij} = \frac{N_{sh}(i,j)}{N_{sh}(i,j) + N_{sm}(i,j) \times N_{ps}} \tag{6}$$

where $N_{rh}(i,j)$ and $N_{rm}(i,j)$ respectively represent the number of times of satisfaction and dissatisfaction when node i delivers packets to node j. $N_{sh}(i,j)$ and $N_{sm}(i,j)$ respectively indicate the number of times of satisfaction and dissatisfaction when node j delivers packets to node i. As can be seen from the above formula, the cooperative tendency assessment includes sending trust and receiving trust. The value range of R_{ij} and S_{ij} is [0, 1], and the initial value is set to 0.5. When the value is higher than 0.5, the node is regarded to be trustworthy, and the closer to 1, the higher the integrity. Similarly, the closer the value is to 0, the more anomalous of the node. At the same time, the formula introduces the penalty coefficient N_{ps}, so that the reputation value decreases faster than the rising rate.

Global cooperative tendency describes the comprehensive evaluation of satisfaction generated when a node interacts with any other node in the network. In this paper, the weighted average method is used to evaluate the global cooperation tendency of nodes. The receiving trust MR_i and sending trust MS_i of node i about global cooperation tendency are shown as follows:

$$MR_i = \frac{\sum_{j \in Gs_i} MS_j^w \times \left(1 - e^{-\frac{N_s(j,i)}{5}}\right) \times R_{ji}}{\sum_{j \in Gs_i} MS_j^w \times \left(1 - e^{-\frac{N_s(j,i)}{5}}\right)} \tag{7}$$

$$MS_i = \frac{\sum_{j \in Gr_i} MR_j^w \times \left(1 - e^{-\frac{N_r(j,i)}{5}}\right) \times S_{ji}}{\sum_{j \in Gr_i} MR_j^w \times \left(1 - e^{-\frac{N_r(j,i)}{5}}\right)} \tag{8}$$

where Gs_i represents the set of nodes to which node i has sent messages and Gr_i represents the set of nodes that send messages to node i. $N_r(j,i) = N_{rh}(j,i) + N_{rm}(j,i)$ and $N_s(j,i) = N_{sh}(j,i) + N_{sm}(j,i)$. w represents the weighting coefficient. When $w = 0$, the global cooperation tendency of the node is equal to the average value of the local cooperation tendency. Experiments have shown that when the number of anomalous nodes in the network is more than 35%, it is more appropriate to take $w \in [0, 0.5)$, otherwise taking $w \in (0.7, 1.2)$.

Formulas (7) and (8) are non-linear equations. Frequent computation will increase additional communication and resource overhead. Therefore, we divided time segments for iterative calculation.

$$MR_i(k+1) = \frac{\sum_{j\in Gs_i(k)} (MS(k))_j^w \times \left(1-e^{\frac{-N_s(j,i)}{5}}\right) \times R_{ji}}{\sum_{j\in Gs_i(k)} (MS(k))_j^w \times \left(1-e^{\frac{-N_s(j,i)}{5}}\right)} \tag{9}$$

$$MS_i(k+1) = \frac{\sum_{j\in Gr_i(k)} (MR(k))_j^w \times \left(1-e^{\frac{-N_r(j,i)}{5}}\right) \times S_{ji}}{\sum_{j\in Gr_i(k)} (MR(k))_j^w \times \left(1-e^{\frac{-N_r(j,i)}{5}}\right)} \tag{10}$$

When the network is initialized, the global cooperation tendency of each node is set to:$MR_i(0) = 0.5$, $MS_i(0) = 0.5$. It is obvious that the value range of the receiving trust and sending trust are [0, 1]. When the value of cooperative tendency of the node is closer to 0, we can consider that the node is anomalous. Therefore, in order to prevent the spread of the virus and packet loss caused by anomalous nodes, we defined threshold δ, which is set to 0.115 in the experimental environment of this paper. When the value of receiving trust or sending trust is less than the threshold δ, the node is marked as an anomalous node. Regardless of the strength of forwarding capability, the node does not participate in the transmission of messages.

The calculation of the $k + 1$ round is based on the calculation results of the k round. The sending trust and receiving trust are calculated at the beginning of the equal interval time slice. The value is calculated iteratively, which makes the algorithm have stronger convergence and stability.

2) CALCULATION OF ENCOUNTER STRENGTH

When selecting a relay node, the message-carrying node should consider not only the cooperation tendency, but also the encounter probability of nodes. As we all know, in opportunistic networks, most routing protocols are based on the assumption that once a link is established between two nodes, the data they need to transmit can be delivered completely and successfully. However, in the actual environment, because of the limited encounter duration, bandwidth and signal strength, not all nodes meeting each other can complete the data delivery. Therefore, if the average duration between nodes is longer, there will be a higher transmission success rate between the two nodes. Therefore, this paper proposes the concept of encounter strength by combining the encounter probability and the duration of encounter between nodes. As shown in Figure 2, the length of time for each encounter between a pair of nodes may be different, which directly affects the efficiency of transmission. According to the historical encounter sequence table of node pairs, we can calculate the total duration of node encounter $T_{total}(i,j)$ between node i and node j in the preparation phase.

$$T_{total}(i,j) = \sum_{k=1}^{n} T_k(i,j), \quad k \le n \tag{11}$$

Among them, $T_k(i,j)$ represents that the duration of the kth encounter between node i and node j. In order to avoid packet loss during delivery process and better measure the transmission performance of nodes, we calculate the average encounter duration $T_{avg}(i,j)$ between node i and node j.

$$T_{avg}(i,j) = \frac{\sum_{k=1}^{n} T_k(i,j)}{n}, \quad k \le n \tag{12}$$

where n represents the number of times that the two nodes have met from the beginning to the end of the preparation phase. When choosing relay nodes, the source node should prefer to deliver data to the nodes with the long average encounter duration of destination node, which can increase the success rate of message transmission.

The probability of encountering and the length of established connection time have great influence on the transmission efficiency of routing algorithms. Therefore, in this paper, the effects of encounter probability and duration on transmission efficiency are fully considered. Combining the advantages and characteristics of the two factors, the concept of encounter strength is proposed. Different from the probability of encounter, the value of the encounter strength can be greater than 1. When any node i and j in the network meet, the encounter strength $S(i,j)$ is updated as follows:

$$S(i,j) = S(i,j)_{old} + \left[\begin{array}{l}(1 - S(i,j)_{old}) \\ \times \left(S_{init} + \eta * min\left(\frac{T_{avg}(i,j)}{T_{\max}}, 1\right)\right)\end{array}\right] \tag{13}$$

$$S(i,j) = S(i,j)_{old} * \gamma^k \tag{14}$$

Among, $S(i,j)$ represents the encounter strength of node i and j after update, and $S(i,j)_{old}$ represents the encounter strength of node i and node j before update. S_{init} represents the initial encounter strength between two nodes in the network. At the initial point, all $S(i,j)$ are initialized to S_{init}. η is the weight factor, and the greater its value is, the greater the impact of encounter duration on encounter strength. $\gamma \in (0, 1)$ is the attenuation constant. The smaller the value is, the faster the numerical attenuation speed of the formula will be. k represents the number of unit time. $T_{\max} = 2 * CacheLength/Bandwidth$ represents the maximum effective duration of the encounter, where $CacheLength$ is the cache size of the node, and $Bandwidth$ is the bandwidth of the internet. $min\left(T_{avg}(i,j)/T_{\max}, 1\right)$ means taking the minimum value in the expression.

As we all know, the opportunistic network has the characteristics of intermittent communication and frequent changes in topology, so the forwarding capability value of the node changes with time. Similarly, the strength of the encounter strength between nodes also need to be updated with time during the preparation period. Different from the PRoPHET algorithm [32], the encounter strength takes into account the impact of the encounter duration and bandwidth of nodes on the message transmission.

