Modeling homogeneous contact distribution of nodes and its application in routing in Mobile Social Networks

Saeed Moradi[1], Jamshid Bagherzadeh Mohasefi[2,*], Ebrahim Mahdipour[1]
[1] Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran
[2] Department of Computer Engineering, Urmia University, Urmia, Iran

Emails: saeed-moradi@srbiau.ac.ir, j.bagherzadeh@urmia.ac.ir, mahdipour@srbiau.ac.ir
*Corresponding author: Tel: +098 4431124202

Received: 11-04-2021, Revised: 29-01-2022, Accepted: 18-04-2022.

Abstract
Different types of contact, including contact between node pairs, any-contact of nodes, and contacts of the entire network, are used to characterize social relations in mobile social networks. Different modes of routing, from the point of view of message delivery semantics, encompass unicasting, multicasting, any-casting, and broadcasting. Studies have shown that using probability distribution functions of contact data, which is mainly assumed to be homogeneous for nodes, improves the performance of these networks. However, there exists an important challenge in studies on distributions. A lot of works apply the distribution of one type of contact to other types. Hence in routing applications, it causes to use of the distribution of one type of contact for any mode of routing. This study provides a complete solution to model each type of homogeneous contact data distribution and to use them in different modes of routing. We propose a routing algorithm that uses this new model. Results show that our solution improves the average latency of comparing methods Epidemic, TCCB, and DR about 3.5-times, 30%, and 45%, respectively. It achieves a delivery rate of about 5% and 6%, and average latency about 6% and 8% better than that of DR and TCCB, respectively.

Keywords: Mobile social networks, types of contact, homogeneous contact, contact data distribution, modeling of contacts.

1. Introduction

The application of various telecommunication capabilities and powerful and lightweight means approximates the vision of "ubiquitous networks" to reality. Mobile Social Networks (MSN) is one of the solutions, which applies a wide range of communication means, (e.g., smartphones and mobile computers) with processing capability. MSN uses opportunistic communication, delay-tolerant networks (DTN) architecture, and the store-carry-forward mechanism [2,4].

MSN is used in many cases. For example, where there is no fixed network infrastructure, governments have cut it off, it is destroyed by natural causes such as floods and earthquakes, or there is a possibility of eavesdropping or filtering. The techniques of social networking are being applied in the field of communication and information technologies to provide efficient solutions for content exchange, and also delivery services [25]. MSN is also used to leverage cellular links by offloading mobile traffic via device-to-device communications. Another application of MSN is its role in 5G networks as 5G MSN to handle the highly increasinging content demand of mobile users [16].

In the MSN, the network is permanently prone to fragmentation due to unstable and disrupted links and frequent disconnections. When two nodes come in the communication range of each other and are connected via wireless links, we can say they are in contact. Different contact types are used to characterize contact and social relations in the MSN literature [4,10]. We categorize them using our notation as follows:

- *Node pairs contact (inter-contact) times /frequency*: Hereafter we call this type of contact CTYPE1 and define it as follows. If we assume *i* and *j* denote node pairs from the set of nodes N of the network, then a *contact time* of the node pairs *(i,j)* is defined as the duration of contact between them. All *contact times* that occur between node pairs *(i,j)* are denoted as *CT(i,j)*. Similarly, *inter-contact time* is defined as the duration between two consecutive contacts of the same node pairs. We denote *inter-contact times* of node pairs *(i,j)* as *ICT(i,j)*. *Contact frequency* and *inter-contact frequency* of node pairs *(i,j)*, which show the number of contacts and inter-contacts, are denoted as *CF(i,j)* and *ICF(i,j)*, respectively.
- *Any-contact (any inter-contact) times/frequency (*CTYPE2*)*: We define the *any-contact time* of a node *i* as the duration which node *i* encounters any node belonging to a specific subset of nodes. All *any-contact times* that occur between node *i* and those nodes are denoted as *ACT(i)*. Similarly, *any-inter-contact times*, *any-contact frequency,* and *any inter-Contact frequency* of node *i*, are denoted as *AICT(i)*, *ACF(i)*, and *AICF(i)*, respectively.
- *Entire network contact (inter-contact) times/ frequency (CTYPE3)*: we define the *entire network contact time* of the network as the duration of

contacts between all nodes. All contact times that occur between all nodes of the network are denoted as *ENCT*. Similarly, *entire network inter-contact times*, *entire network contact frequency*, and *entire network inter-contact frequency* are denoted as *ENICT*, *ENCF,* and *ENICF*, respectively.

Because the latency and long queues are the main features of this network, taking advantage of contact opportunities is crucial. Accordingly, routing attracts special attention in this type of network [3,4].

Although a lot of efforts work on congestion control and buffer management [5,6], their goal is to improve network performance in packet forwarding and routing. To optimize performance, [6] and some works in [5] try to control the workload of the network by limiting the number of copies, but they neglect to take into account the contacts information.

The dominant routing solution in MSN is called "social-aware routing" [8]. Methods of this solution utilize different social features to make decisions in the forwarding of messages. The main object of routing includes maximizing the delivery rate while minimizing the overhead and the delivery latency [9].

The social features of nodes can be achieved in two ways. Some of them (e.g., interest, nationality, and language) are obtained from the registered profile of the nodes. Other features (e.g., similarity, centrality, trust, tie strength, closeness, and distance) can be calculated from the contact information of the nodes. Thus, nodes log their contact information and that of others. In TCCB [26] forwarding procces, using the average and variance of *ICT* in a certain period, a metric called temporal closeness is applied to forward messages as the first step. As the next step, using the temporal closeness of node pairs, a metic called temporal centrality is applied taking into account all nodes in the network. The message is forwarded to nodes with higher temporal closeness with the destination.

It is believed that using contacts probability distributions, more accurately, describes and predicts important quantities such as contact time, inter-contact time, the number of contacts, and their related social features [1,12,13]. Many studies in the literature use probability distribution functions for CTYPEs [4,14,15]. On the other hand, from the point of view of message delivery semantics, there are various routing modes in the literature [8][9] as follows:

- Unicast concentrates on the forward messages to a single destination.
- Multicast involves the distribution of a message to a group of nodes.
- Anycast focuses on the forward messages to any member of a group of nodes.
- Broadcast: it deals with the distribution of messages to entire network nodes.

Some works such as [17] assumed regularity in mobile node movements and sought to predict their location and time to design a routing algorithm. Nevertheless, the assumption of order in motion is not applicable in all circumstances. Therefore, exploiting the probability distribution functions in estimating the ON/OFF periods of links and contacts is considered in various wireless mobile networks [18]. In [7], the authors investigated a delay-tolerant offloading solution and proposed an optimization mechanism that utilizes user movement pattern prediction to formulate a model obtaining the optimal placement of complimentary WiFi networks and their bandwidth allocation. Since the MSN can help cellular links by offloading mobile traffic via device-to-device communications, such mechanisms can benefit from the capabilities of this network.

