# An Effective Clustering Routing Algorithm Based on Social-Interest Similarity in Mobile Opportunistic Networks*

Feng Zeng
School of Software, Central South University
Changsha, Hunan, China 410075
fengzeng@csu.edu.cn

Jie Peng
School of Software, Central South University
Changsha, Hunan, China 410075
pengjie949794576@qq.com

Wenjia Li
Department of Computer Science, New York Institute of Technology
New York, New York, USA 10023
wli20@nyit.edu

## ABSTRACT

In this paper, we consider social attribute and interest jointly to measure the relation between nodes. First, the real-life phenomena are analyzed to show the necessity of the combination of nodes' social attribute and interest, and the definitions of social and interest similarity between nodes are given to show the degree of the relation between nodes. Then, we proposed an effective routing algorithm based on social-interest similarity. In the proposed algorithm, each node maintains a local cluster according to node similarity, keeps updating the regional cluster to ensure cluster members having the best node similarity with it, and forwards message only to the encounter nodes whose local cluster contains the destination node. At last, simulation is done. The results show that the proposed algorithm has better performance than other three algorithms on the whole.

## CCS CONCEPTS

• **Networks** → **Network algorithms**;

## KEYWORDS

mobile opportunistic network; routing; clustering; social relationship; interest similarity

**ACM Reference format:**
Feng Zeng, Jie Peng, and Wenjia Li. 2017. An Effective Clustering Routing Algorithm Based on Social-Interest Similarity in Mobile Opportunistic Networks. In *Proceedings of ACM conference, Chongqing, China, July 2017 (MOBIMEDIA2017),* 6 pages.
DOI: 10.475/123_4

## 1 INTRODUCTION

Mobile Opportunistic Networks [1] (MONs) can be formed by wireless portable devices such as iPad, PDA, smart phone, etc., which are usually carried by the human beings. Owing to the random mobility of nodes, there is no fixed link between any two nodes, and a completely connected path will never exist between the source and the destination. The node can only transfer the message based on storage-carry-and-forward data transmission mode, which means that the node can only opportunistically communicate with each other within the range of wireless coverage. When the destination node cannot be directly accessed, the node carrying the message will choose to forward messages to neighbor nodes, which have the opportunity to meet the destination. Obviously, the random mobility of nodes leads to the frequent disconnection between nodes, and the nodes should make full use of the links adjacent to find the next hop and forward the message with the expectation of reaching the destination quickly and accurately.

In MONs, since wireless devices are usually carried by human beings, the relationship between nodes is affected by many factors in our real lives, and the real-life interpersonal social relationships can be mapped to the MONs, determining the relationship between nodes based on human behavior and activities. For example, the nodes with the strong social relationship are more likely to meet each other than those without any social relation, which is shown to be correct in [2, 3]. However, the social relationship is not the only factor to catch people together. The common interest can also put an impact on the relationship between nodes, as the saying goes, "Like attracts like". Common interests between nodes improve the success rate of data transmission in MONs, as is shown in [4–6]. Consequently, in MONs, it is not only the social relationship but also interest between nodes having relation with the action of message forwarding. For instance, if a good friend is very interested in the message forwarded to him, he may be more positive forwarding the messages. On the contrary, even having a good social relationship with the message sender, he would not be happy to forward the messages which he is not interested in. Therefore, in order to improve the performance of opportunistic transmission between source and destination, we should take both social relationship and interest between nodes into considerations to design message forwarding scheme in MONs.

In this paper, we proposed an effective routing algorithm based on social-interest similarity. The proposed algorithm takes the interpersonal social relationship and the interest as the judgment basis to measure the similarity between two nodes. Then, based on the similarity value, the nodes with high similarity are clustered together, as is helpful to

*Drs. Feng Zeng and Wenjia Li are both corresponding authors for this paper. This work is supported by the Fundamental Research Funds for Central Universities of the Central South University(No.2017zzts611).



MOBIMEDIA 2017, July 13-14, Chongqing, People's Republic of China
Copyright © 2017 EAI
DOI 10.4108/eai.13-7-2017.2270328

facilitate selection of the next-hop in cluster-based routing. In our best knowledge, there is no related work considering social relationship and interest jointly to measure the similarity between nodes, and our work is the first time to design opportunistic routing scheme using social relationship and interest similarity together.

The rest of this paper is structured as follows. In section 2, we describe and analyze the related work. In section 3, we propose several definitions and computing methods to explain the node similarity in detail. In section 4, Routing Algorithm based on Social-Interest Similarity is to be submitted and analyzed. In section 5, simulation results are to be present. The last section concludes this paper.

