IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 21, 2019, accepted May 4, 2019, date of publication May 7, 2019, date of current version May 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2915359

Interest Characteristic Probability Predicted Method in Social Opportunistic Networks

JIA WU, (Member, IEEE), GENGHUA YU, AND PEIYUAN GUAN, (Student Member, IEEE)
School of Computer Science and Engineering, Central South University, Changsha 410083, China

Corresponding author: Genghua Yu (cnygh821@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61672540, and in part by the Hunan Provincial Natural Science Foundation of China under Grant 2018JJ3299 and Grant 2018JJ3682.

ABSTRACT In social opportunistic networks, mobile devices can be regarded as socialization nodes. Furthermore, they carry and store useful information. Mobile devices can select destination nodes and deliver messages through social opportunistic networks because messages can be securely and conveniently stored, carried, and transmitted with nodes. According to such characteristics, communities can be established. However, many communities may deliver messages only depending on one or two nodes. If those nodes are not enough cache and over-flooding, data transmission in communities may wait for a long time. To find many cooperation neighbor nodes can share much useful information. This paper proposes a method that characteristic interests with neighbors are selected or pushed service by users. According to establish interest characteristic probability predicted method, interest neighbors can be selected and receive or send information. Compared with the social opportunistic network algorithms, the new research method achieves better results in accuracy selected and delivery ratio, transmission delay, and routing overhead. According to the simulation experiments, the average delivery ratio of new research method is 0.88, which is 40% higher than that of the Epidemic algorithm; the end-to-end delay reduces 38% with Spray and wait for algorithm.

INDEX TERMS Social networks, interest, neighbor, probability distribution, information transmission.

I. INTRODUCTION

In recent years, as wireless networks have penetrated into our daily lives, the application scale of network has been increasing. As a new type of self-organizing network, it has attracted the attention of researchers at home and abroad [1]–[2]. To get rid of the restriction of establishing the end-to-end communication path to achieve network communication, the concept of social opportunistic network is proposed. This concept has been widely used in animal tracking, vehicle network and other fields [3]–[5]. Social opportunistic networks belong to intermittent connectivity networks. Nodes in social opportunistic networks are characterized by typical mobility, openness, and sparseness. They have low encounter rates and lack fixed and secure connectivity links. Generally, the "Storage-Carrying-Forwarding" mechanism [6] relies on the opportunity brought by node mobility to realize routing. This model requires that all nodes cooperate to forward the routing messages of other nodes in a coordinated manner and realize communication hop by hop through the chances of encounters caused by node movement.

With the rapid growth of the data volume in wireless communication networks, the number of data transmission and the data volume per transmission are also greatly improved, but nodes do not have enough capacity to transmit a large amount of data. Also, nodes need not only to transmit data, but also to calculate some tasks, so nodes consume more energy. At present, many compute intensive mobile applications are deployed on nodes, so the time for nodes to enter the sleep state is faster because of low energy [7]–[10]. As a result, the issue of energy efficiency becomes more serious and challenging in wireless communication networks. Therefore, to improve energy efficiency, we start from reducing energy consumption of nodes. Especially, in order to reduce the energy consumption of nodes in wireless communication networks, we mainly study from the following two aspects.

On the one hand, we propose efficient data packet iteration based on social opportunistic networks. Social opportunistic networks derive from the end-to-end communication of mobile devices carried by people based on

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

the encounter opportunity [11]–[13]. When the transmission domain between users cannot be reached, the social opportunistic network usually adopts the store-and-forward method to complete the data transmission process [9], and the message carrying node forwards the message to a certain relay node or the destination node by means of the encounter opportunity caused by movement to forward. In wireless networks, nodes in an unpredictable geographical location need to communicate with each other at unpredictable geographical intervals. Therefore, as a multi-hop wireless technology, social opportunistic network proposes that end-to-end data transmission can be achieved through ''opportunistic communication''. Such communication relies on the movement of nodes and the effective routing-forwarding algorithm, so as to give a node the opportunity to act as a relay node in some end to end data transmission. In addition, an effective routing forwarding algorithm may select reliable relay nodes for efficient data transmission [14]. Compare with other routing-forwarding algorithms, an effective routing forwarding algorithm have the less number of hops in all successful message forwarding processes, and then the less time and energy nodes spend in the forwarding processes. Therefore, how to select reliable relay nodes is a very important problem in routing-forwarding algorithm.

On the other hand, different routing-forwarding algorithms take different similarity factors of nodes in the network into account, and few algorithms take multiple similarity factors into consideration [15]–[18]. However, each of these similarity factors plays a different role in deliver ratio. Therefore, the routing-forwarding algorithm based on node similarity should consider multiple similarity factors rather than a single factor. Human movement and social attribute usually reflect a special relationship between human, which can be a powerful condition for the choice of relay nodes [19]. In social scene, the corresponding geographic locations of nodes at different times combine to represent the moving trajectories of nodes. People may keep the same movement track almost every day for a certain period of activity, which means they are likely to meet the same people in the same period of time every day, such as from the place of residence to the place of work. In social opportunistic networks, besides the movement of node, each node has a large number of social attributes, such as interest, occupation, place of residence and place of work. The similarity between nodes is calculated by evaluating the social attribute value of nodes [20]. Based on the higher social similarity, nodes that encounter more frequently are divided into the same community and thus more opportunities to exchange information.

This work proposes a method that characteristic interests with neighbors are selected or pushed service by users. According to establish Interest Characteristic Probability Predicted (ICPP) algorithm, we can select interest neighbors to receive or send information. A good performance is achieved with nodes that extend network lifetime reduce delay and overhead in social communication environment.

The study makes the following contributions:

(1) Interest behaviors by neighbors are established with multiple values of nodes in social communication.

(2) By effectively analyzing the attribute preference of nodes in transmission process, we can define the individual preference of nodes as preference similarity, so as to measure their impact on information transmission.

(3) In accordance with the simulation results in the Opportunistic Networking Environment (ONE), we analyze the performance of ICPP and compared it with Epidemic algorithm, Spray and wait algorithm and PRoPHET algorithm, ICPP algorithm shows enhanced performances in increasing the delivery ratio and End-to-end delay and overhead.

The structure in this study is as follow. Related works is section 2; system design is section 3; simulation is section 4 and conclusion is section 5.

II. RELATED WORKS

At present, many study methods in social opportunistic networks have focused on algorithm research. Some algorithms can be adapted to different application fields. In different scenes, effective methods can be established interdisciplinary by improving the available algorithms. Some existing opportunistic network algorithms are discussed as follows.

In recent years, the academic has done a lot of researches around the routing-forwarding algorithms in social opportunistic networks [8], [12], [17], [18], [19], and proposed different effective methods under different application scenarios. In social opportunistic networks, routing-forwarding algorithms are usually divided into two types: context-aware routing-forwarding algorithms and non-context-aware routing-forwarding algorithms. Context-aware routing forwarding algorithm based on the similarity of nodes to select relay nodes through the social relations between nodes and the contextual information related to nodes [7], [12], [17].