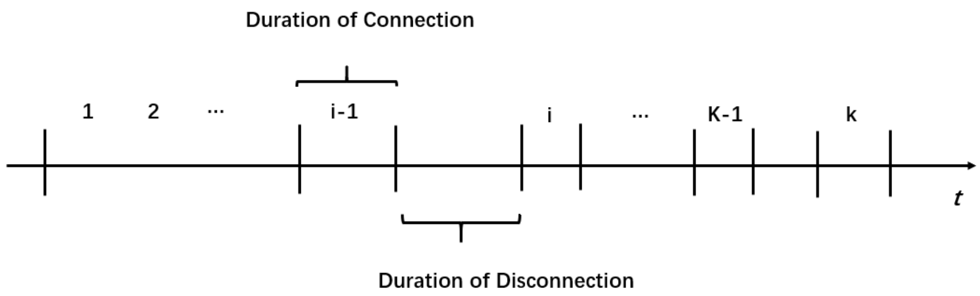

FIGURE 3. Transition of connection states between nodes.

During node initialization, the encounter strength of any pair of nodes is set as S_{init}. In this paper, the value of S_{init} is 0.60. When node i and j meet, the encounter strength is updated according to formula (13). When node i and j do not meet in k time units, the encounter strength is updated according to formula (14). Unlike the probability of encounter, the value of encounter strength can be greater than 1. The node pair with greater encounter strength indicates that the two nodes are closely related.

In order to smooth the encounter strength curve between nodes and prevent route jitter, we use the average encounter strength of nodes in the preparation phase as the basis for nodes to deliver data. The calculation formula for the encounter strength of node i and node j at the kth encounter is:

$$S_k(i,j) = S_{k-1}(i,j) + (1 - S_{k-1}(i,j)) \times (S_{init} + \eta * E(i,j)) \quad (15)$$

Among,

$$E(i,j) = min\left(\frac{T_{avg}(i,j)}{T_{\max}}, 1\right) \quad (16)$$

When the two nodes meet for the first time, the average encounter strength between the nodes is S_{init}. When two nodes meet for the second time, the average encounter strength of the nodes is:

$$S_{avg(2)}(i,j) = \frac{S_1(i,j) \times t_1 + S_2(i,j) \times t_2}{t_1 + t_2} \quad (17)$$

where t_1 represents the time when node i and node j meet for the first time, and t_2 represents the interval between the second encounter and the first encounter.

At the kth encounter, the average encounter strength of node i and node j is:

$$S_{avg(k)}(i,j) = \frac{S_{avg(k-1)}(i,j) \times \sum_{i=1}^{k-1} t_i + S_k(i,j) \times t_k}{\sum_{i=1}^{k} t_i} \quad (18)$$

Among, t_i represents the interval between two nodes meeting at the ith and $(i-1)th$ times.

The characteristics of intermittent communication between nodes determine that not every connection can successfully complete the transmission of messages, so the duration of the encounter should also be taken into account in the routing algorithm. At the same time, We obtain the average value of the encounter strength of nodes in the preparation stage, which can effectively prevent the jitter of the curve.

3) CALCULATION OF FORWARDING CAPABILITY

Based on the above analysis of cooperative tendency and encounter strength of nodes, we can use these indexes to measure the forwarding capability of nodes. The forwarding capability of node i is calculated as follows:

$$\kappa_i = W_1 * MR_i + W_2 * MS_i + W_3 * PD_i \quad (19)$$

where MR_i and MS_i respectively represent the cooperative tendency of receiving trust and sending trust of node i. PD_i represents the average value of encounter strength between node i and destination node in the preparation stage. W is the weight of the three variables, and $W_1 + W_2 + W_3 = 1$. The weight can be solved according to the method of information entropy. The experiment have shown that when the malicious nodes in the network are less than 30%, the value of W_3 is greater than the sum of W_1 and W_2, then the algorithm has a better performance. In particular, if a node is marked as an anomalous node, it will not participate in the delivery of messages, regardless of the encounter strength of the node.

C. MATRIX DECOMPOSITION AND MATRIX COMPLETION

In the preparation phase, nodes collaborate to obtain information about the network. If the length of preparation time is too long, it will increase the network load and huge amounts of data will occupy the excessive cache of nodes. If the time is too short, the information collected by the nodes will not be sufficient to provide adequate theoretical support for the algorithm, resulting in reduction of transmission efficiency of the algorithm. Obviously, not all nodes can collect enough information to evaluate the forwarding capability of each node. Therefore, in order to save network cost and find out all the key nodes during the process of message transmission, this paper uses the matrix decomposition method to decompose the forwarding capability matrix of nodes into three matrices, and then uses the decomposed matrices to predict the missing values of the original matrix.

First, we define the forwarding capability matrix between nodes in the network.

$$F = \begin{bmatrix} -1 & f(1,2) & f(1,3) & \cdots & f(1,n) \\ f(2,1) & -1 & f(2,3) & \cdots & f(2,n) \\ f(3,1) & f(3,2) & -1 & \cdots & f(3,n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f(n,1) & f(n,2) & f(n,3) & \cdots & -1 \end{bmatrix} \tag{20}$$

where $f(i,j)$ represents the forwarding capability of node i to transmit messages to node j. The forwarding capability of a node to forward messages to itself is marked as -1, which is not regarded as the original data in the training process. The matrix of forwarding capability F between nodes is a symmetric matrix with the diagonal value of -1.

In the actual application scenario, the moving trajectory of nodes has the characteristics of periodicity and regularity. Nodes with close social relationships tend to have more frequent connections. In addition, some nodes are active, moving frequently, and some nodes are dull, moving slowly. Therefore, we decompose the original matrix into the form of multiplying two low-rank matrices and adding them to the bias matrix, as shown in formula (21). The calculation method of each element in the original matrix is shown in formula (22).

$$F_{n\times n} = B_{n\times n} + P^T_{n\times m} Q_{m\times n} \tag{21}$$

$$f'_{ij} = b_{ij} + p_i^T q_j = \mu + b_i + b_j + p_i^T q_j \tag{22}$$

where $P^T_{n\times m}$ and $Q_{m\times n}$ are eigenvector matrixes. $B_{n\times n}$is a biased matrix, which is composed of three bias vectors. $p_i^T q_j = \sum_k^F p_{ik}^T q_{jk}$, in which p_{ik}^T and q_{jk} respectively represent the kth eigenvalue in the feature vector. Each element f_{ij} in forwarding capability matrix $F_{n\times n}$ can be decomposed into a form f'_{ij}, in which vector p_i^T is multiplied by vector q_j and added to three offsets. b_i is the offset of forwarding capability of node i, representing the tone of a node's forwarding characteristics. b_j represents forwarding preferences of node j. μ represents the global average value of effective data in the training set, indicating the influence of the network's own attribute parameters on the forwarding capability of nodes. f'_{ij} is a value derived from biases and two vectors.

As we all know, the movement of nodes has preferences. Nodes with stronger social relationships will have more frequent connections and more chances of encounter. The mobility characteristics of a node affect the probabilities of meeting other nodes, which in turn influences the forwarding capability of the node. Therefore, we use the node's preference b_{ij} as an indicator to correct the forwarding capability between nodes.