## 2 RELATED WORK

In MONs, a lot of efforts had been devoted to the research of routing algorithms, and all these algorithms can be mainly classified into two categories: social oblivious routings and social characteristics aware routings [7]. For social oblivious routings, there are many classic algorithms presented. Such as Epidemic [8], Direct Transmission [9], and so on. In Epidemic [8], the nodes flooded the message to all nodes without consideration of routing overhead, and exchange all messages they carried with other nodes they met. Theoretically, Epidemic Routing had the highest success rate of data delivery, but at the price of the highest routing overhead. Direct Transmission [9] can be considered as the simplest routing, the messages carried by the source node can be only delivered to the destination node, which means the source node will keep moving until the destination node is met.

For social characteristics aware routings, the authors in [10] defined the interpersonal relationship of each node as blood relatives, friends and strangers, then used the trigeminal tree to represent the interpersonal relationships of each node and build the optimal dynamic cooperation tree. According to establish dependability, usability and decline factor to count the weight of this topology structure, finally obtain the optimal objects and paths. In [11], the authors determined the social relations of the nodes by the combination of three factors, which are the correlation degree of two nodes, the node similarity of two nodes and the node mobile connectivity degree between two nodes. The node's social relations can be measured based on these three factors. Finally, the best next hop was selected by sorting the values of the social relations of the neighbors.

There are also many routings considering the interests. In [12], the node interest structure was composed of self-interest and second-interest, and the self-interest and the second-interest information stored in the mobile node were used to predict the satisfaction probability of the future arrival destination node for the improvement of the delivery efficiency. In [13], the authors used the interest matrix and the message header to express individual interest and message type respectively. Based on similarity of comparison of the data type and personal interest between two nodes, the nodes with a higher similarity will be put into the corresponding interest communities. Based on the comparison of the similarity between data type and personal interest of two nodes, the nodes with a higher similarity will be put into the corresponding interest communities. With consideration to contact information between nodes, the authors proposed the routing strategies within or between the communities. In [14], the authors selected a suitable next hop according to the node's social relationship, the node's interest and the encounter history information, and then proposed a publish/subscribe routing scheme.

Different from the existing works, in this paper, we will consider the social relations and interest relations of nodes jointly for the selection of the next hop in message forwarding. As mentioned in section 1, both social relationship and interest between nodes has relation with the action of message forwarding. Taking both social relationship and interest between nodes into consideration will be effective to improve the success rate of opportunistic transmission. The proposed algorithm in this paper is a social characteristics aware routing. In our work, the definitions of social and interest similarity between nodes will be presented to show the degree of the relation between nodes, and the nodes with high similarity will be clustered together and become the choices of the next hop in message forwarding.

## 3 NODE SIMILARITY MEASUREMENT

We combine social relation and interest tightly to form node similarity, which may show the accurate relation between the nodes. In this section, after some assumptions are given, the definition of node similarity will be described, and the measurement method will also be discussed.

### 3.1 Assumptions

In section 1, we have mentioned that common interests have the impact on the relationship between nodes, and the authors in [4–6] found that the nodes with common interests are more likely to meet, and people often likes to join groups with similar interests, as is the first assumption in this paper. The second assumption is that the nodes with common friends will meet each other with a high probability, which is verified in [3].

With the two assumptions above, we suppose that there are $n$ types of interest in the network, each message has only one type of interest as its attribute, and each node also has its own interests. It is supposed the interest set is $I$, and $I = \{1, 2, \ldots, k, \ldots, n\}$. Each node may have multiple interests, to a node $s$, its interests can be denoted as a binary string $I_s = x_1^s x_2^s \ldots x_n^s$, where $x_i^s (1 \leq i \leq n)$ is a binary variable, the value of $x_i^s$ is 1 if node $s$ has the interest $i$, otherwise, the value of $x_i^s$ is 0.

### 3.2 Node similarity

Supposed each node has both social attribute related to others and the interests in message, we propose the measurement method of social similarity and interest similarity respectively.

From the above discussion, the degree of relation between nodes lies in two aspects, which are social and interest similarity. Representing the degree of relation between nodes, node similarity can be shown in Fig. 1.