In [1], a general framework based on semi-Markov processes is proposed for modeling delivery in opportunistic networks. The authors consider the exponential, power-law, and hyper-exponential distributions for *ICT* and neglect *CT* in modeling. Authors in [14] conduct a multicasting method using a weighted graph of network contacts. They assume exponential distribution for nodes *ACT* data based on CCDF of the *ENCT* data. Work [11] addresses the problem of relay selection for multicasting data using a centrality metric defined as the Cumulative Contact Probability. It is assumed Poisson process for average *CF* and exponential distribution for *ICT*. Article [15] confirms that power-law distribution with exponential cut-off better models the beginning of the *CT*/*ICT*. The *ICT* is used for community detection and the similarity and friendship features in the forwarding stage.

Work [16] which provides a comprehensive survey of influential nodes discovering methods in MSNs, is in accordance with the power-law assumption for *ICT* data. Label [10] shows using a small label to identify users affiliation community, bring a large improvement in forwarding performance. The authors assume power-law distribution for *ICT*. Article [4] uses social features similarity, centrality, friendship, social strength, and trust each with a variable coefficient as the utility function. It divides the nodes into triple communities, including nodes with high cumulative *CT* and *ICT*, low cumulative *CT* and *ICT*, and the rest of the nodes. The distributions of *contact* and *inter-contact times* of all *CTYPEs* are assumed heavy tail. The Authors in [27] recognized that there is not always a direct relation between *ENICT* and *ICT*. They realize that heavy-tail *ENICT* can emerge also from exponential *ICT* when their rates have a certain characteristic. Authors in [19] exploit the transient node contact patterns to improve data forwarding in DTN. They apply Poisson, exponential, normal distributions to *CF*, *ICT*, and *CT*.

Some articles have examined the distribution of a node's contact with communities. Authors of [28] propose a novel zero-knowledge multi-copy routing algorithm, homing spread (HS), for MSNs. They assume that inter-contact time between any two nodes and between a node and a community home follows independent and identical exponential distributions, respectively. From the point of view of contact distribution, this case can be considered as a type of multi or any–contact. Here, for the sake of brevity, we abstain to bring more references. To summarize, we compare some previous works in Table I.

With recent discussion, it should be clear to some extent that each CTYPE is applicable for a specific routing mode. For example, the CTYPE1 should be used in unicast, the CTYPE2 in anycast, and the CTYPE3 in the broadcast. Therefore, proper use of CTYPEs for routing modes is highly important to improve the performance of

MSN. Hence, it is necessary to provide a model for how to properly extract the contact data distribution and generalize them correctly. Because the application of distributions is the same for anycast and multicast, hereafter we categorize them into the anycast.

However, major challenges remain. Some works generalize the distribution model of one type of contact to other types and use them in several applications such as routing and community detection.

Table I. Comparing some previous works.

Work	Contact model stated	Distribution applied	Generalization	Challenges
[1]	*ICT*	power law, exponential, hyper exponential	No	- neglects CTYPE1(*CT*) in modeling - does not model other CTYPEs
[14]	*CF*, *ICT*	Poisson, exponential	from *ENCT*	- ignores CTYPE1(*CT*) in modeling - does not model other CTYPEs - generalization is not justified
[4]	*CT*, *ICT*, *ACT*, *AICT*	heavy tail	by analysis	- does not model CTYPE3 - generalizations are not justified
[27]	*ICT*, *ENICT*	exponential, heavy tail	No	- does not provide any model for non-generalization - does not model CTYPE1(*CT*), CTYPE2 and CTYPE3(*ENCT*)
[11]	*CF*, *ICT*	Poisson, exponential	No	As work [14]
[15]	*ENCT*, *ENICT*	power-law with exponential cut off	to *CT* and *ICT*	- does not model other CTYPEs - generalizations are not justified
[19]	*CF*,*ICT*,*CT*	Poisson, exponential, normal	from *ENCF*, *ENICT*, *ENCT*	- does not model CTYPE2,3 - generalizations are not justified
[10]	*ICT*, *AICT*, *ENICT*	power law	from *ENICT*, *by analysis*, by analysis	- does not model CTYPE1(*CT*), CTYPE2(*ACT*), and CTYPE3(*ENCT*) - generalization is not justified

For example, some works assume that the inter-contact time distribution of CTYPE3 follows the exponential or Pareto distribution. Then, they apply the same model to CTYPE2 and even to CTYPE1. It causes to use of broadcast data for any- and unicast purposes. The same challenge exists with other contact types. This important issue has not been addressed in the literature, and we try to clarify it.

Moreover, we provide a complete solution for the above-mentioned challenge. Eventually, this study tries to show the implication of this approach through a unicast routing method.

The following contributions can be mentioned:

A. Challenging the abusive generalization of the probability distribution of different CTYPEs to each other and their application in different routing modes.

B. Modeling different types of contacts with homogeneous assumption among nodes, using sums of several random variables theory.

C. Applying the proposed model to various routing modes.

D. Evaluation of the impact of contact distribution on network performance by using a routing method.

The remaining parts of the study are as follows: Section 2 describes the suggested model. Section 3 covers the proposed routing algorithm. Performance evaluations and results are presented in Section 4. Conclusions and future studies are discussed in Section 5. Finally, Section 6 presents the references.

2. Proposed model for different contact types and routing modes

Based on the different contact types discussed, we can be claim that there is an important issue in the literature, which should be addressed. The studies mainly obtain the distribution of one type of contact and generalize it to other types. Accordingly, these methods assume that the probability of observing a particular node by another node is equal to the probability of seeing any neighbours of that node, also to the probability of seeing each node by any other node in the network. In the other words, they calculate different probabilities using the same distribution function.

In this section, we explain our solution on how to model contact types and generalize the distribution of them to each other and use them in different routing modes, assuming homogenous distribution. To model the contact data, we classify the probability distribution functions into three categories:

Class1 includes functions in which the distribution of the sums of several random variables is of a known distribution function with specific parameters.

Class2 includes functions in which the distribution of the sums of several random variables is the same type as the constituent random variables function.

Class3 includes functions in which the distribution of the sums of several random variables is neither of the same type nor of a known distribution function.

