Exploiting Sociality for Collaborative Message Dissemination in VANETs

Weiyi Huang[1,2], Peng Li[1,2(✉)], Tao Zhang[3], Yu Jin[1,2], Heng He[1,2], Lei Nie[1,2], and Qin Liu[4]

[1] College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan, Hubei, China
lipeng@wust.edu.cn

[2] Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan, Hubei, China

[3] Department of Computer Science, New York Institute of Technology, New York, USA

[4] School of Cyber Science and Engineering, Wuhan University, Wuhan, Hubei, China

Abstract. Message dissemination problem have attracted great attention in vehicular ad hoc networks (VANETs). One important task is to find a set of relay nodes to maximize the number of successful delivery messages. In this paper, we investigate the message dissemination problem and propose a new method that aims at selecting optimal nodes as the collaborative nodes to distribute message. Firstly, we analyze the real vehicle traces and find its sociality by extracting contacts and using community detecting approach. Secondly, we propose community collaboration degree to measure the collaborative possibility of message delivery in the whole community. Moreover, we use Markov chains to infer future community collaboration degree. Thirdly, we design a community collaboration (CC) algorithm for selecting the optimal collaborative nodes. We compare our algorithm with other methods. The simulation results show that our algorithm performance is better than other methods.

Keywords: Message disseminations · VANETs · Community collaboration degree

1 Introduction

With the development of vehicular ad hoc networks (VANETs), there are more and more message dissemination applications in VANETs. Message dissemination problem is a difficult problem in VANETs, as it is hard to find a set of nodes

This work is partially supported by the NSF of China (No. 61502359, 61602351, 61572370, and 61802286), the Hubei Provincial Natural Science Foundation of China (No. 2018CFB424), and the Wuhan University of Science and Technology Innovative Entrepreneurship Training Program (17ZRA118).

H. Gao et al. (Eds.): CollaborateCom 2018, LNICST 268, pp. 366–385, 2019.
https://doi.org/10.1007/978-3-030-12981-1_26

to maximize message deliver ratio. Moreover, it is also an interesting problem as the messages are timely disseminated based on designed scheme in VANETs.

Due to the delay and opportunity of VANETs, data forwarding between nodes plays an important role in which the information exchange between nodes in the network. The routing schemes have been widely investigated [1]. Some papers consider sociality-based routing, others consider geographic-based routing. Generally speaking, geographical routing considers hot spots, but ignores the internal correlation between nodes. Although messages is easy to disseminated in hot areas, it is hard to diffuse in non-hot areas, resulting in low forwarding efficiency in some areas. Sociality-based forwarding is concerned about the nodes correlation compared with geographical routing. It takes the contacts between nodes as the link, and measures the probability, frequency of contact, etc. These factors determine which nodes are suitable as forwarding nodes. In this paper, we present a new message forwarding scheme that help forward messages to a destination. Our scheme exploits the vehicles moving pattern based on vehicular sociability.

In order to study the rule of nodes in community, we analyze the real taxis history trace and construct the contact graph. Further, we characterize the contact of nodes. Based on our analyses, we formulate two definitions: delivery probability and community collaboration degree. The delivery probability measures the possibility of a node forwarding messages to another node. The community collaboration degree measures a node's contribution to the community. Our scheme has several advantages as compared with existing approach. First, in our scheme, the delivery probability of nodes considers two factors: direct delivery and indirect delivery. Whether a node chooses to deliver directly messages or indirectly messages depends on the direct delivery probability and the indirect delivery probability. Second, in previous papers, only the contact between nodes is considered, but we present the community collaboration degree to measure the relationship between a community and its nodes.

The contributions of this paper are as follow:

- We formulate the message dissemination problem in VANETs, which selects the optimal collaborative nodes based on delivery probability and community collaboration degree, and prove it to be a NP-hard problem.
- Our paper analyzes real vehicle traces and exploits the sociability of vehicles. On the basis of the delivery probability, we also consider the relationship between the total number of forwarding messages in a community and the number of messages which are forwarded by a node.
- We purpose an scheme based on delivery probability and community collaboration degree, for solving the message dissemination problem in VANETs, which predicts delivery probability and community collaboration degree by utilizing the Markov model in the next slot to select a set of nodes to forward messages. Based on the purposed scheme, we design the CC algorithm to measure those nodes with high delivery probability and community collaboration degree, and then choose the optimal nodes as the collaborative nodes.

– We perform experiments to validate the effectiveness of our scheme and other methods. The experimental results show that our scheme has better performance than prophet in terms of delivery ratio, average delay and delivery cost.

The remainder of this paper is organized as follows. Section 2 gives the related work. Section 3 describes the system and defines the problem. Section 4 investigates vehicular sociality. In Sect. 5, we design the scheme based on community collaboration degree. Section 6 describes performance evaluation. Finally, we give our conclusion and outline the directions for future work in Sect. 7.