In addition, although context-aware routing-forwarding algorithms can improve the transmission environment and improve transmission efficiency, these algorithms usually need to manage a large amount of information and perform computing tasks, thus bringing additional delay and energy consumption to the network. However, non-context-aware routing-forwarding algorithms perform flooding transmission, which brings many redundant messages group copies to the network, and eventually leads to extremely high forwarding delay and energy consumption of the network [8], [18], [19]. It can be seen that both the context-aware routing-forwarding algorithm and the non-context-aware routing-forwarding algorithm will bring some extra delay and energy consumption to the entire wireless network, especially the non-context aware routing-forwarding algorithm.

In context-aware routing-forwarding algorithms, many studies calculate the similar level between nodes to define the relationship between nodes, such as the possibility of a future encounter between nodes, the moving trajectory of nodes, and community partitioning of nodes. In [20], Wang *et al.* innovatively extracts social identity from messages generated by mobile nodes, and proposes the Exclusive OR algorithm that takes into account the multiple social identities of mobile nodes and their corresponding social influences. By the final simulation results, the performance of data transmission can be improved by taking social identity into account. However, the Exclusive OR algorithm does not consider a variety of social attributes.

In [21], Wu *et al.* proposes the SRBRA algorithm, which is based on social relations. Firstly, real-time data generated by nodes are analyzed and summarized, and then specific factors affecting social relations between nodes are extracted to calculate the value of social relations between nodes. Finally, according to the social relation value between nodes, the social relation value between the neighbor node and the destination node is sorted to select the optimal next-hop relay node to complete the transmission of messages. However, the SRBRA algorithm does not take the mobility of nodes into consideration. Besides, in [22], Wu *et al.* studies a framework that takes individual context, society and relationships as matching opportunity predictors. The proposed algorithm based on a series of studies can predict the cooperation opportunities of data transmission between nodes, and then determine the end-to-end communication between nodes in the network according to the cooperation opportunities between nodes.

Some mathematical methods and models are usually used in context-aware routing-forwarding algorithms, such as markov decision model, set theory and graph theory. Of the three related works to be introduced, two uses game theory and the other uses graph theory. In [23], Nguyen and Nahrstedt proposes a new context routing protocol based on game theory to select the most appropriate relay node to forward packets. Through the non-zero cooperation times of two nodes, the context routing protocol builds the game depends on the context information, the distance between the corresponding node and the target node, and the encounter index.

In [24], in order to determine the cost to achieve efficient data transmission, Talipov *et al.* designed a model based on user context replication and the graph theory, which is an online backpack problem. The scheme learns and predicts the context information of each node in order to calculate the data delivery probability of each node, and the number of copied messages is adjusted based on the given delivery threshold. However, the scheme only considers the data information in the process of node transmission, which means the decision accuracy of message transmission needs to be improved in the process of transmission. Besides, in [25], in order to find the vertex cover suitable for the perceptive tasks in the group, Phuong Nguyen and Klara Nahrstedt designed a new context-aware approximation algorithm. At the same time, in order to assign the sensor task to a more "socialized" device for better sensor coverage, a human centered guidance strategy for initial assignment of the sensor device based on participants' meta information was also designed. And the node of this algorithm completes the individual coverage of a social vertex with the human-centered information.

Based on the introduction to social opportunistic networks technology, the next step is designing a model in social opportunistic networks.

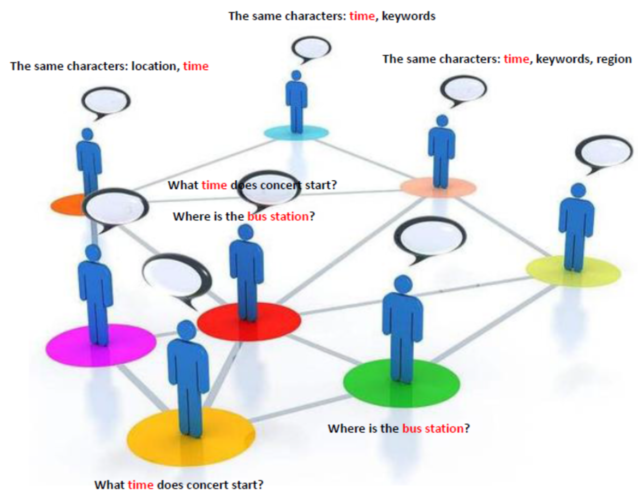

FIGURE 1. The characters and relationships in social networks.

TABLE 1. Data symbols in human activity.

Symbol	Significance
u,v,l,s,t,r	user; interest, location,
U,V	relationship, time, region
M,N	user
K,W	the number of users
R,T	keyword, time period

III. SYSTEM DESIGN

A. DEFINITION OF USER CHARACTERISTICS IN SOCIAL NETWORKS

Social networks usually focus on user information, such as geographic location, time period, region, and keywords.

Fig 1 shows the characters and relationships in social networks. For a person, he used to interest in the same characters when others mention similar topic. If a person establishes focusing on network, he has more 'opportunity' acquiring to 'interest point' or 'help' by mobile devices. He also found a good cooperation by neighbors when they have many similar characters.

In social networks, similar characters usually come from many items. Such as go to the place, buy some sale goods, watch exciting game at the same time. More similar characters can improve messages transmission in social.

In social networks, some data symbols can describe human activity. It assumes that a node in social networks. The node concludes many characters such as user, interest, location, relationship, time, and region. We can establish the data symbols in Table 1.

TABLE 2. Parameter symbols in human activity.

Symbol	Significance
θ_u	Interesting of u, multiple distribution of theme collections
$\theta_l^{'}$	Location of position l, multiple distribution of theme collections
ς_u	Relationship of u, multiple distribution of relationship collections
ϑ_u	Active of u, multiple distribution of motivation collections
ϕ_z	Keywords corresponds to multiple distributions on the theme z
ψ_z	Time correspond to the beta distribution on the theme z
φ_r	Propagation object corresponds to the regional distribution of region r
μ	Average value of location with region r
$\sum$	Covariance of location with region r
λ	Mixed weight of u: the parameter of sampling binary variable α
$b, b^{'}$	Beta priors of u
$\alpha, \alpha^{'}, \beta, \gamma, \tau$	Dirichlet distributed priors

Using the definitions in Table 1, we can analyze a user node's reference to other nodes. When a user node is needed to assess user interest, the behavior presented in Table 2 can be used to determine the characteristic parameters, such as probability of a user in selecting target, for obtaining interested users.