2.1. *Modelling CTYPEs data for unicast routing*

For unicast routing, we propose a criterion, $CloDis_{ij}^{t}$. Given the independence of *contact times* and *inter-contact times of CTYPE1*, *CTYPE2*, *and CTYPE3*, we define the *CloDis* as follows:

$$CloDis(contacts, distributions\ parameters) = (\mu_C, \sigma_C, \tau_C, v_C, \mu_{iC}, \sigma_D, \tau_D, v_D) \quad (1)$$

Where C and D represent the random variables of concerned *contact* and *inter-contact times.* Further, μ, σ, τ, v represent mean, standard deviation, kurtosis, skewness, or other parameters of the related distribution. The letters C and D denote closeness/distance abbreviated in the term *CloDis*.

Furthermore, we define an appropriate definition of *CloDis* for each routing mode. For instance, $CloDis_{ij}^{t}$ holds pairs of nodes i and j contact distributions parameters used in unicast routing, $CloDis_i^t$ holds any-contact distribution parameters of node i used in anycast routing, and $CloDis^t$ holds distribution parameters of entire network contacts used in broadcast routing, in the interval including time t.

Several different states occur to model CloDis_{ij}^{t}, given that only CTYP1 data are available or distributions of other routing modes. If only CTYPE1 data is available, we model it by obtaining its distribution function using best fitting methods. We do not bring here the best fitting methods. However, if distribution functions of other routing modes are available, we can use the solutions described in Sections 2.2 (*i*), 2.2(*ii*), 2.3(*i*), and 2.3(*ii*) that are based on sums of several random variables theory. Modeling of $CloDis_i^t$ and $CloDis^t$ are given below.

2.2. *Modeling CTYPEs data for anycast routing*

This mode of routing is intended to compute $CloDis_i^t$ for each node i in our approach. Obtaining anycast contacts distribution is possible in the following ways:

- Best fitting CTYPE2 data.
- Generalizing of unicast/broadcast contacts distribution if possible.

if we have the unicast/broadcast contacts distribution in hand, we can use the sums of several random variables theory to obtain anycast distribution function while the unicast/broadcast distribution is of the Class-1 or 2. Otherwise, if the distribution is of Class-3, we have to obtain it using best fitting methods. So, depending on the distributions Class, three following states occur. We use character '/' instead 'and/or' in such cases. It is necessary to mention that if the unicast/broadcast distribution is not available, a best fitting method should be applied to CTYPE2 data.

i. Unicast/broadcast distribution belong(s) to Class-1:

In this case, $CloDis_{ij}^{t}$ of every node j belonging to anycast destinations nodes subset and/or $CloDis^t$ are available to node i and we should obtain the $CloDis_i^t$.

Without loss of generality, suppose *ICT(i,j)* follows an exponential distribution for every node i and j, with the mean parameter $1/\lambda$. For node i, we show the set of j indexes as $A_i = \{a_1, a_2, a_3, .., a_s\}$. Therefore, we can take the random variable of the *AICT(i)*, U_i, equal to the sums of independent identical random variables of $ICT(i,j)_{j\epsilon A}$ and express it as follows:

$$U_i = U_{ia_1} + U_{ia_2} + U_{ia_3} + \dots + U_{ia_s} \quad (2)$$

To calculate U_i, we use the torque generating function as follows:

$$M_{U_i}(t) = M_{U_{ia_1}}(t) M_{U_{ia_2}}(t) \dots M_{U_{ia_s}}(t) = \int_0^{\infty} e^{-tU_{ia_1}} \lambda e^{-\lambda U_{ia_1}} dx \dots \int_0^{\infty} e^{-tU_{ia_s}} \lambda e^{-\lambda U_{ia_s}} dx = \frac{\lambda}{\lambda - t} \frac{\lambda}{\lambda - t} \dots \frac{\lambda}{\lambda - t} = \frac{\lambda^s}{(\lambda - t)^s} \quad (3)$$

This is the Gamma random variable torque generator function with parameters (s, λ) with the following density function:

$$f(u_i) = \lambda e^{-\lambda u_i} \frac{(\lambda u_i)^{s-1}}{(s-1)!} \quad , \quad u_i \geq 0 \quad (4)$$

It means sums of several independent random variables of an exponential distribution is a Gamma distribution. Thus, if we want to generalize unicast exponentially distributed data to anycast, we should take a Gamma distribution, not an exponential one.

To obtain the *ACF(i)* distribution, with random variable R_i, without loss of generality, we assume Geometric distribution with parameter p for independent identical random variables of $CF(i,j)_{j\epsilon A}$. Therefore, R_i is obtained from the following relation:

$$M_{R_i}(t) = M_{R_{ia_1}}(t) M_{R_{ia_2}}(t) .. M_{R_{ia_s}}(t) = \quad (5)$$

$$\frac{pe^t}{1-(1-p)e^t} \dots \frac{pe^t}{1-(1-p)e^t} = \left(\frac{pe^t}{1-(1-p)e^t}\right)^s$$

This is the generator function of negative binomial distribution torques with the (p,s) parameters. Similar inferences can be made for *ACT*(i) and *AICF*(i). It should be noted that in this case, the resulting distributions are not the same as the distribution of primary (constituent) random variables.

Here, two conclusions may be obtained. First, the reverse of the presented solution can be used to generalize the distribution of anycast data to unicast. Second, the distribution of anycast contacts can also be obtained from broadcast distribution. We postpone the generalization solution to Sections 2.3(i) and 2.3(ii).

ii. Unicast/broadcast distribution belong(s) to Class-2:

Without loss of generality, we assume *CT*(i,j) follows normal distribution $U_{ij} \sim N(\mu_{ij}, \sigma_{ij}^2)$ and A_i is the set of anycast destinations of node i. The sums of participating random variables, U_i, follows normal distribution $U_i \sim N(\mu_i, \sigma_i^2)$ with the following parameters:

$$\mu_i = \sum_{j=a_1}^{a_s} \mu_{ij} , \ \sigma_i^2 = \sum_{j=a_1}^{a_s} \sigma_{ij}^2 \quad (6)$$

It can be proved through relations similar to section 2.2 (i). We see in this case, the resulting function for *ACT*(i) is the same as the primary functions. Also, if we represent $CF(i,j)_{j\epsilon A}$ with the Poisson distribution Q_{ij} with the parameter of λ_{ij}, the sums of the participating random variables, Q_i as the *ACF*(i), follow a Poisson distribution as well. The density function of Q_i can be expressed as follows:

$$g(q_i) = \sum_{j=a_1}^{a_s} \lambda_{ij} * \exp(-q_i \sum_{j=a_1}^{a_s} \lambda_{ij}) \quad (7)$$

A similar proof can be given for *AICT*(i) and *AICF*(i). It is clear that if the distribution of unicast contacts is of Class-2 functions, it can be applied to anycast contacts and vice versa. As we will see in Section 2.3(*ii*), this relationship is also established between broadcast and anycast modes.

iii.Unicast/broadcast distribution belong(s) to Class-3:

In this case, the distribution of unicast//broadcast random variables cannot be generalized to sums of anycast participating random variables. Moreover, the methods of previous parts of this section are not applicable either. Accordingly, only a best fitting method should be applied.