If we find the suitable matrixes of P, Q and B to minimize the training set error. It is also possible to minimize the prediction error of the forecast set. So we define the loss function as:

$$E(p,q,b_i,b_j) = \min_{p^*,q^*,b^*} \sum_{i,j}^{n} (f_{ij}-\mu-b_i-b_j-p_i^T q_j)^2 \tag{23}$$

If the loss function above is directly optimized, it will lead to learning overfitting. In order to prevent the loss function from overfitting in the process of iterative solution, the regularization factor $\lambda(\left\|p_i^T\right\|^2+\left\|q_j\right\|^2+b_i^2+b_j^2)$ is added, in which λ is the regularization parameter that needs to be obtained through trial and error.

$$E(p,q,b_i,b_j) = \min_{p^*,q^*,b^*} \sum_{i,j}^{n} (f_{ij}-\mu-b_i-b_j-p_i^T q_j)^2 + \lambda(\left\|p_i^T\right\|^2+\left\|q_j\right\|^2+b_i^2+b_j^2) \tag{24}$$

In order to minimize the loss function, the gradient descent method can be used, in which the direction of rapid descent can be obtained by taking the partial derivative of parameters. Unlike ordinary matrix decomposition methods, this method has two bias vectors. Partial derivatives of the four parameters can be obtained:

$$\begin{aligned} \partial E/\partial p_i^T &= -2q_j \cdot e_{ij} + 2\lambda p_i \\ \partial E/\partial q_j &= -2p_i \cdot e_{ij} + 2\lambda q_j \\ \partial E/\partial b_i &= -2e_{ij} + 2\lambda b_i \\ \partial E/\partial b_j &= -2e_{ij} + 2\lambda b_j \\ e_{ij} &= f_{ij} - f'_{ij} \end{aligned} \tag{25}$$

According to equations (24) and (25), the iterative update rules for the four parameters can be obtained by using the stochastic gradient descent as follows:

$$\begin{aligned} p_i^T &:= p_i + \alpha[q_j(f_{ij}-\mu-b_i-b_j-p_i^T q_j) - \lambda p_i^T] \\ q_j &:= q_j + \alpha[p_i(f_{ij}-\mu-b_i-b_j-p_i^T q_j) - \lambda q_j] \\ b_i &:= b_i + \alpha(f_{ij}-\mu-b_i-b_j-p_i^T q_j - \lambda b_i) \\ b_j &:= b_j + \alpha(f_{ij}-\mu-b_i-b_j-p_i^T q_j - \lambda b_j) \end{aligned} \tag{26}$$

where $\alpha = \alpha_0/(1+decay\_rate * t)$ represents the learning rate of four parameters, in which $decay\_rate$ is the super-parameter and t represents the number of iterations of gradient descent with initial learning rate of α_0.

Through continuous iteration, we can finally obtain the offset b_i, b_j and decomposition matrix P, Q. For any missing element in the original matrix F, it can be deduced by formula (22). In particular, for ease of understanding, the detailed steps of matrix decomposition are given below.

Step 1: Initialize the matrix P and Q. Make the values of vectors p_i^T and q_j be any number between 0 and 0.1 and initialize the values of the bias terms b_i and b_j to 0. Calculate the mean value μ of the effective values in the original matrix $F_{n\times n}$.

Step 2: According to the derivation formula (22), the evaluation value f'_{ij} of node i's forwarding capability to node j can be obtained. Then based on the formula $e_{ij} = f_{ij} - f'_{ij}$, calculating the error value between the evaluation value f'_{ij} and the actual value f_{ij}.

Step 3: The gradient descent method is used to minimize the loss function. First, the partial derivatives of p_i^T ,q_j, b_i and b_j are calculated respectively, and obtain the corresponding gradient vector. Then update the gradient vector according to the recursive formula (26), The characteristic variable will be updated along the gradient direction.

Step 4: According to the updated feature parameters, step (2) is executed to calculate the sum of the error values sl in the training set. Calculate the value avg by sl/num where num is the number of valid data in the training set. If avg is less than threshold t, the training is stopped, otherwise step (3) is continued. In this paper, t takes the value of 0.1.

Step 5: The original matrix can be represented by the final trained characteristic parameters, and the missing values in the original matrix can be evaluated by formula (22).

Matrix decomposition is a widely used method in machine learning, which can not only reduce the dimension of the original matrix, but also predict the missing values in the matrix. In this paper, the information collected by the nodes is limited. In order to find out all the key nodes in the transmission process, we adopt the matrix decomposition method to predict the missing values in the matrix. In addition, in a transmission cycle, each node only needs to establish a node forwarding capability matrix as the basis for data delivery, which can reduce the computational complexity and improve the transmission efficiency of the algorithm.

D. COMPLEXITY ANALYSIS OF THE ALGORITHM

On the whole, we propose a transmission prediction mechanism exploiting comprehensive forwarding capability of nodes in opportunistic networks. The algorithm is divided into two stages: the preparation stage and the transmission stage. In the preparation stage, nodes continuously collect and update information in the network by cooperating with each other. In the stage of message transmission, the optimal relay nodes are recommended by quantifying the forwarding capability of nodes. In addition, in order to find out all the key nodes in the transmission process, we use the matrix decomposition method to evaluate the missing values in the forwarding capability matrix. What's more, the proposed strategy is based on a multi-copy routing strategy. In order to control the number of message copies we employ positive transmission method to deliver data, which can make the data spread in the direction of increasing forwarding capability of nodes. For enhancing the understandability of the algorithm, the specific steps of message transmission are as follows:

1) In the preparation phase, each node creates an archive file to collect information about the network and constructs a list of information sequences from the collected information. At the same time, each node shares the list to the nodes it encounters.
2) The node carrying the message comprehensively evaluates the forwarding capability of nodes in the network based on the list of information sequences, and then establishes the forwarding capability matrix. The measurement of forwarding capability is composed of cooperation tendency and encounter strength, which can be used to effectively identify anomalous nodes and obtain the optimal recommendation of relay nodes.
3) Obviously, not all nodes can collect enough information to evaluate the forwarding capability of each node. Therefore, in order to save network cost and find out all the key nodes during the process of message transmission, we use the method of matrix decomposition to predict the missing values in the forwarding capability matrix.
4) The node carrying the message finds the appropriate relay node from its neighbors according to the forwarding capability matrix. When the node has a new transmission task within the same transmission cycle, the node does not need to re-establish a new forwarding capability matrix, and the matrix created for the first time can still be used. At the same time, in order to control the number of copies, we deliver data in the direction of increasing forward capability of nodes.

Algorithm 1 is established to detail the process of the algorithm. In particular, during the preparation stage, nodes achieve the purpose of collecting network information by sharing state sequence list with each other. The time complexity of this phase is $O(\log_2^n)$. Based on the collected information, the source node calculates the cooperation tendency and the strength of encounter between nodes in the network. In this process, the time complexity is $O(n)$. Finally, according to the strategy of matrix decomposition, the forwarding capability matrix of nodes is completed. During a transmission cycle, each node only needs to establish a forwarding capability matrix. When there are multiple transmission tasks, the matrix created for the first time can still be used, making the result that it does not require repeated calculations in the subsequent transmission tasks. So the time complexity of this stage is $O(n)$. In short, after strict mathematical verification and analysis, the time complexity of our proposed TPMEC algorithm is $O\left(\log_2^n + n + n\right) = O(n)$. The time complexity of Spray and Wait algorithm is $O(n\log_2^n)$, while that in ETNS algorithm is $O(n)$.