Interested users not only can send data information to each other but also can pay attention to the updates on data and information in a timely manner. From Table 1 and Table 2, we can establish interest model by users in social networks. The main ideas and work steps are shown in the following flow chart:

We consider that the properties of a node in communication can be used as a reference value for similar nodes. By analyzing and predicting the similarity of their individual preferences, appropriate nodes can be selected to cooperate effectively so as to complete the information transmission. The main way to build a community of interest feature models is shown in Figure 2. First, the nodes communicate with each other, and after establishing a communication connection, they can see each other's preference attributes. Then compare their interests and preferences with each other, if the fit is very high, the two nodes are divided into a community. If the fit is low, the two nodes do not constitute a community. Do the same with the other nodes. We judge the feature fitting degree of nodes by calculating the transmission probability and the prediction cooperation probability between nodes.

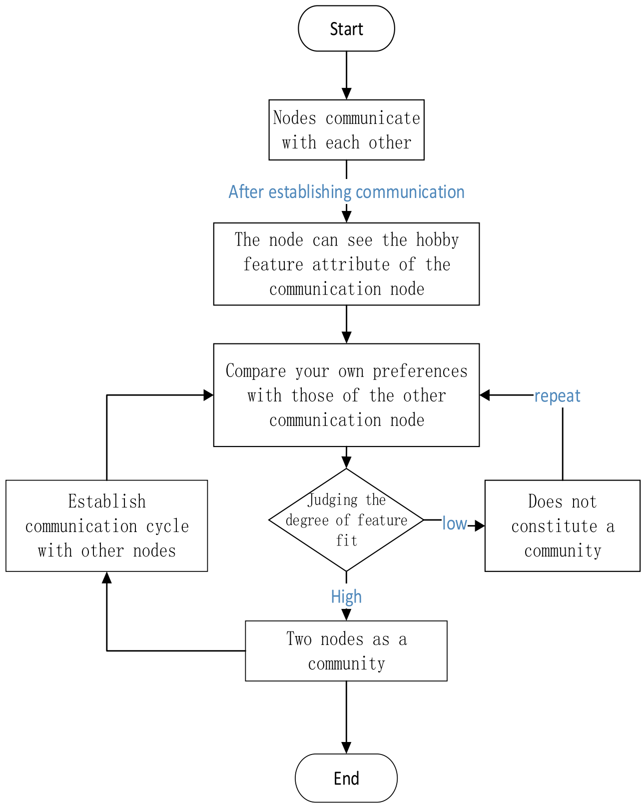

FIGURE 2. Main steps of the interest feature model.

Among them, the transmission probability is used to indicate whether it is possible to establish cooperation and transmission between two nodes. The prediction indicates whether the node can establish the probability of cooperation because there are more identical attributes. The difference between the two is that the former is implemented in the calculation, and the latter is realized by examining the commonality of the attributes.

B. ITERATIVE UPDATE OF INTERESTED USERS

To evaluate the characteristics of user interest, we establish the probability distribution model of θ_u, and this model is represented by parameter α

$$P(\theta_u|\alpha) = \frac{\Gamma(\sum_z \alpha)}{\prod_z \Gamma(\alpha)} \prod_z \theta_{u,z}^{\alpha-1} \tag{1}$$

$\Gamma(\sum_z \alpha)$ is a function. The polynomial interest parameter of α is shown as follows.

$$P(\theta_l{'}|\alpha') = \frac{\Gamma(\sum_z \alpha')}{\prod_z \Gamma(\alpha^{'})} \prod_z \theta'^{\alpha^{'}-1}_{l,z} \tag{2}$$

$$P(\phi_z|\beta) = \frac{\Gamma(\sum_w \beta)}{\prod_w \Gamma(\beta)} \prod_z \phi_{z,w}^{\beta-1} \tag{3}$$

$$P(\varsigma_u|\delta) = \frac{\Gamma(\sum_s \delta)}{\prod_s \Gamma(\delta)} \prod_s \varsigma_{u,s}^{\delta-1} \tag{4}$$

$$P(\vartheta_u|\gamma) = \frac{\Gamma(\sum_\gamma \gamma)}{\prod_\gamma \Gamma(\gamma)} \prod_\gamma \varsigma_{u,\gamma}^{\gamma-1} \tag{5}$$

$$P(\varphi_r|\tau) = \frac{\Gamma(\sum_v \tau)}{\prod_v \Gamma(\tau)} \prod_v \varphi_{r,v}^{\tau-1} \tag{6}$$

Given that all interests are in an independent distribution, we calculate the joint distribution of the polynomial interest parameter of α using Formulas (2)–(6).

$$\begin{aligned} &P(v, l_v, w_v, t, z, r, s|\alpha, \alpha', \beta, \gamma, \tau, \phi, \mu, \textstyle\sum) \\ &\quad = P(z|\alpha)P(z|\alpha')P(r|\gamma)P(s|\delta) \\ &P(w_v|z, \beta)P(v|r, \tau)P(t|z, \phi)P(l_v|r, \mu, \textstyle\sum) \\ &\quad = \int P(z|\theta)P(\theta|\alpha)d\theta \int P(z|\theta')P(\theta'|\alpha)d\theta' \\ &\qquad \times \int P(r|\vartheta)P(\vartheta|\gamma)d\vartheta \int P(z|\varsigma)P(\varsigma|\delta)d\delta \\ &\qquad \times \int P(w_v|z, \phi)P(\phi|\beta)d\phi \\ &\qquad \times \int P(v|r, \varphi)P(\varphi|\tau)d\varphi P(t|z, \phi)P(l_v|r, \mu, \textstyle\sum) \end{aligned} \tag{7}$$

For each user, interest collection represents (u, v, l_v, w_v, t). We obtain the value of user interest behavior by acquiring the social relevance S, potential theme Z, and potential region R of the user. Using Formula (7) and Bayes' rule, we can obtain the conditional probability. Social correlation s can be expressed as

$$P(s|s_{-u,v}, z, r, v, l_v, w_v, t, u) = \frac{n_{u,s'}^{-u,v} + \delta}{\sum_{s'} n_{u,s'}^{-u,v} + \delta} \tag{8}$$

In the equation above, $s_{-u,v}$ represents the social relevance of all current user relationships; $n_{u,s'}$ represents the number of social correlation s sampled from the social correlation distribution of user u; $-u, v$ represents the number of entries except for current attendance records. According to the principle of subject consistency, we analyze the probability sampling (0, 1):

$$\begin{aligned} &P(a = 1|a_{-u,v}, z, u) \\ &\quad = \frac{n_{u,z'}^{-u,v} + \alpha}{\sum_{z'} n_{u,z'}^{-u,v} + \alpha} \cdot \frac{n_{u,a1}^{-u,v} + b}{n_{u,a1}^{-u,v} + n_{u,a0}^{-u,v} + b + b'} \end{aligned} \tag{9}$$

$$\begin{aligned} &P(a = 0|a_{-u,v}, z, u) \\ &\quad = \frac{n_{l,z'}^{-u,v} + \alpha'}{\sum_{z'} n_{l,z'}^{-u,v} + \alpha'} \cdot \frac{n_{u,a0}^{-u,v} + b'}{n_{u,a1}^{-u,v} + n_{u,a0}^{-u,v} + b + b'} \end{aligned} \tag{10}$$

where $n_{u,a1}^{-u,v}$ represents the number of visits to the user literature during $a = 1$; $n_{u,a0}^{-u,v}$ indicates the number of visits to the user literature during $a = 0$. Z represents the interest of choice. By choosing z, when $a = 1$, we obtain

$$\begin{aligned} &P(z|a = 1, z_{-u,v}, s, r, v, l_v, w_v, t, u) \\ &\quad = \frac{n_{u,z'}^{-u,v} + \alpha}{\sum_{z'} n_{u,z'}^{-u,v} + \alpha} \cdot \frac{(1-t)^{\frac{1}{\phi_{z,1}}} t^{\frac{1}{\phi_{z,2}}}}{B(\varphi_{z,1}, \varphi_{z,2})} \cdot \prod_{w \in w_v} \frac{n_{z,w'}^{-u,v} + \beta}{\sum_{w'} n_{z,w'}^{-u,v} + \beta} \end{aligned} \tag{11}$$