2.3. *Modeling CTYPEs data for broadcast routing*

In our approach, we need to calculate $CloDis^t$ for this mode of routing. Thus, depending on which class the distribution *Unicast/anycast* belongs to, three following states occur. It is necessary to mention that if the unicast/broadcast distribution is not available, a best fitting method should be applied to CTYPE3 data.

i. Unicast/anycast distribution belong(s) to Class-1:

In this case, broadcast contacts distribution that is related to $CloDis^t$, can be obtained using any of the unicast and anycast contacts distributions. Through unicast distributions, with similar assumptions of Section 2.2(i) about *CT*(i,j) and the number of network nodes m, we can show the entire network contacts random variable, U, as follows:

$$U = \sum_{i=1}^{m-1} \sum_{j=2}^{m} U_{ij} \quad (8)$$

Therefore, equation (4) changes as follows:

$$f(u) = \lambda e^{-\lambda u} \frac{(\lambda u)^{m(m-1)/2}}{((\frac{m(m-1)}{2})-1)!} \quad ,u \geq 0 \quad (9)$$

In relation (5), the parameter s changes to $m(m-1)/2$, and concerned discussion is also established. Similar relations are obtained to $CloDis^t$ calculating through anycast distributions.

Some results are as follows. First, unicast/anycast contacts distribution should not be applied always to the distribution of entire network contacts as the broadcast contacts. Second, the reverse of the present solution can also be used to generalize the distribution of broadcast to unicast/anycast modes.

ii. Unicast/anycast distribution belong(s) to Class-2:

Recall from Section 2.2(*ii*), we showed that the resulting function for $CloDis_i^t$ is the same as the primary functions. This inference also applies to the present state $CloDis^t$. Using unicast distributions, we can argue that the distribution of *ENCT* as sums of entire network nodes random variables is normal U~(μ,σ^2) with parameters as follows:

$$\mu = \sum_{i=1}^{m-1} \sum_{j=i}^{m} \mu_{ij} \ , \ \sigma^2 = \sum_{i=1}^{m-1} \sum_{j=i}^{m} \sigma_{ij}^2 \quad (10)$$

A similar proof can be made using anycast contacts distribution. Proofs for *ENICT*, *ENCF*, and *ENICF* are not brought here for brevity.

We end this part with a conclusion. If the distribution of unicast/anycast contacts is of Class-2, it can be applied to broadcast and vice versa.

iii) Unicast/anycast distribution belong(s) to Class-3:

In this case, as we have argued before, the distribution of unicast/anycast contacts cannot be generalized to sums of them as the broadcast contacts. Accordingly, only a best fitting method is applied.

It is worth mentioning that applying the distributions of CTYPE3 to CTYPE2 and/or to CTYPE1 is incorrect in some cases and has no theoretical basis. Therefore, their use in any application such as routing may lead to wrong results. We summarize the proposed solutions of this section in Table II.

3. Proposed routing algorithm

In our method called *CloDis* too, first, the network graph is constructed and the community detection process is done using *CloDises* and a community detection method such as the modularity method[20]. Then, the forwarding decision-making phase acts as a community-based way to forward messages. In this way, a message is routed according to a global ranking of nodes between the communities until it reaches the destination community (inter-community routing). It is then routed to the destination node based on a local ranking of nodes as the intra-community routing.

In the community detection process, for every pair of nodes i and j, the inverse of the average *ICT(i,j)* and the cumulative *CT(i,j)* are included to form communities. This causes nodes, which see each other sooner, to be more likely in a common community, thus possibly reducing delivery latency.

To rank nodes, a criterion is required to evaluate the ability of nodes to forward messages. This criterion consists of multiple different social features.

Numerous studies have demonstrated that some social features do not significantly differ between the nodes [4]. In other words, more distinctive criteria are needed to rank the nodes. Therefore, it is better to use multiple social criteria in the utility function with different and dynamic coefficients. These features are further applied as the core phase of our utility function as done in [4]. The pseudo-code of the *Clodis* algorithm is represented in Algorithm 1.

Table II. Summarization of proposed modeling solutions for different routing modes

Routing	Model	Known Distributions	Class	Solution	Section
Unicast	$CloDis_{ij}^{t}$	-	-	Best fitting of unicast (CTYPE1) contacts	2.1
		anycast / broadcast	1	Generalization to constituent distributions	2.2 (*i*) / 2.3(*i*)
			2	Generalization to the same distribution	2.2(*ii*) / 2.3 (*ii*)
			3	Best fitting of unicast (CTYPE1) contacts	2.1
Anycast	$CloDis_{i}^{t}$	-	-	Best fitting of anycast (CTYPE2) contacts	2.2
		unicast / broadcast	1	Generalization from/to constituent distributions	2.2 (*i*) / 2.3 (*i*)
			2	Generalization to the same distribution	2.2 (*ii*) / 2.3 (*ii*)
			3	Best fitting of anycast (CTYPE2) contacts	2.2
Broadcast	$CloDis^{t}$	-	-	Best fitting of broadcast (CTYPE3) contacts	2.3
		unicast / anycast	1	Generalization from constituent distributions	2.3 (*i*)
			2	Generalization to the same distribution	2.3 (*ii*)
			3	Best fitting of broadcast (CTYPE3) contacts	2.3

Algorithm 1: *Clodis algorithm*

```
Begin
CloDises← extractDistributionsParameters(contacts)
 NetworkGraph ← extractGraph(CloDises)
 IAICT ←inverseOfAverage(ICT)
 CCT ← cumulative(CT)
 //Communities: a list of sets
 Communities ← communityDetection(IAICT,CCT)
 // forwarding stage
CandidateNodes←currentNode.EncounteredNodes()
 AbortingProbThreshold ← 0.5
 for Candidate : CandidateNodes do
     //Required time to send message m
     y ← MessageSize(m)/NetworkInterfaceSpeed
     x1 ← CurrentContactElapsedTime
     // X : contact time random variable
     AbortingProb ← P(X - x1 < y)
     if (AbortingProb < AbortingProbThreshold)
        if (CommunityOf(currentNode) ==
            CommunityOf(destination))
              if ((CommunityOf(Candidate) ==
                 CommunityOf(destination)) and
                 (IntraCommunityRankOf(Candidate) >
IntraCommunityRankOf(currentNode))
                          SendMessageTo(m,Candidate)
        else if ((CommunityOf(Candidate) ==
            CommunityOf(destination)) or
               (InterCommunityRankOf(Candidate) >
InterCommunityRankOf(currentNode))
                          SendMessageTo(m,Candidate)
 end for
end.
```

In the Algorithm-1, we used the same framework presented in our previous work [29] to detect the communities. To calculate *IntraCommunityRankOf*(i) and *InterCommunityRankOf*(i) for every nodes i and destination d, we use the following equation:

$$w_1\text{*similarity}(i,d) + w_2\text{*centrality}(i) + w_3\text{*trust}(i) \; w_4\text{*friendship}(i, d) + w_5\text{*social\_strength}(i) \tag{11}$$

where w_1 to w_5 are dynamic coefficients. If the nodes belong to a common community, the metric *IntraCommunityRankOf* is used as the local ranking and so the social features of nodes of the community are used to calculate that metric. In cases that nodes are in distinct communities, *InterCommunityRankOf* is used as the global ranking and calculated using social features of the whole network nodes.