IV. SIMULATION AND ANALYSIS

In the experiment of this paper, the simulation tool of Opportunistic Network Environment (ONE) is used to evaluate the performance of the algorithm. In order to comprehensively measure the performance of the strategy, TPMEC algorithm will be compared and analyzed with the four algorithms: ITPCM algorithm [33] (Information Transmission Probability and Cache Management Method), ETNS algorithm [20] (Effective Data Transmission Strategy Based on Node Socialization), PRoPHET algorithm [32] and Spray and Wait algorithm [34]. PRoPHET and Spray and Wait algorithms are typical and traditional methods, which are very representative in routing algorithms. The other two algorithms are the latest routing algorithms in opportunistic networks. Among them, PRoPHET is a routing transmission algorithm based on the probability of encounter between nodes and Spray and Wait

Algorithm 1 A Transmission Prediction Mechanism Exploiting Comprehensive Node Forwarding Capability

Input: A graph $G(V, E)$, source node S and destination node D;
Output: optimal path for data delivery;
Create a network data archive for each node
Collect information about the network
Initialize the threshold δ for anomalous nodes
if $S.isMessageCarrier()$ **then**
  for *each node* $i \in$ collection of nodes V **do**
    Calculate the forwarding trust degree MS
    Calculate the receiving trust degree MR
    if $MR < \delta$ or $MS < \delta$ **then**
      Mark the node as an anomalous node
    end if
    Evaluate the encounter strength value between the node and the destination node according the information collected in preparation stage.
  end for
end if
Establish the forwarding capability matrix
Decompose the original matrix according to the method of matrix decomposition
for *each missing value* v in forwarding capability matrix **do**
  Predicting the forwarding capability value using feature parameters
end for
for *each node* $j \in$ neighbor nodes of S **do**
  if $j.Ability() \geq S.Ability()$ and $j \notin$ abnormal node **then**
    S sends data packets to j
  end if
end for

is an algorithm composed of two stages: Spray stage and Wait stage. ITPCM is an algorithm based on node transmission probability and routing cache management method, and ETNS is a routing transmission algorithm using the social relationships and similarities of nodes.

According to the actual application environment, the map of Helsinki, capital of Finland, is used as the simulation scene, and the simulation area is 4500*3400m. Different from other simulation environments, the simulation map has a reasonable layout of parks, schools, shopping malls, housing, etc. They can simulate real environments and form close-knit online communities. The specific experimental parameters of this simulation are as follows: the number of nodes is 500-1300, and the movement characteristics of nodes are based on the way defined by the HCMM [35] (home-cell community-based mobility model). In order to compare the performance effects under different mobile models, the HCMM is compared with the mobile models of random work, random waypoint, and gauss-markov in the same environment. The simulation time of the experiment is 12h. The generation interval of the message is 35-60s. The node moves at a speed of 1-5m/s and the cache size is 10-40Mb. The transmission mode is broadcast and the maximum length of node transmission is 10m. The transmission speed between nodes is 0.2-1.0Mb/s, and the transmission speed varies in different environments.

In this experiment, TPMEC algorithm will be compared with PRoPHET, Spray and Wait, ITPCM and ETNS algorithms in the same experimental environment. The indicators we pay attention to in the experiment are as follows:

1.Delivery ratio: The ratio of the messages generated by source nodes are successfully received by destination nodes, which can be used to measure the effectiveness of the routing algorithm. It can be calculated as $D_{node} = (Num_r/Num_c) \times 100\%$, where Num_c represents the number of messages generated and Num_r represents the number of messages received by destination node.

2.Average end-to-end delay: Average time consumed by receiving messages from source nodes to destination nodes, including the sending time and waiting time. It can be calculated by the formula: $E_{del} = E_{sum}/S_{node}$, where E_{sum} represents the total delay of network transmission, and S_{node} represents the number of destination nodes that successfully receive messages.

3.Routing overhead: The indicator represents the network overhead consumed when messages are successfully transmitted between nodes and is used to measure the power consumption and efficiency of transmission protocols. Its calculation formula is $R_{over} = (T_{total\_time} - T_{succeed\_time})/T_{total\_time}$, where T_{total_time} represents the total time of packet delivery, and $T_{succeed\_time}$ represents the time of successful packet delivery.

A. ANALYSIS OF EXPERIMENTAL RESULTS

1) INFLUENCE OF PREPARATION TIME ON TPMEC ALGORITHM

This chapter mainly discusses the impact of preparation time on TPMEC algorithm in terms of delivery ratio and average end-to-end delay. In the preparation phase, nodes collaborate to collect information in the network. Each node in the network has an archive file that records information about encounters with other nodes. If the length of preparation time is too short, the node has not yet obtained adequate network information. If the length of preparation time is too long, the archive file will take up a large node cache. Therefore, proper preparation time is critical to the transmission performance of the algorithm. According to the experimental results, when the simulation time is 12h and the preparation phase is set to 80-120min, TPMEC algorithm has the best performance in terms of transmission performance.

Firstly, we compare the influence of preparation time on delivery ratio of the algorithm. As shown in fig.4, the broken line keeps rising with the passage of time, and then slowly declines after rising to a certain stage. This is because

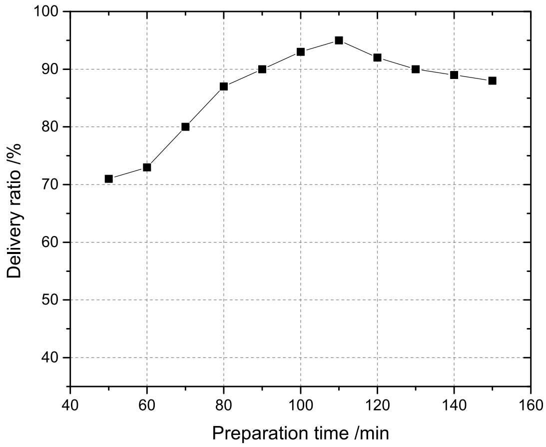

FIGURE 4. Packet delivery ratio.

the message transmission of TPMEC algorithm is based on valid information about the network acquired by nodes in the preparation phase. At the beginning, nodes don't obtain sufficient and reliable network information, resulting in low success rate of data delivery. With the passage of time, the network information obtained by the nodes becomes more and more perfect, and the success rate of data delivery becomes higher. However, too long preparation time will reduce the transmission success rate of the algorithm. This is because the information in the network collected and updated by a node for a long time will consume a large amount of cache and energy of the node, resulting in that the node cannot carry more copies of messages in the process of message transmission. In addition, when a node needs to select a suitable relay node, a large amount of network information will increase the computing time and complexity of the algorithm, which affects the transmission efficiency of messages.

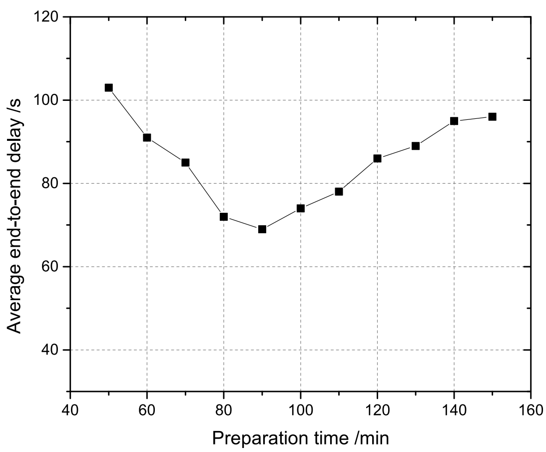

FIGURE 5. Average end-to-end delay.

Next, we compare the effect of preparation time on TPMEC algorithm in terms of average end-to-end delay. As shown in figure 5, the broken line decreases with the passage of time, and when it reaches a certain value, the broken line increases slowly. As described in the previous paragraph, nodes need to collect sufficient and valid network information to make the best decision during message transmission. Thus, at the beginning, the end-to-end average delay of data delivery is decreasing over time. However, when the preparation time is too long, the average end-to-end delay increases slowly. This is because the nodes hold too much network information, which takes up a large number of node cache and increases the computational complexity. Therefore, the excessive preparation phase will also reduce the performance of the algorithm.

TABLE 1. Simulation parameter settings.

Simulation Environment	Description
Simulator	The ONE
Simulated time $/h$	12
Communication area $/m^2$	4500m*3400m
Number of nodes	100-1300
Node movement speed $/(m \cdot s^{-1})$	1-5
Node cache size $/Mb$	10-40
Maximum transmission distance $/m$	10
Transmission mode	Broadcast
Transmission speed $/(Mb \cdot s^{-1})$	0.2-1

2) INFLUENCE OF NODE CACHE ON ROUTING ALGORITHMS

This section examines the impact of node cache size on the performance of five algorithms. In the process of data delivery, messages need to occupy the cache space and consume energy of nodes. However, the cache size of a node is limited, and the cache size directly affects the efficiency of routing algorithms. Therefore, we compare the performance of the five algorithms in terms of message transmission by setting different node cache sizes.