When $a = 0$

$$\begin{aligned} &P(z|a = 0, z_{-u,v}, s, r, v, l_v, w_v, t, u) \\ &\quad = \frac{n_{u,z'}^{-u,v} + \alpha'}{\sum_{z'} n_{u,z'}^{-u,v} + \alpha'} \cdot \frac{(1-t)^{\frac{1}{\phi_{z,1}}} t^{\frac{1}{\phi_{z,2}}}}{B(\varphi_{z,1}, \varphi_{z,2})} \cdot \prod_{w \in w_v} \frac{n_{z,w'}^{-u,v} + \beta}{\sum_{w'} n_{z,w'}^{-u,v} + \beta} \end{aligned} \tag{12}$$

We pass the probability sampling region r:

$$\begin{aligned} &P(r|r_{-u,v}, z, r, v, l_v, w_v, t, u) \\ &\quad = \frac{n_{u,r}^{-u,v} + \gamma}{\sum_{r'} n_{u,r'}^{-u,v} + \gamma} \cdot \frac{n_{r,v}^{-u,v} + \tau}{\sum_{v'} n_{r,v'}^{-u,v} + \tau} \cdot P(l_v|\mu_r, \textstyle\sum_r) \end{aligned} \tag{13}$$

We can update parameters μ_r and $\sum_r$ using the matrix generated by the potential theme z and potential region r in the iteration

$$\mu_r = E(r) = \frac{1}{|s_r| - 1} \sum_{v \in s_r} l_v \tag{14}$$

$$\sum\nolimits_r = D(r) = \frac{1}{|s_r| - 1} \sum_{v \in s_r} (l_v - \mu_r)(l_v - \mu_r)^T \tag{15}$$

where s_r represents the interest collection of potential regional r. With the interest set, we can update the ϕ parameter of the next interest point:

$$\phi_{z,1} = t_z\left(\frac{t_z(1 - t_z)}{s_z^2} - 1\right) \tag{16}$$

$$\phi_{z,2} = (1 - t_z)\left(\frac{t_z(1 - t_z)}{s_z^2} - 1\right) \tag{17}$$

where t_z and s_z^2 represent the sampling mean and covariance of the time stamps configured on theme z, respectively.

By updating the iterative social correlation s and potential theme z and region r, the next moment of interest feature model can be expressed as

$$\theta_{u,z}^{(t+1)} = \theta_{u,z}^{(t)} + \frac{n_{u,z'} + \alpha}{\sum_{z'} n_{u,z'} + \alpha} \tag{18}$$

$$\theta'^{(t+1)}_{l,z} = \theta_{l,z}^{(t)} + \frac{n_{l,z'} + \alpha'}{\sum_{z'} n_{l,z'} + \alpha'} \tag{19}$$

$$\phi_{z,w}^{(t+1)} = \phi_{z,w}^{(t)} + \frac{n_{z,w} + \beta}{\sum_{w'} n_{z,w'} + \beta} \tag{20}$$

$$\varsigma_{u,s}^{(t+1)} = \varsigma_{u,s}^{(t)} + \frac{n_{u,s'} + \delta}{\sum_{s'} n_{u,s'} + \delta} \tag{21}$$

$$\vartheta_{u,r}^{(t+1)} = \vartheta_{u,r}^{(t)} + \frac{n_{u,r} + \gamma}{\sum_{r'} n_{u,r'} + \gamma} \tag{22}$$

$$\varphi_{r,v}^{(t+1)} = \varphi_{r,v}^{(t)} + \frac{n_{r,v} + \tau}{\sum_{v'} n_{r,v'} + \tau} \tag{23}$$

$$\mu_z^{(t+1)} = \mu_z^{(t)} + \mu_z \tag{24}$$

$$\sum\nolimits_r^{(t+1)} = \sum\nolimits_r^{(t)} + \sum\nolimits_r \tag{25}$$

$$\lambda_u = \frac{n_{us1} + b}{n_{us1} + n_{us0} + b + b'} \tag{26}$$

C. USER INTEREST QUERY AND RECOMMENDATION

If we need to query whether user u currently matches our requirements for matching degree, then we can analyze it by querying the probability joint distribution of this user. Suppose we add the Ω feature vector to our regular feature, which represents our newly-added goal. Through probability matching model analysis, we obtain

$$\begin{aligned} &P(v, l_v, w_v, t|u, r, \Omega) \\ &\quad = P(l_v|r, \Omega)P(v|r, \Omega) \sum_s (\prod_{u \in s} P(s|u, \Omega))^{\frac{1}{|w_v|}} \\ &\quad\quad \cdot \sum_z P(z|u, \Omega)P(z|l_v, \Omega)P(t|z, \Omega) \\ &\quad\quad \cdot (\prod_{w \in w_v} \sum_z P(w|w, \Omega))^{\frac{1}{|w_v|}} \end{aligned} \tag{27}$$

By updating the correlation s society, underlying themes r and z, and potential area, we can track and carry out the Ω eigenvector user change data process. Using the joint probability distribution value will decide whether the current user is added to the interested users.

If the current user conditions meet the requirements of our interested users, that is, social relevance s and underlying themes z and r conform to the target users and potential areas, then we can recommend data information to the user.

Assuming that the user carries q data information in line with the conditions of the current interested user, we can set an interest rate $S(q, v)$, which indicates the v user with q data information scores. The rank of interest is expressed as

$$S(q, v) = \sum_r w(q, r)F(v, r) \sum_s w(q, s) \sum_z w(q, z)F(v, z) \tag{28}$$

where w represents the weight of potential theme z and region r in terms of social relevance s. Among them,

$$w(q, r) = P(l_q|\mu_r, \sum\nolimits_r) \tag{29}$$

$$w(q, s) = (\prod_{u \in w_v} \varsigma_u, s)^{\frac{1}{|w_v|}} \tag{30}$$

$$w(q, z) = \theta_{u,z}\theta_{l_v,z}\theta_{z,t} \tag{31}$$

By calculating $S(q, v)$, we can rank the user nodes in the social network and finally select the node of the corresponding score as the interest node and prioritize such data information.