For every candidate node, we define a prerequisite phase to select it as the possible relay node by influencing the contact distribution parameters of that node as follows.

Using *CloDis*, we calculate the probability of whether the message can be completely sent in the remaining of the current *contact time* or not. If the aborting probability exceeds a threshold, sending the message to the candidate node will not be examined. This technique prevents the incomplete sending of messages due to insufficient *contact time* and thus reduces the aborted messages. Consequently, network costs can be significantly decreased as well.

4. Simulations and results

In this section, our method is compared to some well-known routing methods. The software ONE [21] is used as the simulator. It is based on Java programming language and implements some popular routing protocols such as Epidemic [22]. In addition to artificial movement models, it supports any real traces e.g. Infocom05 [23], MIT Reality Mining [24].

4.1. Comparing protocols

The conventional method for showing the efficiency of an algorithm, model, or framework is to implement it in a simulation environment and compare the results with benchmark protocols in that field. Among various algorithms, the following protocols were employed.

- Epidemic [22] is a knowledgeless and blind method that uses the flood spreading of messages in routing. It causes messages to reach their destination in the shortest possible time with the highest delivery rate when there is enough memory. Instead, it can produce the highest overhead.

- DR [4] is a social-aware algorithm that utilizes communities and a utility function consisting of several social features.
- TCCB [26] uses predicted temporal social contact patterns including temporal closeness and centrality in data forwarding.

4.2. Datasets

The following three real network traces are used for experiments. These networks are used in studies to evaluate the performance of MSN solutions.

- Infocom05 trace was collected at the INFOCOM 2005 conference for three days using small portable electronic devices equipped with Bluetooth. About 41 people from different countries took part in it.
- MIT Reality Mining is a real and long trace with duration of approximately one year which makes it highly suitable for evaluating the mobility of MSN nodes over a long period. In this trace, near 97 students and staff of the University of MIT have participated using Bluetooth-enabled smartphones.
- Infocom06 trace was recorded at the INFOCOM 2006 conference via Bluetooth-enabled devices participating about 98 people in three days.

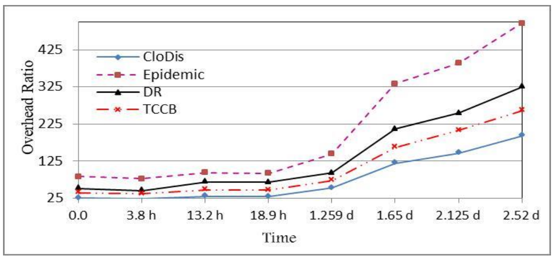
4.3. Performance metrics

The following metrics are used to evaluate the methods.

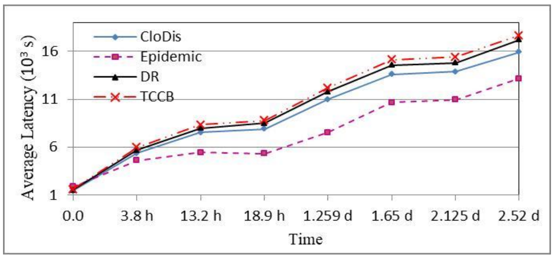
- Average Overhead: The ratio of the number of relayed messages to the number of delivered messages.
- Average delivery Latency: The average time elapsed between the generation of messages and their delivery to the destinations.
- Delivery Rate: The ratio of delivered messages to the total number of generated messages.

4.4. Simulation setup

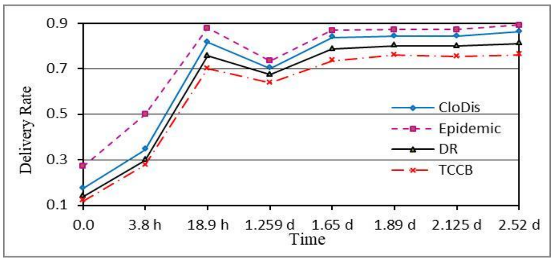
In the experiments, in our method, a day is divided into 5 time periods, including the intervals of 8-12, 12-14, 14-18, 18-24, and 0-8. The *Modularity* method is applied for community detection with resolution parameters as 1. Moreover, aborting probability threshold is set to 0.5, as a tradeoff between delivery rate and latency. For TCCB, the suggested intervals of a day are used including 0-4, 4-8, 8-12, 12-16, 16-20, and 20-24. The parameters g and w are set to 0.8 and 0.5, respectively. We use the Recent Weighted Method (*RWM*) as the prediction method. In DR, the features similarity, trust, and the like are used with the same scenarios and dynamically adjusted coefficients.



(a)



(b)



(c)

Fig. 1. Comparing methods over time constraints for Infocom06 trace. (a) Overhead Ratio, (b) Average Latency, (c) Delivery Rate

Two sets of experiments are conducted in two different scenarios. First, the results are obtained by varying the time. In the second set, the effect of traffic loads is investigated by varying the buffer space. It is noteworthy that the effect of other parameters such as time to live (TTL) of messages can also be examined through the changes made in their values. For investigated traces, the variables in TABLE III are used in the first set of experiments.

TABLE III. Variables setting for the traces in the time constraints test (s: seconds)

Parameters	Infocom06	MIT	Infocom05
Simulation time	342916	2.5 months	274884s
Warm-up	114,305s	4 weeks	91,628s
Update interval	0.2s	10s	0.2s
Device buffer	600MB	500MB	500 MB
Messages TTL	18 hours	12 days	1 day
Messages interval	50-100s	600-1800s	50-100s

For the second set of experiments, the buffer size variable is changed from 20 to 1200 (MB), 10 to 530 (MB), and 10 to 1250 (MB) in Infocom06, MIT, and Infocom05 traces, respectively.