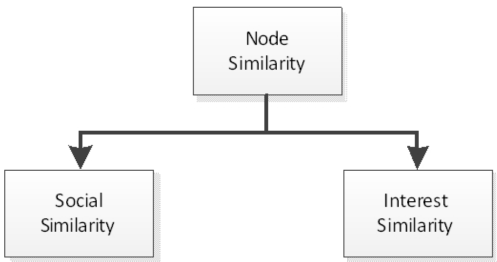


**Figure 1: Similarity structure of a node**

*Definition 3.1.* Social similarity is defined as the ratio of the number of common friends of two nodes (such as $A$ an $B$) to the total number of friends of both, as is shown in Equation 1.

$$Sim1(A,\ B) = \frac{N(A) \cap N(B)}{N(A) \cup N(B)} \tag{1}$$

Where, $N(A)$ on behalf of $A$'s friends, $N(B)$ on behalf of $B$'s friends, $N(A) \cap N(B)$ represents the number of common friends, and $N(A) \cup N(B)$ is the number of all friends of node $A$ and $B$.

*Definition 3.2.* Since the nodes' interests are the sets of binary variables, which are asymmetric binary attribute, the similarity of two nodes' interest can be measured by Jaccard coefficient shown in Equation 2.

$$Sim2(A,\ B) = \frac{q}{q+r+s} \tag{2}$$

Supposed the interests of $A$ and $B$ are $x_1^A x_2^A \dots x_n^A$ and $x_1^B x_2^B \dots x_n^B$ respectively, the $q$, $r$, and $s$ are defined in Equation 3, 4 and 5 respectively. In Equation 2, $q$ is the number of common interests of two nodes, $r$ represents the number of interests which are only for $A$, and $s$ is the number of interests only for $B$.

$$q = \mid \{(x_i^A, x_i^B) \mid x_i^A = 1, x_i^B = 1, 1 \le i \le n\} \mid \tag{3}$$

$$r = \mid \{(x_i^A, x_i^B) \mid x_i^A = 0, x_i^B = 1, 1 \le i \le n\} \mid \tag{4}$$

$$s = \mid \{(x_i^A, x_i^B) \mid x_i^A = 1, x_i^B = 0, 1 \le i \le n\} \mid \tag{5}$$

*Definition 3.3.* According to the above definitions, node similarity between node $A$ and $B$ is shown in Equation 6.

$$Sim(A,\ B) = \alpha Sim1(A,B) + \beta Sim2(A,B) \tag{6}$$

Where, $\alpha \in [0,1]$, $\beta \in [0,1]$, and $\alpha + \beta = 1$.

# 4 ROUTING ALGORITHM BASED ON SOCIAL-INTEREST SIMILARITY

In order to improve the success rate of data transmission in MONs, we propose a Routing Algorithm based on Social-Interest Similarity (RASIS), which includes two parts, the cluster construction and update, and the message delivery.

## 4.1 Cluster construction and update

### 4.1.1 Cluster construction

. In MONs, each node maintains a cluster locally, which is used to collect the nodes with high similarity to it, so as to find the suitable next hop in message forwarding. When meeting the other node, the node searches the local cluster to find whether the met node is in its local cluster. If the met node exists in the local cluster, then it is ignored. Otherwise, the node similarity of two nodes is computed according to the social and interest relations of two nodes based on Equation 6. Then given that $\delta \in [0,1]$ is the threshold to determine whether a node will be added to the cluster, if the similarity of the two nodes is greater than the threshold $\delta$, the node is added to the local cluster.

### 4.1.2 Cluster update

. Because of the movement of nodes and the meetings between the nodes are random, the social relationship and interests of the nodes will change with time. Therefore, in order to keep high similarity with the nodes in the local cluster, we define the change degree of the nodes, which is used to measure the change degree between two nodes over time. In order to calculate the change degree of nodes, we introduce the change degree of common friends and interests. Specific definitions are shown as follows.

*Definition 4.1.* The change degree of the common friends of two nodes ($A$ and $B$) is intensity of variation of the common friends between $A$ and

$$C1(A,\ B) = |Sim1_{old}(A,B) - Sim1_{new}(A,B)| \tag{7}$$

In Equation 7, $Sim1_{old}(A,B)$ and $Sim1_{new}(A,B)$ represent the social similarity between $A$ and $B$ before $T$ time interval and social similarity between $A$ and $B$ after $T$ time interval respectively.

*Definition 4.2.* Similar to the change degree of common friends, the change degree of the common interests is intensity of variation of the common interests between $A$ and $B$ in a period of time ($T$), which is calculated as Equation 8.

$$C2(A,\ B) = |Sim2_{old}(A,B) - Sim2_{new}(A,B)| \tag{8}$$

In Equation 8, $I_{old}(A,B)$ and $I_{new}(A,B)$ represent the social similarity between $A$ and $B$ before $T$ time interval and the social similarity between $A$ and $B$ after $T$ time interval respectively.