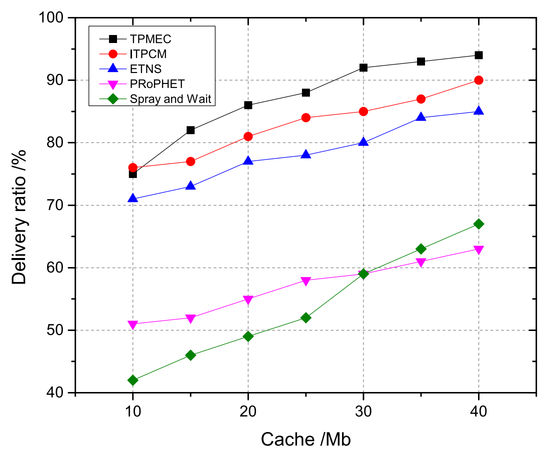

FIGURE 6. Packet delivery ratio.

Firstly, we compare the impact of the cache size of nodes on the delivery ratio. As shown in figure 6, with the increase of node cache, the delivery ratio of the five algorithms also increases correspondingly. Among them, the Spray and Wait algorithm always has the lowest transmission rate. This is because that the algorithm blindly spreads messages, which will produce a large number of copies and consume a lot of cache space in the spray stage. It has the most serious dependence on the cache size of nodes. However, the ITPCM algorithm has a cache management mechanism, which can reduce the dependency on the size of the node cache. It is obvious that TPMEC algorithm always has the highest transmission success rate, because the algorithm only collects the network information in the preparation stage, and does not require repeated collections during the same transmission cycle, so it has a small dependence on the cache size of nodes.

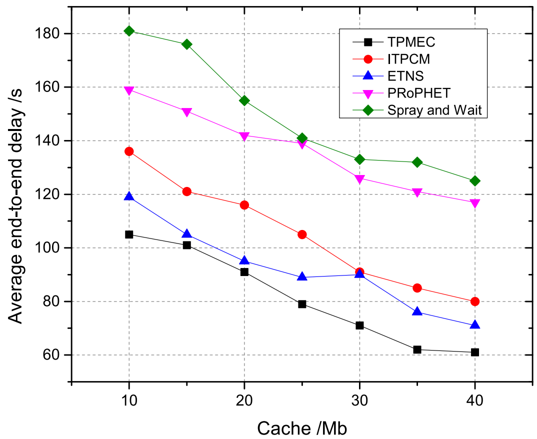

FIGURE 7. **Average end-to-end delay.**

Next, we compare the impact of the cache size of nodes on the average end-to-end delay of the five algorithms. As shown in figure 7, the broken line in the figure decreases in different range as the node cache increases. Since the Spary and Wait and the PRoPHET algorithms do not take into account the context information and the duration of encounter between nodes, the average end-to-end delay of the two algorithms is relatively large. ITPCM and ETNS algorithms take into account the context information and social attributes of nodes, so their performance is better. TPMEC algorithm measures the forwarding capability of nodes comprehensively based on the cooperative tendency and the encounter strength of nodes. Therefore, the performance of the algorithm is optimal.

Finally, the impact of node cache size on the routing overhead is shown in figure 8. Spray and Wait algorithm relies heavily on the cache size of nodes and blindly spread messages in the form of spray, so when the cache space is small, its transmission performance is relatively low. ITPCM algorithm ignores the fact that information with low priority may be blocked, which will cause the source node generates more message copies and increase the routing overhead during data delivery. TPMEC algorithm adopts the positive transmission strategy, so the routing overhead is minimal.

Figure 10 and figure 11 are box charts, which can clearly show the scope and data distribution of local or all data. These two figures respectively show the changes in the delivery ratio and routing overhead of the five algorithms when the node cache varies. It can be seen from the figures that the Spray and Wait algorithm is the most sensitive to the cache size of nodes,

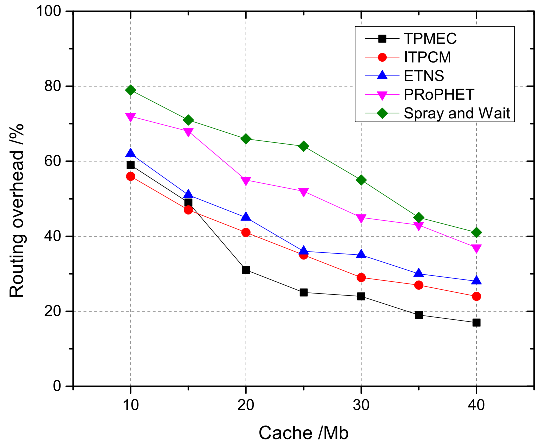

FIGURE 8. **Routing overhead.**

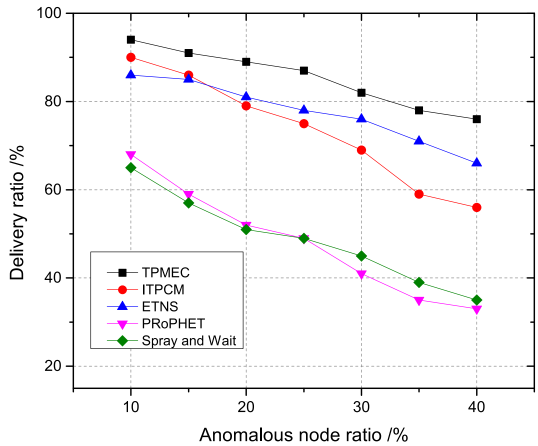

FIGURE 9. **Delivery ratio.**

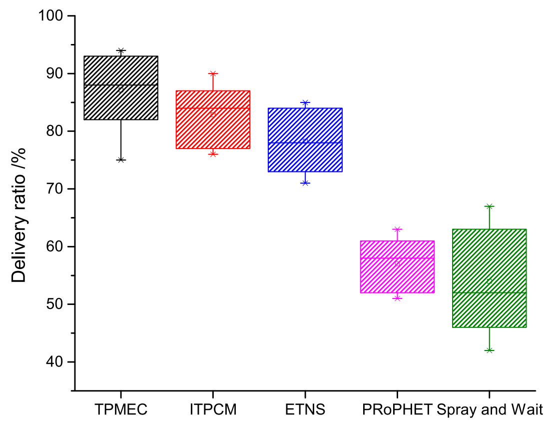

FIGURE 10. **Packet delivery ratio.**

because this algorithm relies on the spray method for message transmission, which requires a large amount of node cache size and increases the routing overhead. TPMEC algorithm only collects information in the preparation stage and finds the next hop by comprehensively evaluating the forwarding capability of nodes, which occupies a small cache space, so its performance is optimal.

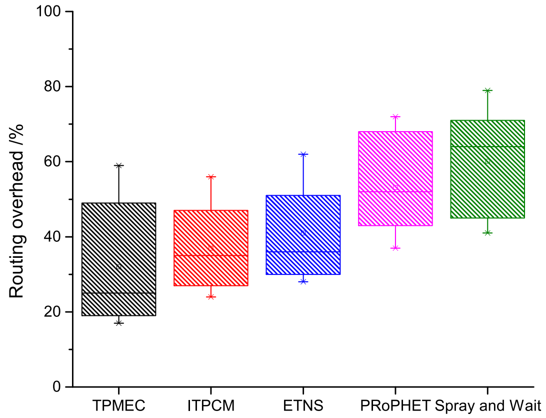

FIGURE 11. Routing overhead.