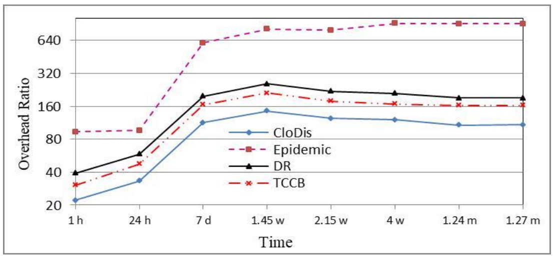
4.5. Simulation results

In the experiments letters h, d, w, and m represent hour, day, week, and month respectively.

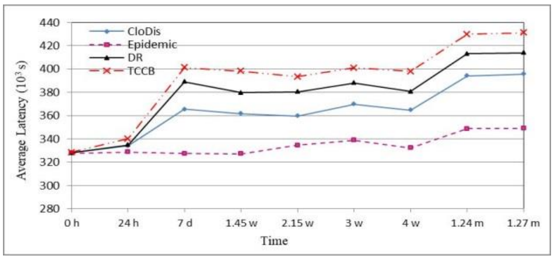
4.5.1. Impact of the time constraints

Fig. 1 illustrates the results of the variable time for the Infocom06 trace. It can be observed that *CloDis* outperforms TCCB and DR in all criteria, and the Epidemic in terms of the overhead ratio. Fig. 1(a) shows that *CloDis* acts better than Epidemic, TCCB, and DR, respectively, about 2.8-times, 27%, and 48% in terms of overhead ratio. It improves average latency and delivery rate, , by an average of 6.5% and 4.6%, as well as 5%

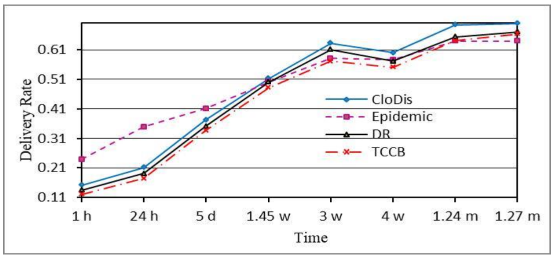
and 4% respectively compared to TCCB and DR as depicted in Figs. 1 (b) and (c). The results of the variable time for MIT trace are presented in Fig. 2. As shown in Fig.2 (a), CloDis works better than all methods in terms of overhead ratio and improves Epidemic, TCCB, and DR by an average of approximately 5-times, 33%, and 41%, respectively.



(a)



(b)

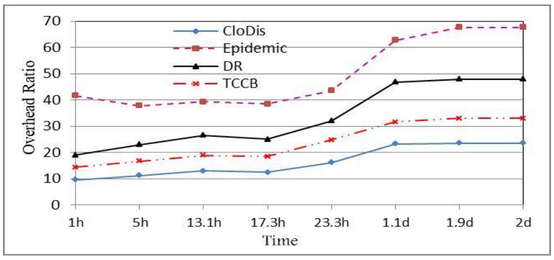


(c)

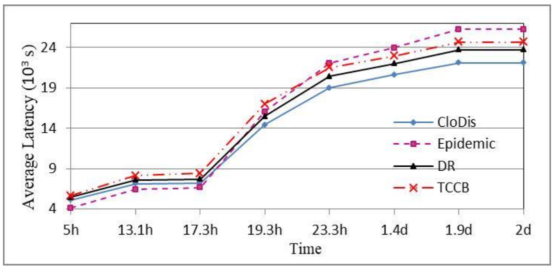
Fig. 2. Comparing methods over time constraints for MIT trace. (a) Overhead Ratio, (b) Average Latency, (c) Delivery Rate

According to Figs. 2 (b) and (c), *CloDis* improves TCCB and DR by about 7.5% and 5% in average latency, as well as 5.3% and 5% in terms of delivery rate, respectively.

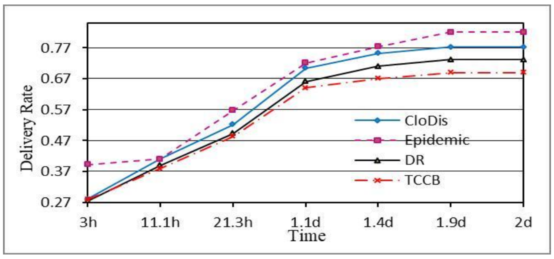
Based on the results, our improvement in terms of overhead is extremely higher than other criteria, especially in contrast to the Epidemic method.



(a)



(b)



(c)

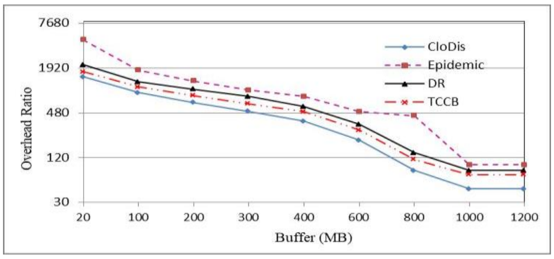
Fig. 3. Comparing methods over time constraints for Infocom05 trace. (a) Overhead Ratio, (b) Average Latency, (c) Delivery Rate

Fig. 3 illustrates results with the time variable for the Infocom05 trace. Fig. 3(a) shows that *CloDis* works better about 3-times, 30%, and 50%, respectively, compared to Epidemic, TCCB, and DR in terms of overhead ratio. Our method improves the average latency and delivery rate, respectively, by an average of 8.5% and 7.6%, as well as 7% and 5.7% compared to TCCB and DR depicted in Figs. 3 (b) and (c).

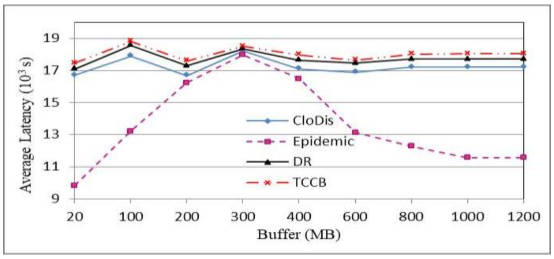
4.5.2. Impact of the buffer space constraints

The results of the comparing methods in different buffer spaces for all traces are illustrated in Figs. 4-6. As shown in Figs. 4 (c), 5 (c), and 6 (c), increasing the buffer size improves the network delivery rate. The reason is that the messages can stay longer time in memory and have more opportunities to be delivered to the destination before their removal. On the other hand, increasing the buffer space reduces the overhead as shown in Figs. 4 (a), 5 (a), and 6 (a). The reason is obvious because the overhead is inversely related to the delivery rate and increasing the delivery rate reduces the number of sent messages. It is known that the number of forwards is correlated with overhead. Thus, the overhead represents a significant reduction.