The pseudo-code with ICPP is in Table 3.

TABLE 3. Interest characteristic probability predicted (ICPP) algorithm.

Algorithm 1: Interest Characteristic Probability Predicted (ICPP) algorithm
Input: source node S[i] Output: Community C[i];

Start
source node S[i] moves at random;
While(S[i] != NULL)
{
S[i] send similar characters;
For i from 1 to n;
If characters .S[i+1]= characters .S[i]
Community C[i]=< S[i], S[i+1]>;
}
End While
While(C[i]!= NULL)
{
S[i] sends messages to neighbors *N* ;
neighbors *M* response S[i];
neighbors *M*< neighbors *N*;
Calculate neighbors *M* -> C[i];
Calculate structure list;
}
End While
End

The time complexity in ICPP is O(n). Nodes in community could be established a list. If neighbors can be selected by nodes, the list could be added. Nodes cannot wait for response and move to other path. It is not like Epidemic algorithm and Spray and wait algorithm, nodes must send and receive all messages. There is no community can afford redundant neighbors or cache. So, the time complexity in Epidemic algorithm and Spray and wait algorithm are O(n^2).

IV. SIMULATION

In this paper, The One Simulator is used to simulate the proposed algorithm, and some opportunistic network classical routing algorithms are compared. The performance of ICPP algorithm is evaluated from the aspects of delivery ratio, overhead, and transmission delay. We use an open street map to edit city maps in ONE. In the simulation, Shortest Path Map-Based Movement (SPMBM) is used by calculated the movement with nodes by collecting real map data according to the minimum distance coordinate. In addition, the simulation adopts an open street map to edit city maps in Helsinki. Different parks, streets, and shops are established in the map [25]. They can exhibit a real environment.

The parameters can be settled based on the random models social networks. The parameters adopted in the experiment are set as follows. The simulation time is 100 minute to 400 minute, and the simulation area is 4500 m × 3400 m in the map. The involved nodes are 2000. The transmission pattern is broadcast, the maximum transmission area of each node is 10 m^2, and the sending frequency of a data packet is 25 s to 35 s. The data packet type is random array. Moreover, a node consumes one Joule unit energy when it sends a data packet; initial energy for node is 100J. Each node carries 10 data packets, and the transmission pattern of nodes is a social model. Furthermore, the transmission speed of the node is 0.5–1.5 m/s, and the cache of each node is 5 MB. In parameter setting, weight w is marked signw_{new},w_{delete}, w_{change}, w_{old}; other parameter of node concludes: connect time t, relationship list r, location l.

ICPP is compared with types of classical algorithm mentioned to verify its performance. This study focuses on the following parameters:

(1) Selected and delivery ratio: This parameter explains to the probability of selecting a relevance node in delivering messages.

(2) Average end-to-end on delay: This parameter includes the delay of route selecting, waiting delay in the data classification queue, transmission delay, and redelivering in MAC.

(3) Overhead on average: This parameter shows the cache space in source node when information is transmitted. In this paper, we can design the node which delivers data to its neighbors is source node.

In simulation, we firstly compare with node and its neighbors and find a conclusion how many neighbors are the most suitable for node to transmit messages in ICPP. We count average value in 2,000 nodes.

Fig 3 shows the relationship between delivery ratio and neighbors. We found from 1 neighbor to 4 neighbors, node in ICPP can improve its delivery ratio. When the cooperation neighbors become 4, the delivery ratio is 0.88. That is to say, more neighbors can find same interesting point nodes easily. However, over 5 neighbors joined in transmission, the delivery ratio can reduce because more redundancy information have delivered by neighbors. Interesting points may have not been received by neighbors when node sent.

Fig 4 shows the relationship between delay and neighbors. More neighbors can reduce delay if nodes can share transmission mission. From 1 neighbor to 3 neighbors, delay can be reduced and nodes can acquire the best transmission environment. Over 3 neighbors take apart in transmission;

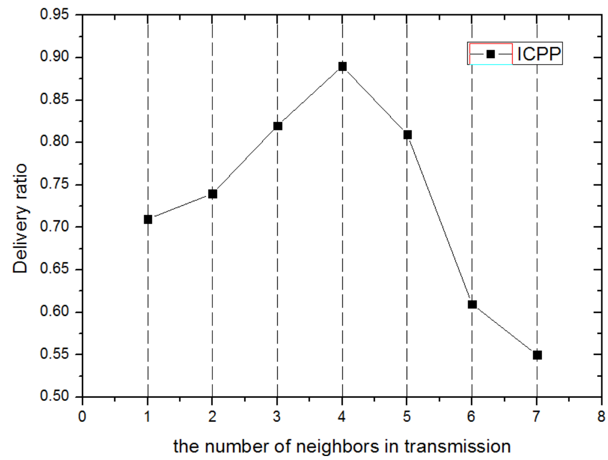

FIGURE 3. The relationship between delivery ratio and neighbors.

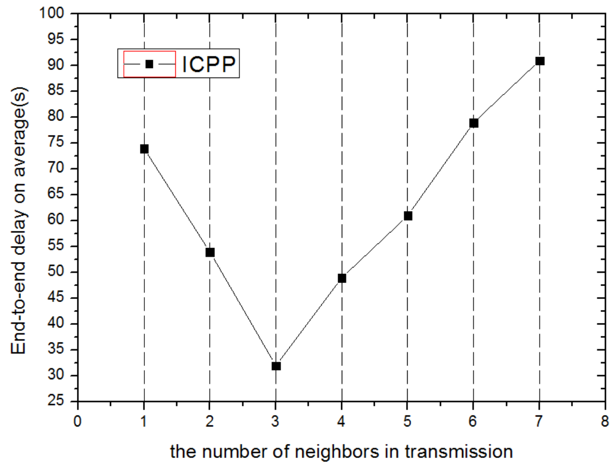

FIGURE 4. The relationship between delay and neighbors.

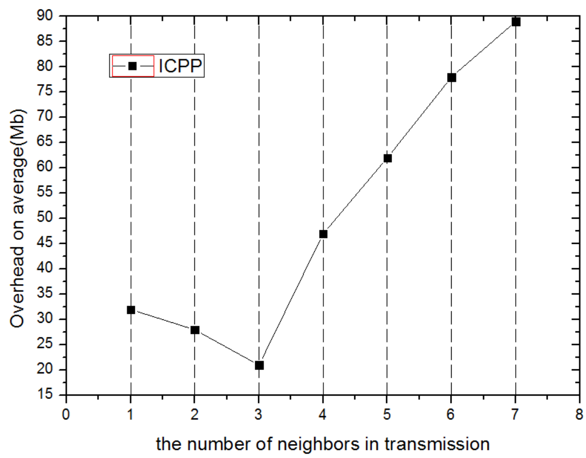

FIGURE 5. The relationship between overhead and neighbors.

redundancy information may affect quality in transmission. The same condition is as Fig 5, much cache have been consumed. That is to say, over 3 neighbors joined in transmission, overhead and delay would increase.