2 Related Work

Message dissemination is a key component of vehicular ad hoc networks (VANETs). Many schemes have been developed for VANETs, which are different in their protocols characteristics, frame of network [1]. There are many protocols investigated in previous literature [2–4]. The epidemic is used to relay messages by flooding or restricted flooding [5]. FirstContact [3] is a single-copy routing protocol, which consumes only a little buffer and delivery cost. However, forwarding only one copy may cause messages transmission failure, so the performance of this protocol in delivery ratio and delay will be worse than multiple copy protocols. Baiocchi et al. [4] presented an analytical model to evaluate the performance of the distributed beaconless dissemination protocols in VANETs. However, the beaconless protocols may produce occurrence of duplicated message transmissions.

Recently, there are some researches based on social structure [6–13]. In [11], authors presented SimBet, an algorithm for forwarding data in delay-tolerant MANETs based on the node's centrality to choose the next forwarding node. They predicted the delivery by estimating the centrality of the node. They have demonstrated through real trace data that SimBet had better delivery performance compared with epidemic. Also, SimBet achieves the goal that finding a route due to the low connectivity of the sending and receiving nodes. In [12], it introduced a centralized heuristic algorithm to learn delivery probability of different paths and choose the best path. Instead of [12], the work of [13] focused on buffer management and messages schedule. The later the message is created, the higher priority it gets. In contrast, our strategy does not focus on buffer management, but we consider buffer size in the simulation experiment. Besides the delivery probability, we also consider the proportion of node forwarding numbers in the overall forwarding numbers of the community.

Moreover, the game theory based routing protocol is also emerging. There have been many articles published on the theory. Abdelkader et al. [14] proposed a distributed game theoretic approach that computes a node utility function to achieve fair cooperation. Compared with other DTN protocols in performance, this approach shows that fairness among the nodes is improved, and delivery cost is reduced. Cai et al. [15] presented an efficient incentive compatible routing protocol (ICRP) with multiple copies for two-hop DTNs based on the algorithmic game theory. The protocol considered the behaviors of selfish nodes that did not forward other nodes' messages and rely on the nodes to forward its own messages. Besides, the optimal sequential stopping rule and Vickrey Clarke Groves auction strategy are also adopted for selecting optimal relay nodes. Further, they proposed the realistic probability model to find the optimal stopping time threshold and optimizes selection of relay nodes.

In addition, we have noticed some interesting studies [16–21] about message dissemination. He et al. [19] first proposed a store-and-forward framework with extra storage to solve the scalability and high-mobility issues. Then he proposes an optimal link strategy based on the dependence of the delay components. The experimental results show performance of the proposed solution is much higher in delivery ratio compared with the state-of-the-art solutions. Li et al. [20] presented an scheme to ensure the minimal budget to deliver the message to the vehicle in a given geographical area and a given a piece of a message. Further, they considered to utilize the optimal RSUs to forward the message based on the proposed quadtree model. Bi et al. [21] proposed an urban multi-hop broadcast protocol which aimed at delivering emergency messages so that lowering transmission delay and reducing message redundancy.

Although, there are plenty of different approaches in message dissemination, most of them are probably limited to a specific traffic scenarios including the topological structure of street, vehicle speed, etc. Our work aims to propose the message dissemination problem and design a scheme about how to select the optimal collaborative nodes, which is independent of specific scenarios. Compared with [5,22], our scheme considers the proportion of forwarding numbers of a node in the whole community besides delivery probability. Then, we utilize Markov chains to infer the future nodes forwarding states.

3 System Description and Problem Definition

3.1 System Description

In the traditional urban vehicular network, infrastructures are often used as relay nodes for message dissemination because it has larger coverage and longer service time. However, it is difficult to deploy them on a large scale due to its high price and deployment costs. Besides, the dissemination effect is not ideal since it is hard to move flexibly when it is built, which results in a large coverage blind area throughout the network. By using the short-distance communication technology of on-board equipment such as Bluetooth and WiFi, vehicles can become "mobile infrastructures" to accomplish dynamic transmission. Although they only have a small transmission range and short working cycle, they can rely on quantity advantage to compensate for the coverage blindness caused by the infrastructures to some extent.

In urban environment, we can divide dynamic nodes into potential collaborative nodes and high quality collaborative nodes according to whether they are willing to participate in forwarding messages. In general, potential collaborative nodes such as pedestrians, shared bikes, private cars, often request to other nodes for forwarding its own data in buffer but do not always forward data from other nodes for privacy reasons. High quality collaborative nodes are voluntary to forward messages and have a wide range of movement than pedestrians, private cars or shared bikes. These public vehicles cannot refuse to provide forwarding services for privacy reasons. Nevertheless, the number of such high quality collaborative nodes is significantly smaller than the former. Therefore, the research focuses on the potential collaborative nodes.

We assume that there is a potential collaborative node i, if it helps forward the message of node a, then i becomes the collaborative node of node a. However, potential collaborative nodes may be less likely to forward messages for their own reasons and just send their own requests to other nodes. In order to allow more potential collaborative nodes to participate in the forwarding service, we consider to give the nodes that are willing to be collaborative nodes a certain reward to inspire more potential collaborative nodes to participate in the forwarding service. The main advantages of utilizing potential collaborative nodes for message dissemination lies in two aspects. On the one hand, it effectively reduces the deployment and forwarding costs compared with fixed infrastructures. On the other hand, it can increase the likelihood of successful messages delivery based on its high mobility.