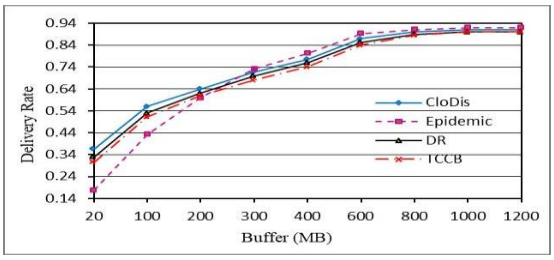
In the case of average latency (Figs. 4 (b), 5 (b), and 6 (b)), by the initial increase of the buffer, the latency increases relatively, but then decreases and remains constant to a certain extent. The reason for the initial increase is that the messages gradually remain in the memory longer, and therefore, increase the delivery latency since the delivery latency is calculated for the delivered messages rather than all created messages. However, it reduces delivery latency by the continuation of this process because the delivery rate demonstrates an increase.



(a)



(b)

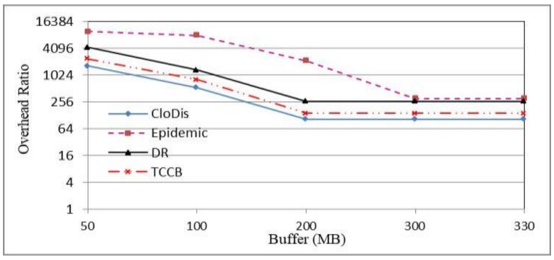


(c)

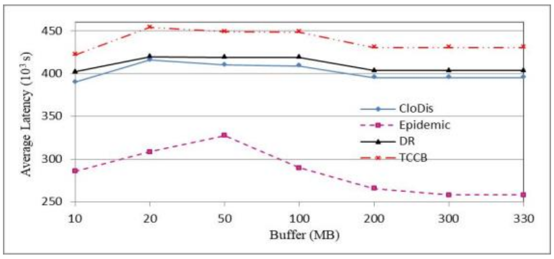
Fig. 4. Comparing methods over buffer constraints for Infocom06 trace. (a) Overhead Ratio, (b) Average Latency, (c) Delivery Rate.

On the other hand, the Epidemic having sufficient memory acts as an upper bound on the delivery rate (Figs. 1 (c) and 3 (c)). However, this method shows less efficiency in low memory. For example, in the above experiments, this method works worse when the memory is reduced from 300 to 200 in Infocom06, 400 to 300 in MIT, and 500 to 400 in Infocom05. Simultaneously, this method has a higher overhead and less delay in most cases.

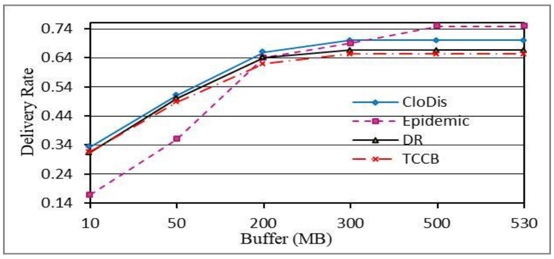
Finally, similar to the first set of experiments, the superiority of our method in improving network criteria compared to other methods is maintained by changing the variable of buffer space.



(a)



(b)

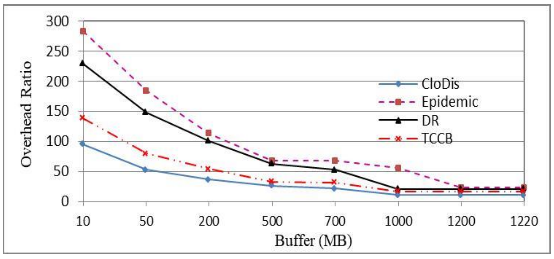


(c)

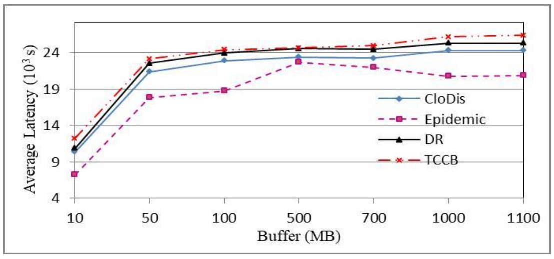
Fig. 5. Comparing methods over buffer constraints for MIT trace. (a) Overhead Ratio, (b) Average Latency, (c) Delivery Rate.

5. Conclusions and future works

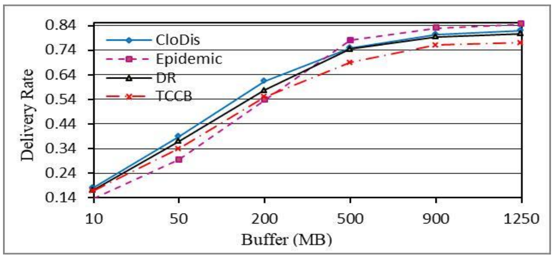
To the best of our knowledge, no study has so far addressed modeling the homogeneous distribution function of node's contact types and exploiting them proportionally in various routing modes. We modeled various contact types with the distribution function of the sums of several random variables. In this regard, it was proven that the generalizations and methods of their use in different modes of routing in the existing literature have fundamental problems. Therefore, a complete model (i.e., *CloDis*) was presented for maintaining the contact information. Using this criterion, it is possible to calculate different probabilities for the occurrence of each contact time and inter-contact time and theirs remaining part-time. Accordingly, we could prevent unsuccessful transmissions and reduce overhead in the proposed routing algorithm. Additionally, the overhead was improved by the CloDis compared to the three strategies Epidemic, TCCB, and DR, as well as the delivery rate and average latency compared to the last two protocols.



(a)



(b)



(c)

Fig. 6. Comparing methods over buffer constraints for Infocom05 trace. (a) Overhead Ratio, (b) Average Latency, (c) Delivery Rate.

The implication of the proposed ideas on reducing energy consumption and increasing network lifetime can be evaluated through a cross-layer approach. There are more challenges about contacts distribution modeling that we will address in future works. We will examine the improvement of community detection algorithms and their modeling, as well as community-independent routing methods using presented solutions.

6. References

[1] C. Boldrini, M. Conti, A. Passarella, "Performance modeling of opportunistic forwarding under heterogeneous mobility", *Computer Communications*, vol. 48, no. 5, pp. 56-70, 2014.

[2] Vinícius F.S. Mota, Felipe D. Cunha, Daniel F. Macedo, José M.S. Nogueira, Antonio A.F. Loureiro, "Protocols, mobility models and tools in opportunistic networks: A survey", *Computer Communications*, vol. 48, no. 1, pp. 5-19, 2014.