The next step, we would compare with delivery ratio, delay and overhead between ICPP and other algorithms.

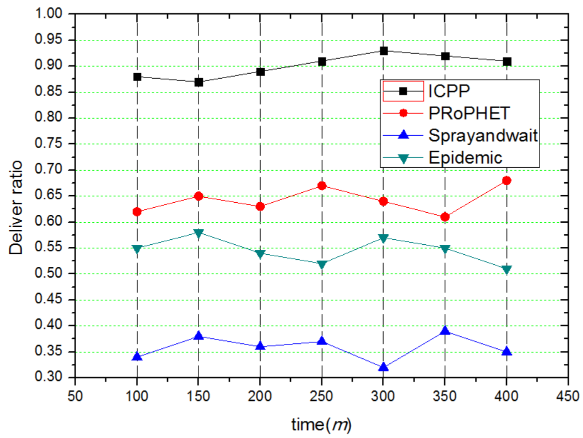

FIGURE 6. Delivery ratio and simulation time.

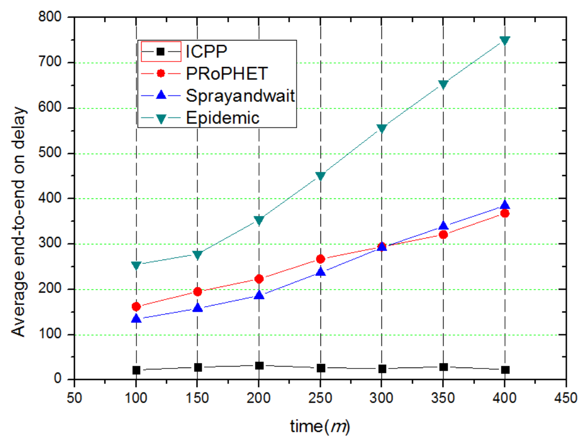

FIGURE 7. End-to-end delay on average and simulation time.

Fig 6 shows the relationship between delivery ratio and simulation time. Spray and wait algorithm has the lowest delivery ratio, it is only 0.32 to 0.38. The reason is that algorithm adopted over-flood spray to transmit information. Each node has to send much information. Epidemic algorithm improves information transmission and reduces the number of data copies. It is better than Spray and wait algorithm. Epidemic algorithm adopts a packet queue according to transmission cost. It determines the replication order according to the priority of packets, which not only avoids congestion, but also save resources by blindly copying message. PRoPHET algorithm adopted probabilistic prediction. Much cache can be used in transmission. The target node can be founded easily. So the delivery ratio is better than Epidemic algorithm and Spray and wait algorithm. The ICPP algorithm has the highest transmission deliver ratio, reaching 0.87 to 0.93. Precisely, it uses a combination of features to calculate the social network trust relationship to select trusted nodes, which reduces congestion and information replication, and then improves the efficiency of the selected node and the reachability of the destination node. It also effectively improves delivery ratio.

Fig 7 shows the relationship between End-to-end delay on average and simulation time. From the figure, the average delay of the Epidemic algorithm reaches over 700 when the simulation time has over 400 minutes. The Epidemic routing algorithm uses flooding to deliver packets. As time increases, more and more data packets are transmitted. The resources have been consumed in large amounts. PRoPHET algorithm and Spray and wait algorithm are better than Epidemic algorithm. The average delay has been controlled from 130 to 350. Effective methods can limit nodes delivering messages at random. The average delay in ICPP is very low. It has only no more than 50. Because this method adopted to interesting point transmission. If characteristics are not the same or forcing, information cannot be accepted. It avoids a number of redundancy data transmitting between node and its neighbors.

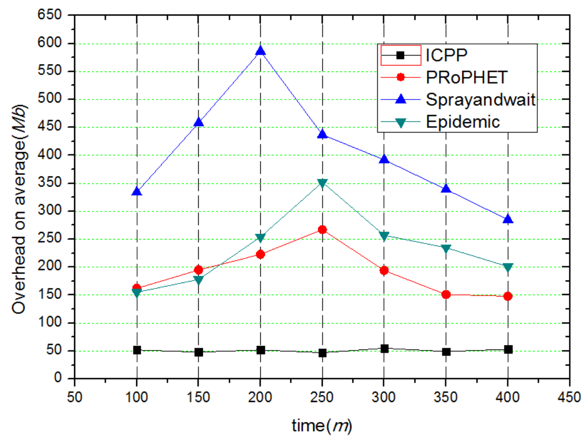

FIGURE 8. Overhead on average and simulation time.

Fig 8 shows the relationship between overhead and simulation time. From the figure, we can see that overhead of the ICPP algorithm is not affected by the time. Spray and wait algorithm has the highest overhead because over-flood spraying. The top overhead is 580. After this moment, many nodes cannot send and receive messages because caches have not enough. The overhead can reduce slowly. The same condition appeared to PRoPHET and Epidemic algorithm. ICPP algorithm is not affected when the simulation time increased. Because many interesting point nodes receive and send messages by neighbors. Much useful information can be transmitted immediately. There are enough cache to delivering next data packets. So, it is good performance in simulation.

Fig 9 shows the relationship between the number of selected neighbor nodes and the delivery ratio. It can be seen from the figure that their trend is that as the number of neighbor nodes increases, the transmission success rate increases and then decreases. Due to the number of cooperative neighbor nodes selected by the node, there are many redundant data packets in the network, which occupy the resources of the network and cause network congestion. Therefore, it will degrade network performance and reduce the transmission success rate. As can be seen from the figure, the Spay and wait algorithm and the Epidemic routing algorithm have the lowest transmission success rate, and the transmission success rates

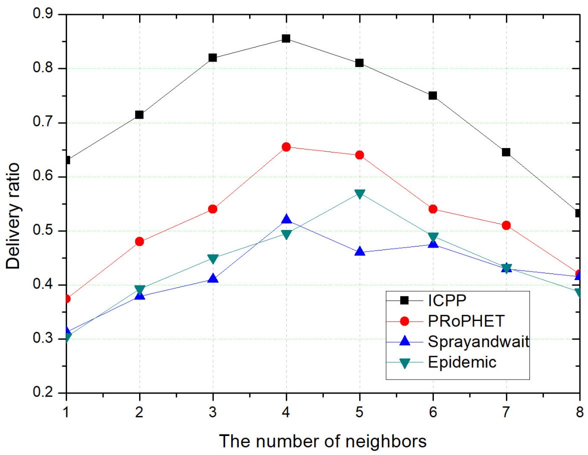

FIGURE 9. The relationship between the number of neighbors and delivery ratio.