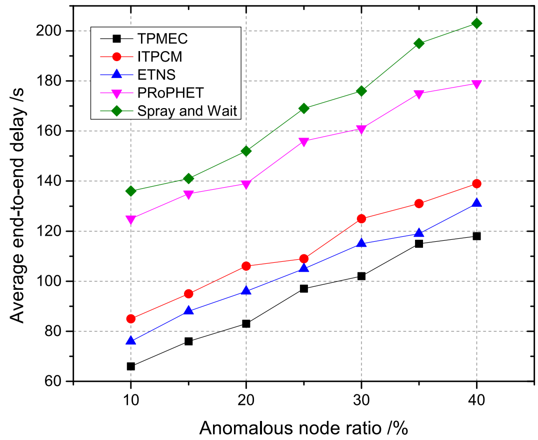

FIGURE 12. Average end-to-end delay.

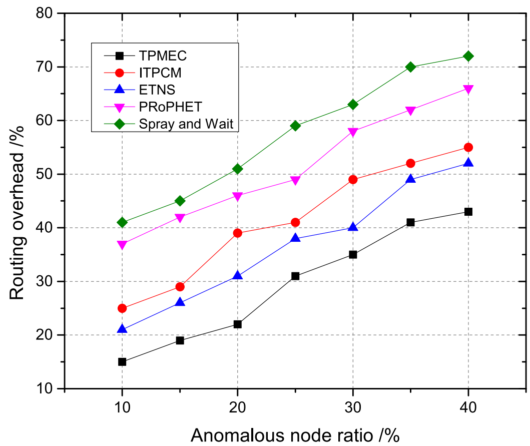

FIGURE 13. Routing overhead.

3) INFLUENCE OF ANOMALOUS NODES ON ROUTING ALGORITHMS

In this chapter, we will discuss the influence of the ratio of anomalous nodes on the transmission efficiency of the five algorithms. It is well known that nodes have limited resources, including storage and energy. Not all nodes are willing to help other nodes store and forward messages. In addition, viruses in the network will also cause harm to the network security, so effective screening of anomalous nodes is very important for the performance of message transmission.

Firstly, we compare the influence of the anomalous node ratio on the delivery ratio. As is shown in figure 9, The Spray and Wait and PRoPHET algorithms do not take any effective measures to identify anomalous nodes, so when the ratio of anomalous nodes is high, the delivery ratio of the two algorithms is relatively low. However, ITPCM and ETNS algorithms formulate forwarding strategies according to the context information and social relations of nodes. As a result, their transmission rates are relatively high. TPMEC algorithm adopts the cooperative tendency to measure the node's receiving and forwarding trust degree, which can avoid sending messages to anomalous nodes, so the transmission success rate is the highest.

Next, the influence of the ratio of anomalous nodes on the average end-to-end delay of the five algorithms is shown in figure 12. The Spray and Wait and PRoPHET algorithms will inevitably deliver the message to anomalous nodes, resulting in a significantly larger average end-to-end average delay. The Spray and Wait algorithm is based on the multi-copy transmission mode, so when nodes have enough cache space, the average end-to-end delay of the algorithm is better than that of PRoPHET. TPMEC algorithm can effectively identify anomalous nodes and reduce the average end-to-end delay by comprehensively evaluating the trust degree of nodes based on the collected information in the preparation phase.

Finally, the influence of the ratio of anomalous nodes on the routing overhead is shown in figure 13. TPMEC algorithm calculates the cooperative tendency index based on the information collected in the preparation phase, which can effectively avoid the loss of data packets, so the routing overhead of the algorithm is the lowest. The other four algorithms do not take effective measures to identify the trust degree and cooperation tendency of nodes, which will inevitably deliver packets to the anomalous nodes, causing the loss of data packets and increasing the routing overhead the network.

Figure 14 and figure 15 respectively show the performance of the five algorithms under different anomalous node ratios. PRoPHET and Spray and Wait algorithms are sensitive to the anomalous nodes. This is because these two algorithms do not take effective methods to identify the anomalous nodes. ITPCM and ETNS algorithms can relatively avoid data packets being delivered to anomalous nodes through the social attribute of nodes. There is no doubt that the TPMEC algorithm performs best in the experiment. This is because that the algorithm adopts an effective mechanism to identify anomalous nodes, which can effectively improve the efficiency of message transmission.

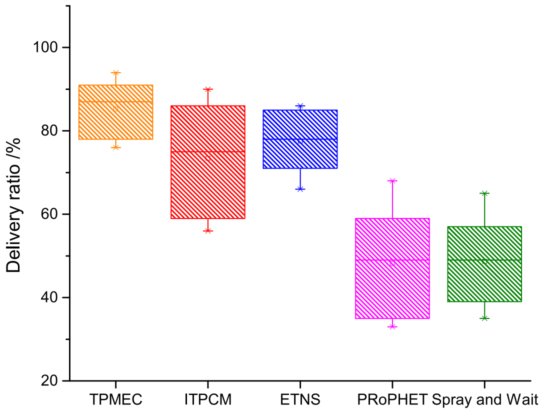

FIGURE 14. Packet delivery ratio.

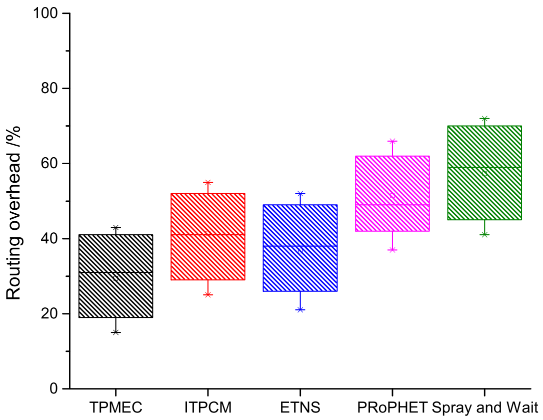

FIGURE 15. Routing overhead.

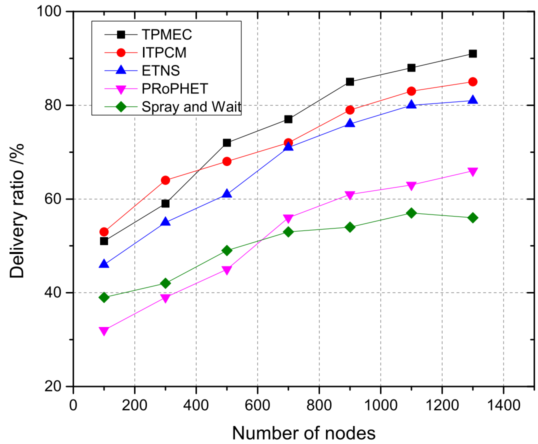

FIGURE 16. Packet delivery ratio.

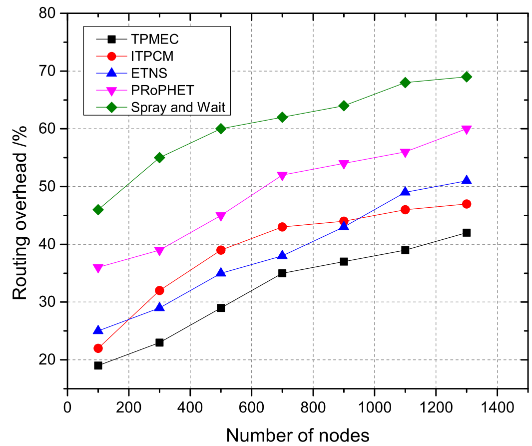

FIGURE 17. Routing overhead.

4) INFLUENCE OF NODE NUMBER ON ROUTING ALGORITHMS

In this section, we discuss the impact of the number of nodes on the algorithms in terms of delivery ratio and routing overhead.

Figure 16 shows the influence of the number of nodes on the delivery ratio in the five algorithms. As can be seen from the figure, Spray and Wait algorithm always has the lowest delivery ratio, because the algorithm blindly delivers messages and produces plenty of copies during the Spray phase. When the number of nodes increases, a mass of copies will heighten the load on the network, which affects the efficiency of algorithm transmission. PRoPHET algorithm transmits messages based on the encounter probability of nodes, so when the number of nodes increases, its performance is higher than the Spray and Wait algorithm. ITPCM and ETNS algorithms adopt the social and mobile characteristics of the nodes to select the appropriate relay nodes purposefully. TPMEC algorithm comprehensively measures the forwarding ability of nodes based on their cooperative tendency and meeting strength, so the algorithm has the highest success rate of message transmission.