*Definition 4.3.* According to the above two definitions, we can get the change degree of the nodes during $T$ time interval, as is shown in Equation 9.

$$C(A,\ B) = \alpha C1(A,B) + \beta C2(A,B) \tag{9}$$

Where, $\alpha \in [0, 1]$, $\beta \in [0, 1]$, and $\alpha + \beta = 1$.

In Equation 9, we define the change degree of the nodes, and according to the the definition, we propose a **mechanism of cluster updating**:

We make the local cluster of nodes update every $T$ time interval in RASIS, in order to ensure the nodes in the cluster are with high similarity to the node that owns the cluster. Every time cluster update, the change degree of nodes is calculated as Equation 9. Given that $\tau \in [0, 1]$ is the threshold to determine whether a node will be removed from the cluster, if the value of a node in the cluster is greater than $\tau$, the node will be removed from the cluster.

## 4.2 Message delivery

It is supposed that node $S$ carries the messages(a message list) sent to node $D$ and every node has a local cluster which was illustrated When node $S$ encounters node $E$ at time $t$, actions will happen as follows. The delivery will be divided into 2 steps: (1) Request for the node $E$'s local cluster; (2) Determine the relationship between the node E and the destination node D, and then decide whether or not to transmit the message to the E.

**Algorithm 1** Message Delivery Algorithm of RASIS

```
When S meet E
if NotInLocalCluster(E) then
    RequestLocalClusterOf(E)
end if
while LocalCluster.hasNext() do
    getCurrentMessage as m
    if DestinationOf(m)=E then
        Deliver(m,E)
        Delete(m)
    else if E.LocalCluster contains DestinationOf(m) then
        CopyAndDeliver(m,E)
    end if
    move m to next message in the message list
end while
```

# 5 SIMULATIONS

## 5.1 Simulation platform and Experimental steps

The simulation platform used in this paper is MobEmu [3], the same as [14]. MobEmu is an opportunistic network simulator that can replay the encounter trajectory of a node based on a data set and apply the required routing algorithm when the nodes meet, and finally collects and gives the statistics of the experimental results. All the experiments in this paper are based on the real data set $UPB$ [15]. The data set $UPB$ contains 66 nodes and 5 interest types, with a duration of 64 days, and can be downloaded from CRAWDAD (http://www.crawdad.org/).

We compare the performance of the proposed algorithm RASIS with Epidemic, BubbleRap, ONSIDE, and the following performance metrics are used in the performance comparison.

(1) $HitRate$: the ratio of the number of messages arrived at the destination to the total number of messages sent.
(2) $DeliveryCost$: the ratio of the total number of messages exchanged to the number of messages generated.
(3) $DeliveryLatency$: the average amount of time passed when messages transmitted from source to destination.
(4) $HopCount$: the number of intermediate nodes in the shortest path from source to destination for a successful transmission.

## 5.2 Results comparison

In the experiments, we compared the performance of the four routings in the same conditions using the dataset $UPB$. The experimental results are shown in Fig. 2 to Fig. 5.