[3] Kumaram, Shubham & Srivastava, Shefali & Sharma, Deepak, "Neural network-based routing protocol for opportunistic networks with intelligent water drop optimization: IWDNN", *International Journal of Communication Systems*, vol. 33, issue 8, e4368, 2020.

[4] Ciobanu RI, Dobre C, Reina DG, Toral SL, " A dynamic data routing solution for opportunistic networks", 14th International Conference on Telecommunications (ConTEL), June 2017, Zagreb, Croatia, pp. 83-90.

[5] Aloizio P. Silva, Scott Burleigh, Celso M. Hirata, Katia Obraczka, "A survey on congestion control for delay and disruption tolerant networks", *Ad Hoc Networks*, vol. 25, Part B, no. 15, pp. 480-494, 2015.

[6] فاطمه مشاک، علیرضا کشاورز حداد، شاپور گلبهار حقیقی، « ارتقاء عملکرد شبکه‌های سیار متحمل تأخیر با اولویت‌بندی بسته‌های بافرشده»، مجله مهندسی برق دانشگاه تبریز، جلد ۵۰, شماره ۴, صفحات ۱۸۳۳–۱۸۴۵، ۱۳۹۹.

[7] راضیه نظری، صالح یوسفی، بابک قالب‌ساز جدّی، « ارائه مکانیسمی برای کمینه کردن هزینه های اپراتور سلولی در فرآیند برون‌سپاری ترافیک»، مجله مهندسی برق دانشگاه تبریز, جلد ۴۷، شماره ۴، صفحات ۱۷۴۵–۱۷۵۷، ۱۳۹۶.

[8] Xia F, Liu L, Li J, Ma J, Vasilakos AV, "Socially aware networking: A survey", *IEEE Systems Journal*, vol. 9, issue 3, pp. 904-921, 2015.

[9] S. CC, V. Raychoudhury, G. Marfia, A. Singla, "A survey of routing and data dissemination in Delay Tolerant Networks", *Journal of Network and Computer Applications*, vol. 67, no. 9, pp. 128-146, 2016.

[10] P. Hui and J. Crowcroft, "How Small Labels Create Big Improvements", *Fifth Annual IEEE International Conference on Pervasive Computing and Communications Workshops (PerComW'07)*, March 2007, White Plains NY, USA, pp. 65-70.

[11] A. Roy, S. Bose, A. Tamaghna, S. Dasbit, " Social-based energy-aware multicasting in delay tolerant networks", *Journal of Network and Computer Applications*, vol. 87, no. 12, pp. 169-184, 2017.

[12] A. Passarella, M. Conti, "Analysis of individual pair and aggregate inter-contact times in heterogeneous opportunistic networks", *IEEE Transactions on Mobile Computing*, vol. 12, issue 12, pp. 2483–2495, 2013

[13] W. Gao, Q. Li, B. Zhao, G. Cao, "Multicasting in delay tolerant networks: a social network perspective", *Proceedings of the tenth ACM international symposium on Mobile ad hoc networking and computing ACM (MobiHoc'09)*, May 2009, LA New Orleans, USA, pp. 299–308.

[14] D. Zhang, H. Ma, and D. Zhao," Social-Aware Backbone-Based Multicast Routing in Mobile Opportunistic Networks", *3rd International Conference on Big Data Computing and Communications (BIGCOM)*, Aug. 2017, Chengdu, China, pp. 31-38.

[15] N. Gondaliya, D. Kathiriya, "Community Detection Using Inter Contact Time and Social Characteristics Based Single Copy Routing in Delay Tolerant Networks", *International Journal of Ad hoc, Sensor & Ubiquitous Computing*, vol. 7, no. 1,pp. 21-35, 2016.

[16] M. M.Tulu, M. E. Mkiramweni, R. Hou, S. Feisso, T. Younas," Influential nodes selection to enhance data dissemination in mobile social networks: A survey", *Journal of Network and Computer Applications*, vol. 169, no. 2, Article 102768,2020.

[17] Q. Xu, Z. Xiong, N. Xiao, Y. Xie, D. Yong, M. Deng, H. Min, "Intelligent distributed routing scheme based on social similarity for mobile social networks", *Future Generation Computer Systems*, vol. 69, no. 19, pp. 472-480,2019.

[18] X. Wang, X. Zhong, L. Li, S. Zhang, R. Lu, T. Yang, “TOT: Trust aware opportunistic transmission in cognitive radio Social Internet of Things”, *Computer Communications*, vol. 162, no. 1,pp. 1-11, 2020.

[19] W. Gao, G. Cao, T. La Porta, and J. Han, “On Exploiting Transient Social Contact Patterns for Data Forwarding in Delay-Tolerant Networks”, *IEEE Transactions on Mobile Computing*, vol. 12, no. 1, pp. 151-165, 2013.

[20] L. Waltman, NJ. Van Eck, “A smart local moving algorithm for large-scale modularity-based community detection”, *The European Physical Journal B*, Vol. 86, no. 11, pp. 1-4, 2013.

[21] A. Keränen, J. Ott, T. Kärkkäinen, “The one simulator for dtn protocol evaluation”, *in Proceedings of the Second International Conference on Simulation Tools and Techniques* (*Simutools'9*), vol. 55, Brussels, Belgium, March 2009, pp. 1–10.

[22] A. Vahdat, D. Becker, “Epidemic routing for partially connected ad hoc networks”, *Handbook of Systemic Autoimmune Diseases*, Elsevier, 2000.

[23] Haggle, Haggle project, 2013. <http://www.haggleproject.org>.

[24] N. Eagle, A. Pentland, and D. Lazer, “Inferring Social Network Structure using Mobile Phone Data”, *Proceedings of the National Academy of Sciences*, vol. 106, no. 36, pp. 15274-15278, 2009.

[25] R. Akhtar, Y. Shengua, Z. Zhiyu, Z. Khan, I. Memon, S.Rehman, S. Awan, “Content distribution and protocol design issue for mobile social networks: a survey”, *EURASIP Journal on Wireless Communications and Networking,* 2019, no. 128, 2019.

[26] 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”, *in IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 10372-10383, 2017.

[27] A. Passarella, M. Conti, “Analysis of individual pair and aggregate intercontact times in heterogeneous opportunistic networks”, IEEE Transactions on Mobile Computing, vol. 12, no. 12, pp. 2483–2495, 2013.

[28] M. Xiao, J. Wu and L. Huang, “Home-Based Zero-Knowledge Multi-Copy Routing in Mobile Social Networks”, in *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 5, pp. 1238-1250, 2015.

[29] Moradi. S, Mohasefi. JB, Mahdipour. E, “MSN-CDF: New community detection framework to improve routing in mobile social networks”, *International Journal of Communication Systems*, 2021; vol. 34, no. 18 , e4989, 2021.