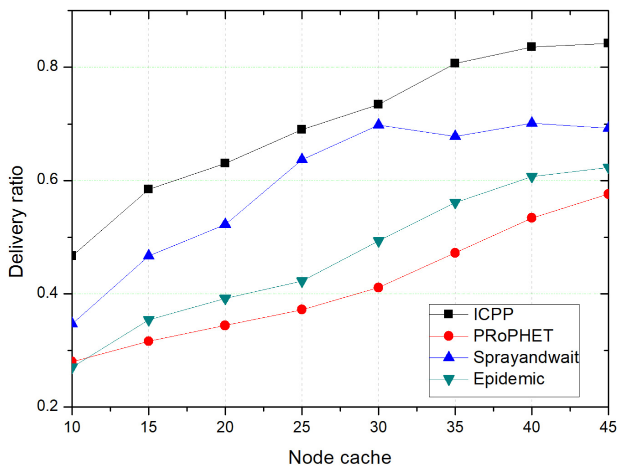

FIGURE 10. The relationship between node cache and delivery ratio.

of the two algorithms are between 0.3 and 0.6. The ICPP algorithm is always better than other algorithms. Since the neighbor nodes we choose are not random, the transmission of information through effective nodes can reduce the redundancy of a large amount of data in the network. However, the number of effective nodes is not large. As the number of neighbor nodes increases, some nodes that have little influence on data transmission are selected as neighbor nodes, which has no gain for network performance improvement. Therefore, the transmission success rate will also decrease.

Fig 10 shows the relationship between node cache and delivery ratio. As can be seen from the figure, when the node cache is small, the impact on the delivery ratio is large. From the trend point of view, increasing the cache of nodes can effectively improve the success rate of data transmission. Among them, the transmission success rate of the ICPP algorithm is higher than other algorithms, and tends to be stable as the cache increases to 35M. Its trend is similar to the Spay and Wait algorithm, which is independent of the data cache size when considering neighbor nodes. Therefore, when the cache reaches a certain value, the cache size has less impact on our algorithm. Regarding the Epidemic algorithm that delivers messages based on flooding mode can be seen from Fig 10, the size of the cache has a greater impact on it. The larger the node cache, the more messages can be cached. The delivery ratio increases due to a reduction in congestion conditions due to a large amount of data.

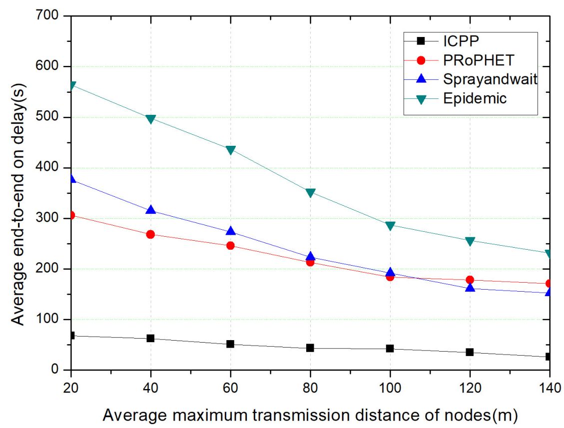

FIGURE 11. The relationship between average distance and end-to-end delay on average.

Fig 11 shows the relationship between the average maximum transmission distance of the nodes and the average end-to-end delay. We can see from the figure, the delay decreases with the increase of the average maximum transmission distance of the nodes. This is because the increase of the data transmission distance can increase the probability of encountering with the target node and reduce the data transmission time. Among them, the Epidemic algorithm has the highest average delay. The high delay of this algorithm is due to the use of flooding mode to transfer data, which will cause a lot of data redundancy to the network. Therefore, it is greatly affected by the transmission range of the node, and the descending speed is fast. The average delay of the ICPP algorithm is always lower than other algorithms and is less affected by the transmission range. Because the method we proposed effectively improves the way to filter nodes, the larger the transmission range, the more nodes that can provide filtering. Their contribution to the data transfer path will be more helpful in reducing latency.

The simulation in the ICPP algorithm uses different mobile models and cache to demonstrate the performance. The simulation uses the SPMBM, random walk (RM), and random way point (RWP) models.

Fig 12 explains three different topological structure models in communication map. In the RM model, the structure is sparse and the performance is limit, because, the effective data cannot be delivered to arrive at destination nodes. In this condition, the selected and delivery in RM is only 0.65. In RWP model, the delivery ratio is better than RM when messages are delivering in data packets because the density for nodes on the map is larger RM model. The delivery ratio in RWP exceeds over 0.75. However, both RWP and RM

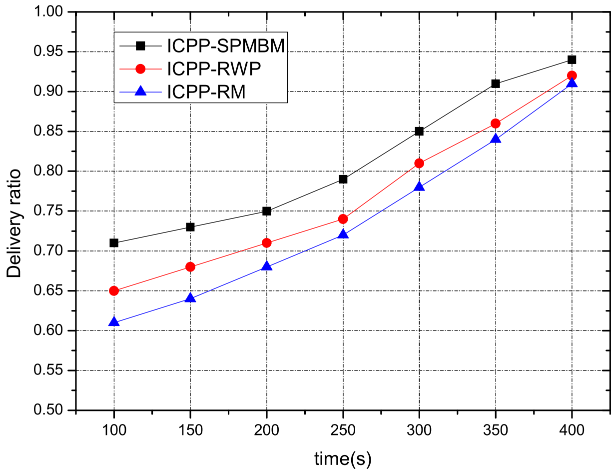

FIGURE 12. Delivery ratio in different models.

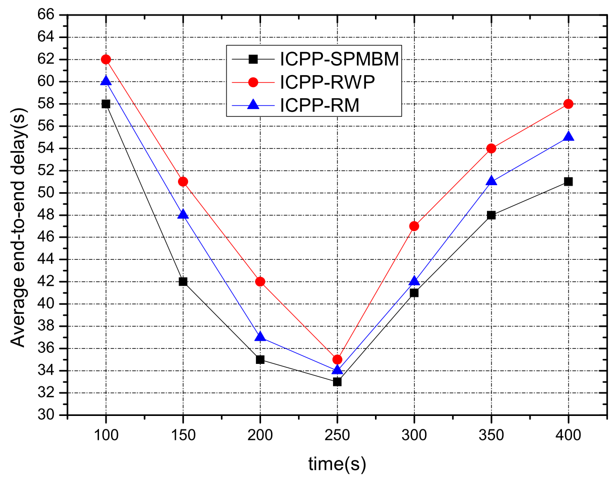

FIGURE 13. Average end-to-end delay of different models.

adopt a non directional and simple movement. The weight among nodes transmission messages depend on movement and meeting. So, transmission delay in different model may occur in finding a neighbor and in waiting to send. As shown in Fig 13, the delay in RWP and RM method are higher than that in SPMBM. The ICPP algorithm uses the SPMBM model, which can record mobile routing. It shows a real node movement model.