Figure 17 shows the influence of the number of nodes on the routing overhead in the algorithms. Spray and Wait algorithm is based on a flooding spread strategy that takes up a large amount of cache space of nodes in the network. PRoPHET algorithm selects the next hop based on the probability of encounter, and its transmission load is lower than the Spray and Wait algorithm. ITPCM algorithm has a specific cache management strategy, and when the number of nodes increases, the network load does not increase significantly. In order to control the number of copies, TPMEC adopts the positive transmission method to deliver data, which can make the data spread in the direction of increasing forwarding capability of nodes. In the same transmission cycle, the forwarding capacity of each node is only calculated once, so TPMEC algorithm has the optimal performance.

5) INFLUENCE OF MOBILE MODELS ON ROUTING ALGORITHMS

We discuss and analyze the performance of the algorithm in different mobile models. In the simulation, the mobile models of random walk (RW), random waypoint (RWP), Gauss-Markov (GM), and HCMM (home-cell community-based mobility model) are respectively adopted to analyze the message transmission efficiency.

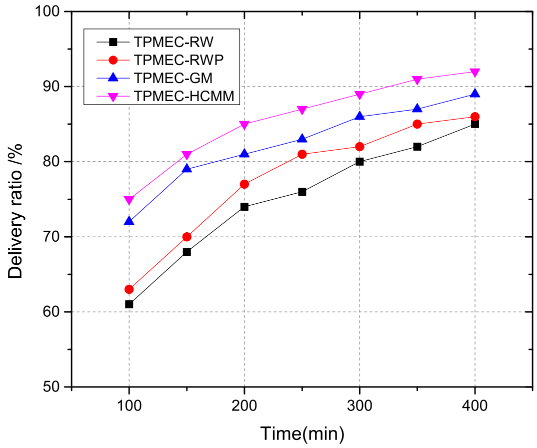

FIGURE 18. Packet delivery ratio.

Figure 18 shows the performance of packet delivery ratio in different mobile models. As can be seen, the curve rises slowly over time and finally tends to be stable. This is because during the transmission phase of the algorithm, the node needs to calculate the information collected during the preparation phase, which consumes more time and resources. In the RM mobile model, the algorithm has the lowest delivery ratio. Because this moving model has no memory and no movement rule, which makes a small accuracy in the prediction of node forwarding ability. The delivery ratio of RWP is higher than that of RW, but both RWP and RW adopt a simple moving mode in random direction, which limits the application of these models to some extent. The GM motion model has a correlation of velocity and direction, in which the movement trajectory of the nodes is smooth. In the HCMM mobile model, it has the highest transmission success rate, because the mobile model considers the mobile periodicity and social attributes of nodes, which is more consistent with the application scenarios of this algorithm.

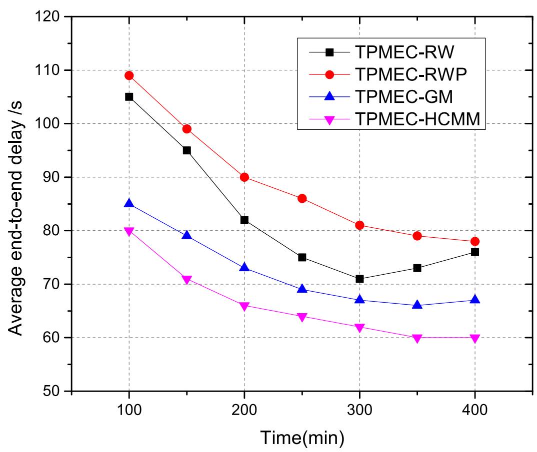

FIGURE 19. Average end-to-end delay.

Figure 19 shows the performance of different mobile models in the average end-to-end delay. As can be seen from the figure, the curve has a relatively large rate of decline at the beginning, and then gradually flattens out. This is because the computing power and cache of nodes are limited. RW and RWP have high transmission delay, because in these two modes, the moving direction and speed of nodes are random and irregular. GM mobile model has a correlation between direction and speed at any time, but the movement of nodes is largely not a straight line. HCMM mobile model has the best performance, because the model matches the moving pattern of the node in the actual situation. In the proposed strategy, the theory of matrix decomposition is used to analyze the transmission preference of nodes, which is more suitable for the mobile model.

V. CONCLUSION

This paper proposes a transmission prediction mechanism to explore the comprehensive forwarding capability of nodes in opportunistic networks. The strategy is divided into two phases: the preparation phase and the transmission phase. In the preparation phase, each node establishes an archive file to collect network information, and then updates the status sequence list through cooperation. In the transmission phase, the algorithm quantifies the forwarding capability of nodes by the cooperation tendency and the encounter strength, which can avoid deliver packets to the anomalous nodes and improve the success rate of data delivery.

At the same time, it is difficult for the source node to evaluate the forwarding capability of all nodes through the information collected. Therefore, in order to find out all the key nodes in the transmission process, we use matrix decomposition method to predict and supplement in the missing value of forwarding capability, which can improve the efficiency of message transmission. In future work, we will continue to improve our algorithm according to the actual application scenarios, and improve the privacy and security of messages in the process of transmission. In future works, we will continue to optimize the proposed algorithm based on the actual application environment and improve the privacy and security of information in the transmission process.

REFERENCES

[1] J. Dede, A. Förster, E. Hernández-Orallo, J. Herrera-Tapia, K. Kuladinithi, V. Kuppusamy, P. Manzoni, A. bin Muslim, A. Udugama, and Z. Vatandas, "Simulating opportunistic networks: Survey and future directions," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 1547–1573, 2nd Quart., 2018.

[2] A. Roy and T. Deb, "Performance comparison of routing protocols in mobile ad hoc networks," *Int. J. Eng. Sci. Technol.*, vol. 2, p. 279, 2018.

[3] N. Li, J.-F. Martínez-Ortega, and V. H. Díaz, "Cross-layer and reliable opportunistic routing algorithm for mobile ad hoc networks," *IEEE Sensors J.*, vol. 18, no. 13, pp. 5595–5609, Jul. 2018.

[4] Y. Wang, Y. Liu, J. Zhang, H. Ye, and Z. Tan, "Cooperative store–carry–forward scheme for intermittently connected vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 777–784, Jan. 2017.

[5] S. Trifunovic, S. T. Kouyoumdjieva, B. Distl, L. Pajevic, G. Karlsson, and B. Plattner, "A decade of research in opportunistic networks: Challenges, relevance, and future directions," *IEEE Commun. Mag.*, vol. 55, no. 1, pp. 168–173, Jan. 2017.

[6] Y. Yan, Z. Chen, J. Wu, and L. Wang, "An effective data transmission algorithm based on social relationships in opportunistic mobile social networks," *Algorithms*, vol. 11, no. 8, p. 125, 2018.

[7] D. Tuncer, V. Sourlas, M. Charalambides, M. Claeys, J. Famaey, G. Pavlou, and F. De Turck, "Scalable cache management for ISP-operated content delivery services," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 8, pp. 2063–2076, Aug. 2016.

[8] J. Wu and Z. Chen, "Human activity optimal cooperation objects selection routing scheme in opportunistic networks communication," *Wireless Pers. Commun.*, vol. 95, pp. 3357–3375, Aug. 2017.

[9] Z. H. Kumiawan, L. V. Yovita, and T. A. Wibowo, "Performance analysis of dLife routing in a delay tolerant networks," in *Proc. Int. Conf. Control, Electron., Renew. Energy Commun.*, Bandung, Indonesia, Sep. 2017, pp. 41–46.