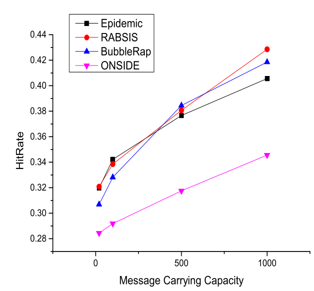


**Figure 2: HitRate**

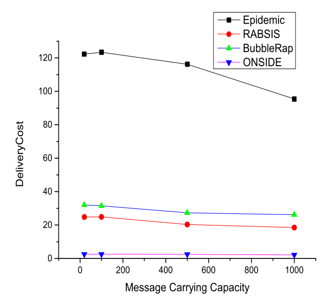


**Figure 3: DeliveryCost**

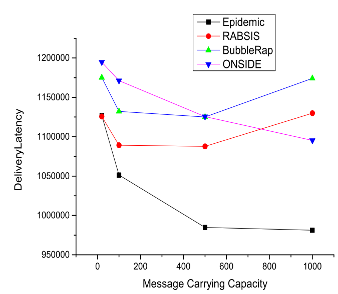


**Figure 4: DeliveryLatency**

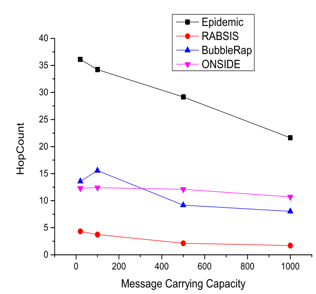


**Figure 5: HopCount**

As can be seen from the Fig. 2, the $HitRate$ of the four algorithms increases as the message carrying capacity increases. Compared with Epidemic, BubbleRap and ONSIDE, RASIS has the $HitRate$ increased by 1.2%, 2.1%, 15.2% respectively.This means that after the social and interest are taken into account, the predictions of the relationships between nodes are more accurate(this will be confirmed in Fig. 5), resulting in higher success rates.

The comparison of $DeliveryCost$ is shown in Fig. 3. Due to the calculation of node similarity and maintenance of local clusters requiring a process, compared with ONSIDE, RASIS has the $DeliveryCost$ increase by 16%. However, ONSIDE's low delivery cost is at the expense of low message delivery success rate. Therefore, considering the message passing success rate and delivery overhead, we think RASIS is more advantageous because it have a higher level of success in delivering messages. And compared with Epidemic and BubbleRap, RASIS has the $DeliveryCost$ decreased by 34% and 8% respectively.

The comparison of $DeliveryLatency$ is shown in Fig. 4. Since nodes in Epidemic Routing flooded the messages to all nodes, the delivery latency will decrease rapidly when the message carrying capacity increase. For this reason, compared with Epidemic, RASIS has the $DeliveryLatency$ increased by 8%. But compared with ONSIDE and BubbleRap, the RASIS has the $DeliveryLantency$ decreased by 3.7% and 3.2% respectively.

The comparison of $DeliveryLatency$ is shown in Fig. 5. Because of the nodes in RASIS choose the next hop which has high similarity to destination node, the $HopCount$ is lower than others, which means our node similarity measurement is effective on the judgment of the relationship between nodes. Compared with Epidemic, BubbleRap and ONSIDE, RASIS has the $HopCount$ decreased by 88%, 28.3%, 30.9% respectively.

## 5.3 Impact of parameters on RASIS

In the experiment for the selection of parameters, we use the same $\alpha$ and $\beta$ in determining node similarity and the change degree of the nodes. In order to choose the appropriate value for $\alpha$, we select different values of $\alpha$ for testing, and we chose the experimental results of a better experiment, which are shown in Figure 6 to 9.

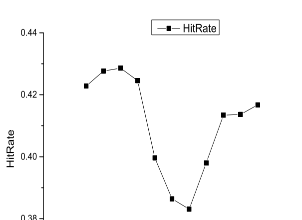


**Figure 6: HitRate with various $\alpha$**

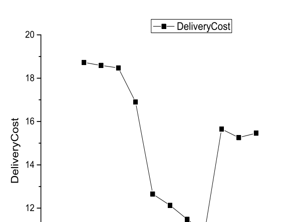


**Figure 7: Message delivery cost with various $\alpha$**

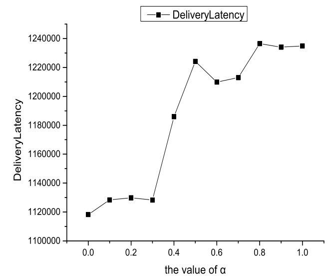


**Figure 8: Message delivery latency with various $\alpha$**

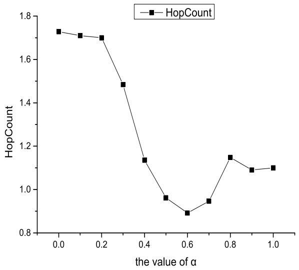


**Figure 9: HopCount with various $\alpha$**

In the test, we set the message capacity to 1000. From Fig. 6 to Fig. 9, we can see that RASIS has the better performance of $HitRate$ and $DeliveryLatency$ when $\alpha$ is about 0.2 than other values of $\alpha$. But, the $DeliveryCost$ and $HopCount$ are not in the best situation when $\alpha$ is about 0.2. Although 0.2 is not the best in every respect, but it is in the HitRate and message delivery lantency aspects of the performance is indeed very prominent. From Fig. 2 to Fig. 5, we can see that RASIS in $DeliveryCost$ and $HopCount$ has great advantage. Therefore, in order to make RASIS more advantageous in terms of $HitRate$ and $DeliveryLatency$, we chose 0.2 that will not be optimal for $DeliveryCost$ and $HopCount$. However, even so, with $\alpha = 0.2$, RASIS has the better performance of $DeliveryCost$ than BUBBLE and Epedemic shown in Fig. 3, and has the best performance shown in Fig. 5 in term of $HopCount$. Consequently, in simulation, we chose 0.2 as the value of $\alpha$ and 0.8 for $\beta$. In some situations, we can choose different values of $\alpha$ to highlight distinctive advantages of RASIS.