V. CONCLUSION

In this study, we contribute a social networks algorithm that characteristic interests with neighbors are selected or pushed service by users to solve the problems in social networks. With satisfactory results from simulation and comparison with some existing algorithms, the new method is found to not only decrease energy consumption but also improve the delivery ratio and overhead in social networks.

In future work, we will consider the energy consumption of nodes in the case of large-scale data transmission in 5G network. Explore more effective data transmission methods to analyze social attributes and network structure of nodes to reduce energy consumption and improve transmission performance. Moreover, the security of data transmission in big data communication network is considered.

REFERENCES

[1] J. Wu, Z. Chen, and M. Zhao, "Effective information transmission based on socialization nodes in opportunistic networks," *Comput. Netw.*, vol. 129, pp. 297–305, Dec. 2017.

[2] M. Rajesh and J. M. Gnanasekar, "Path observation-based physical routing protocol for wireless ad hoc networks," *Wireless Pers. Commun.*, vol. 11, no. 3, pp. 1267–1289, Nov. 2017.

[3] S. Han, W. K. Ng, L. Wan, and V. C. S. Lee, "Privacy-preserving gradient-descent methods," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 6, pp. 884–899, Jun. 2010.

[4] C. Phua, K. Smith-Miles, V. Lee, and R. Gayler, "Resilient identity crime detection," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 3, pp. 533–546, Mar. 2010.

[5] J. M. Aliotta *et al.*, "Microvesicle entry into marrow cells mediates tissue-specific changes in mRNA by direct delivery of mRNA and induction of transcription," *Experim. Hematology*, vol. 38, no. 3, pp. 233–245, Mar. 2012.

[6] A. Y. Lokhov, M. Mézard, H. Ohta, and L. Zdeborová, "Inferring the origin of an epidemic with a dynamic message-passing algorithm," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 90, no. 1, Jan. 2014, Art. no. 012801.

[7] P. Guan and J. Wu, "Effective data communication based on social community in social opportunistic networks," *IEEE ACCESS*, vol. 7, pp. 12405–12414, 2019. doi: 10.1109/ACCESS.2019.2893308.

[8] S. Misra, T. Ojha, and A. Mondal, "Game-theoretic topology controlfor opportunistic localization in sparse underwater sensor networks," *IEEE Trans. Mobile Comput.*, vol. 14, no. 5, pp. 990–1003, May 2015.

[9] J. Wu, Z. Chen, and M. Zhao, "Information cache management and data transmission algorithm in opportunistic social networks," in *Proc. Wireless Netw.*, Feb. 2018, pp. 1–12.

[10] J. Wu, Z. Chen, and M. Zhao, "Weight distribution and community reconstitution based on communities communications in social opportunistic networks," *Peer-to-Peer Netw. Appl.*, vol. 12, no. 1, pp. 158–166, Jan. 2019. doi: 10.1007/s12083-018-0649-x.

[11] M. Grossglauser D. N. C. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Trans. Netw.*, vol. 10, no. 4, pp. 477–486, Aug. 2002.

[12] G. Wang, B. Wang, and Y. Gao, "Dynamic spray and wait routing algorithm with quality of node in delay tolerant network," in *Proc. Int. Conf. Commun. Mobile Comput.*, vol. 3, Apr. 2010, pp. 452–456.

[13] 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.*, Aug. 2005, pp. 252–259.

[14] J. Leguay, T. Friedman, and V. Conan, "DTN routing in a mobility pattern space," in *Proc. ACM SIGCOMM Workshop Delay-Tolerant Netw.*, Aug. 2005, pp. 276–283.

[15] J. Burgess, B. Gallagher, D. Jensen, "MaxProp: Routing for vehicle-based disruption-tolerant networks," *INFOCOM*, vol. 6, pp. 1–11, Apr. 2006.

[16] J. Leguay, M. Lopez-Ramos, K. Jean-Marie, and V. Conan, "An efficient service oriented architecture for heterogeneous and dynamic wireless sensor networks," in *Proc. 33rd IEEE Conf. Local Comput. Netw. (LCN)*, Oct. 2008, pp. 740–747.

[17] B. Burns, O. Brock, and B. N. Levine, "MORA routing and capacity building in disruption-tolerant networks," *Ad Hoc Netw.*, vol. 6, no. 4, pp. 600–620, Jun. 2008.

[18] V. Kavitha and E. Altman, "Analysis and design of message ferry routes in sensor networks using polling models," in *Proc. 8th Int. Symp. Modeling Optim. Mobile, Ad Hoc, Wireless Netw.*, May/Jun. 2010, pp. 247–255.

[19] J. Wu and Z. Chen, "Sensor communication area and node extend routing algorithm in opportunistic networks," *Peer-to-Peer Netw. Appl.*, vol. 11, no. 1, pp. 99–100, Jan. 2018.

[20] R. Wang *et al.*, "Social identity–aware opportunistic routing in mobile social networks," *Trans. Emerg. Telecommun. Technol.*, vol. 29, no. 5, p. E3297, May 2018.

[21] X.-H. Wu, X.-F. Gu, and S. Poslad, "Routing algorithm based on social relations in opportunistic networks," in *Proc. 12th Int. Comput. Conf. Wavelet Active Media Technol. Inf. Process.* Dec. 2016, pp. 146–149.

[22] J. Wu, Z. Chen, and M. Zhao, "SECM: Status estimation and cache management algorithm in opportunistic networks," *J. Supercomput.*, vol. 75, no. 5, pp. 2629–2647, May 2019. doi: 10.1007/s11227-018-2675-0.
[23] P. Nguyen and K. Nahrstedt, "Context-aware crowd-sensing in opportunistic mobile social networks," in *Proc.12th Int. Conf. Mobile Ad Hoc Sensor Syst.*, Oct. 2016, pp. 477–478.
[24] 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.
[25] A. Keränen, J. Ott, and T. Kärkkäinen, "The ONE simulator for DTN protocol evaluation," in *Proc. 2nd Int. Conf. Simulation Tools Techn.*, Mar. 2009, Art. no. 55.

JIA WU received the Ph.D. degree in software engineering from Central South University, Changsha, Hunan, China, in 2016, where he is currently an Associate Professor with the School of Software. Since 2010, he has been an Algorithm Engineer with IBM Company, Seoul, South Korea, and in Shanghai, China. His research interests include wireless communications and networking, wireless networks, big data research, and mobile health in network communication. He is a Senior Member of the China Computer Federation (CCF), and a member of the IEEE and ACM.

GENGHUA YU received the master's degree in information science and engineering from Central South University, Changsha, Hunan, China, in 2017. She is the 2017 outstanding graduate of Nanchang University. Her research interests include wireless communications and networking, big data research, wireless networks, and data mining.

PEIYUAN GUAN received the master's degree from the Software College, Central South University, Changsha, China, in 2016, where he is currently pursuing the Ph.D. degree in information science and engineering. His research interests include edge computing, the Internet of Things, and game theory. He is a Student Member of IEEE and CCF.