[10] V. V. Neena and V. M. A. Rajam, "Performance analysis of epidemic routing protocol for opportunistic networks in different mobility patterns," in *Proc. Int. Conf. Comput. Commun. Inform.*, Coimbatore, India, Jan. 2013, pp. 1–5.

[11] S. J. Borah, S. K. Dhurandher, S. Tibarewala, I. Woungang, and M. S. Obaidat, "Energy-efficient prophet-prowait-EDR protocols for opportunistic networks," in *Proc. Global Commun. Conf. (GLOBECOM)*, Singapore, Dec. 2017, pp. 1–6.

[12] M. Y. Mir, C.-L. Hu, and S.-Z. Huang, "Data forwarding with finite buffer capacity in opportunistic networks," in *Proc. 27th Wireless Opt. Commun. Conf. (WOCC)*, Hualien, Taiwan, Apr./May 2018, p. 1-5. doi: 10.1109/WOCC.2018.8372708.

[13] J. Wu and Z. Chen, "Sensor communication area and node extend routing algorithm in opportunistic networks," *Peer-Peer Netw. Appl.*, vol. 11, no. 1, pp. 90–100, Jan. 2018.

[14] R. Drãgan, R.-I. Ciobanu, and C. Dobre, "Leader election in opportunistic networks," in *Proc. Int. Symp. Parallel Distrib. Comput.*, Jul. 2017, pp. 157–164.

[15] A. Thakur, R. Sathiyanarayanan, and C. Hota, "STEEP: Speed and time-based energy efficient neighbor discovery in opportunistic networks," *Wireless Netw.*, vol. 25, no. 6, pp. 3275–3296, 2019.

[16] E. Talipov, Y. Chon, and H. Cha, "User context-based data delivery in opportunistic smartphone networks," *Pervasive Mobile Comput.*, vol. 17, pp. 122–138, Feb. 2015.

[17] X. Chen, C. Shang, B. Wong, W. Li, and S. Oh, "Efficient multicast algorithms in opportunistic mobile social networks using community and social features," *Comput. Netw.*, vol. 111, pp. 71–81, Dec. 2016.

[18] J. Tao, H. Wu, S. Shi, J. Hu, and Y. Gao, "Contacts-aware opportunistic forwarding in mobile social networks: A community perspective," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Barcelona, Spain, Apr. 2018, pp. 1–6. doi: 10.1109/WCNC.2018.8377216.

[19] D. Li, L. Ma, Q. Yu, Y. Zhao, and Y. Mao, "An community detection algorithm based on the multi-attribute similarity," in *Proc. Int. Conf. Commun. Circuits Syst.*, Nov. 2013, pp. 118–121.

[20] Y. Yan, Z. Chen, J. Wu, L. Wang, K. Liu, and Y. Wu, "Effective data transmission strategy based on node socialization in opportunistic social networks," *IEEE Access*, vol. 7, pp. 22144–22160, 2019.

[21] K. Jang, J. Lee, S. K. Kim, J.-H. Yoon, and S.-B. Yang, "An adaptive routing algorithm considering position and social similarities in an opportunistic network," *Wireless Netw.*, vol. 22, pp. 1537–1551, Jul. 2016.

[22] D. K. Sharma, S. K. Dhurandher, I. Woungang, A. Bansal, and A. Gupta, "GD-CAR: A genetic algorithm based dynamic context aware routing protocol for opportunistic networks," in *Proc. Int. Conf. Netw.-Based Inf. Syst.* Cham, Switzerland: Springer, 2017.

[23] K. Liu, Z. Chen, J. Wu, and L. Wang, "FCNS: A fuzzy routing-forwarding algorithm exploiting comprehensive node similarity in opportunistic social networks," *Symmetry*, vol. 10, no. 8, p. 338, 2018. doi: 10.3390/sym10080338.

[24] Y. Lin, Z. Chen, and J. Wu, "Efficient forwarding strategy for opportunistic network based on node similarity," in *Proc. Nat. Conf. Theor. Comput. Sci.* Singapore: Springer, 2017, pp. 86–100.

[25] P. Nguyen and K. Nahrstedt, "Crowdsensing in opportunistic mobile social networks: A context-aware and human-centric approach," 2017, *arXiv:1704.08598*. [Online]. Available: https://arxiv.org/abs/1704.08598

[26] S. K. Dhurandher, S. J. Borah, I. Woungang, A. Bansal, and A. Gupta, "A location prediction-based routing scheme for opportunistic networks in an IoT scenario," *J. Parallel Distrib. Comput.*, vol. 118, pp. 369–378, Aug. 2018.

[27] E. H. Lee, D. Y. Seo, and Y. W. Chung, "An efficient routing protocol using the history of delivery predictability in opportunistic networks," *Appl. Sci.*, vol. 8, no. 11, p. 2215, 2018.

[28] R. Wang, X. Wang, F. Hao, L. Zhang, S. Liu, L. Wang, and Y. Lin, "Social identity—Aware opportunistic routing in mobile social networks," *Trans. Emerg. Telecommun. Technol.*, vol. 29, no. 5, 2018, Art. no. e3297.

[29] H. Zhou, V. C. M. Leung, C. Zhu, S. Xu, and J. Fan, "Predicting temporal social contact patterns for data forwarding in opportunistic mobile networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 10372–10383, Nov. 2017.

[30] H. A. Nguyen and S. Giordano, "Context information prediction for social-based routing in opportunistic networks," *Ad Hoc Netw.*, vol. 10, no. 8, pp. 1557–1569, Nov. 2012.

[31] H. Zhang, Z. Chen, J. Wu, and K. Liu, "FRRF: A Fuzzy reasoning routing-forwarding algorithm using mobile device similarity in mobile edge computing-based opportunistic mobile social networks," *IEEE Access*, vol. 7, pp. 35874–35889, 2019.

[32] T.-K. Huang, C.-K. Lee, and L.-J. Chen, "PRoPHET+: An adaptive prophet-based routing protocol for opportunistic network," in *Proc. IEEE Int. Conf. Adv. Inf. Netw. Appl.*, Apr. 2010, pp. 112–119.

[33] J. Wu, Z. Chen, and M. Zhao, "Information transmission probability and cache management method in opportunistic networks," *Wireless Commun. Mobile Comput.*, vol. 2018, Jan. 2018, Art. no. 1571974.

[34] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Spray and wait: An efficient routing scheme for intermittently connected mobile networks," in *Proc. ACM Workshop Delaytolerant Netw.*, Philadelphia, PA, USA, Aug. 2005, pp. 252–259.

[35] C. Boldrini and A. Passarella, "HCMM: Modelling spatial and temporal properties of human mobility driven by users' social relationships," *Comput. Commun.*, vol. 33, no. 9, pp. 1056–1074, 2010.

PENG ZHENG is currently pursuing the master's degree with the School of Computer Science and Engineering, Central South University. His current research interests include wireless communications, complex networks, and opportunistic networks.

HONGXIAO FEI is an Associate Professor with the School of Computer Science and Engineering, Central South University. His current research interests include information filtering, network security, wireless communications, and complex networks.

JIA WU received the Ph.D. degree from the School of Computer Science and Engineering, Central South University, Chang-sha, Hunan, China, in 2016, where he is a special Associate Professor. He is an Engineer with "Mobile Health" Ministry of Education-China Mobile Joint Laboratory and a distinguished Associate Professor with the School of Computer Science and Engineering, Central South University. He is a Senior Member of CCF, a member of ACM. His research interests include wireless communications and networking, wireless network, and mobile health in network communication.

YEQING YAN is currently pursuing the master's degree with the School of Computer Science and Engineering, Central South University. She is also a Researcher with "Mobile Health" Ministry of Education-China Mobile Joint Laboratory. Her current research interests include communications and networking, complex networks, and opportunistic networks.