# 6 CONCLUSION

In this paper, we consider the social relations and interest relations of nodes jointly for the selection of the next hop in message forwarding, and propose a Routing Algorithm based on Social-Interest Similarity (RASIS) in MONs. In RASIS, each node maintains a local cluster, decides whether to let a node join in its local cluster according to node similarity, and keeps updating the local cluster to ensure cluster members having the best similarity with it. When delivering messages, the first thing is to identify the relationship between the encounter node and destination node, if the cluster of encounter node contains the destination node, the message will be forwarded to the encounter node. Experimental results show that, compared with other algorithms, RASIS has a high level of $HitRate$, and its $DliveryLantency$ and $HopCounts$ are reduced.

# REFERENCES

[1] Leszek Lilien, Zille Huma Kamal, Vijay Bhuse, and Ajay Gupta. Opportunistic networks: the concept and research challenges in privacy and security. 2007.

[2] Chiara Boldrini, M Conti, and A Passarella. Impact of social mobility on routing protocols for opportunistic networks. In *World of Wireless, Mobile and Multimedia Networks, 2007. WoWMoM 2007. IEEE International Symposium on a*, pages 1–6, 2007.

[3] Radu Ioan Ciobanu, Ciprian Dobre, and Valentin Cristea. Social aspects to support opportunistic networks in an academic environment. In *International Conference on Ad-Hoc, Mobile, and Wireless Networks*, pages 69–82, 2012.

[4] Arezu Moghadam and Henning Schulzrinne. Interest-aware content distribution protocol for mobile disruption-tolerant networks. In *World of Wireless, Mobile and Multimedia Networks and Workshops, 2009. WoWMoM 2009. IEEE International Symposium on a*, pages 1–7, 2009.

[5] Alessandro Mei, Giacomo Morabito, Paolo Santi, and Julinda Stefa. Social-aware stateless forwarding in pocket switched networks. In *INFOCOM, 2011 Proceedings IEEE*, pages 251–255, 2011.

[6] Paolo Costa, Cecilia Mascolo, Mirco Musolesi, and Gian Pietro Picco. Socially-aware routing for publish-subscribe in delay-tolerant mobile ad hoc networks. *IEEE Journal on Selected Areas in Communications*, 26(5):748–760, 2008.

[7] Faisal Naeem, Sahibzada Ali Mahmud, and Mohammad Haseeb Zafar. Social interest-based routing in delay tolerant networks. In *International Conference on Emerging Technologies*, pages 1–5, 2015.

[8] Amin Vahdat and David Becker. Epidemic routing for partially-connected ad hoc networks. *Master Thesis*, 2000.

[9] Matthias Grossglauser and D. N. C Tse. Mobility increases the capacity of ad hoc wireless networks. In *INFOCOM 2001. Twentieth Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, pages 1360–1369 vol.3, 2001.

[10] W. U. Jia, Y. I. Xi, Zhi Gang Chen, School Of Software, and Central South University. Algorithm of selecting optimal dynamic cooperation tree in opportunistic networks. *Computer Science*, 2014.

[11] Xiaohua Wu and Song Chen. Identify and measure social relations: Routing algorithm based on social relations in opportunistic networks. In *IEEE International Conference on Computational Science and Engineering*, pages 407–412, 2014.

[12] Liu Li. Interest-based prediction routing protocol in socially-aware opportunistic networks. In *International Symposium on Instrumentation and Measurement, Sensor Network and Automation*, pages 371–374, 2013.

[13] Qilie Liu, Chunfeng Hu, Yun Li, and Weiliang Zhao. An interest community routing scheme for opportunistic networks. pages 4366–4371, 2013.

[14] Radu Ioan Ciobanu, Radu Corneliu Marin, Ciprian Dobre, and Valentin Cristea. Onside: Socially-aware and interest-based dissemination in opportunistic networks. 59(10):1–6, 2014.

[15] Radu Corneliu Marin, Ciprian Dobre, and Fatos Xhafa. Exploring predictability in mobile interaction. In *Third International Conference on Emerging Intelligent Data and Web Technologies*, pages 133–139, 2012.