3.2 Problem Definition

We consider disseminating message within a given budget. Due to the time to live of messages, it is of great importance to be timely forwarded before they are out-of-date. We define our message dissemination problem as follows:

Definition 1. *In the case of giving the budget B and reward r for inspiring vehicles to disseminate messages in VANETs, how to choose the optimal potential collaborative nodes so that the number of successful delivery messages is maximized?*

The formulation of above problem is shown as follows:

$$maxD(x) = \sum_{i=1}^{n} (x_i * m_i) \tag{1}$$

subject to

$$\sum_{i=1}^{n} (x_i * r_i) \leq B \tag{2}$$

$$x_i \in \{0, 1\}, 1 \leq i \leq n \tag{3}$$

where n represents the total number of nodes in the network. x_i indicates whether the node is the collaborative node. If $x_i = 1$, it indicates that this node is the destination. Otherwise, it is not. m_i represents the number of messages that destination has received. Constraint (2) denotes that the total rewards should be less than the given budget B, where r_i represents the reward for inspiring node i to be a collaborative node.

Our message dissemination problem can be modeled as the budgeted maximum coverage problem [23]. We have the following theorem to prove it,

Theorem 1. *The message dissemination problem is NP-hard*

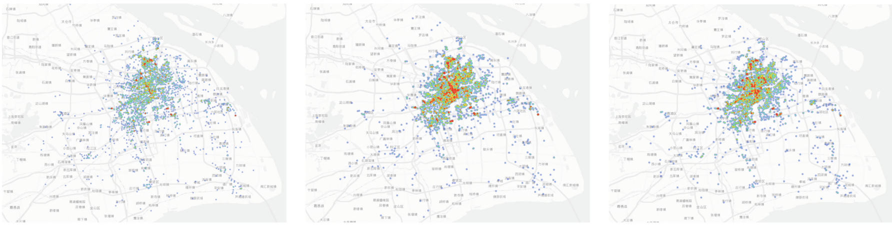
Proof. We can construct the contact graph $G(N, E)$ from the known traces of vehicles, where N is the set of nodes and E is the set of edges. Each node in the network represents a vehicle and each edge indicates that there is at least one communication within the period of time. We assume that $C = \{c_1, c_2, ..., c_n\}$ indicates potential collaborative nodes set with correlative rewards $\{r_i\}_{i=1}^{n}$, $n = |N|$. Similarly, $M = \{m_1, m_2, ..., m_n\}$ indicates a set with correlative the number of received messages. Our goal is to find a set $C^{'} \subseteq C$, such that the number of successful delivery messages are maximized and rewards does not exceed a given budget B. Accordingly, we reduce the message dissemination problem to a budgeted maximum coverage problem, which is a classical NP-hard problem.

4 Scaling Vehicular Sociality

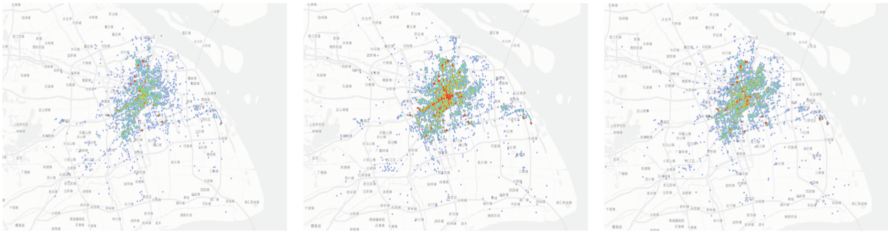
In order to understand the pattern of vehicular mobility and design message dissemination schemes, it is of great importance for us to study historical data in terms of frequency of contacts. After obtaining the messages to be forwarded (which can be implemented by V2V or V2R communication), the vehicle continues to drive along its own path. When two vehicles encounter within the communication range, they can contact with each other but not for long. The sociality between vehicles can be found in a large number of such short and frequent contacts.

4.1 Observation of Hot Spots

We collect the GPS trace of taxis in Shanghai, which is collected between Feb. 1 and Feb. 7, 2007. Due to the interference and loss of wireless signals, we amend the drifted GPS data. Figure 1 is the heat map formed in different time periods after processing the GPS data of 4,316 taxis. As can be seen from the figure, the heat of the whole area is relatively low at 8 a.m., because the majority of office workers travel during this period. At 12 o'clock, the heat increased significantly, and there was a trend that the dispersed heat areas are linked together. Obviously, present vehicle activity is more frequent than 8 o'clock. By 22 o'clock, the heat has dropped compared with 12 o'clock. After analyzing the vehicle trajectory for 7 days, we find that the movement pattern of vehicles presents a concentration and periodicity. On the whole, the areas with large traffic flow are concentrated in Jing 'An district, Xuhui district, Putuo district, Hongkou district, Pudong new area and Minhang district.



(a) Taxi hot spots at 8 a.m. on Feb. 1, 2007 (b) Taxi hot spots at 12 a.m. on Feb. 1, 2007 (c) Taxi hot spots at 22 p.m. on Feb. 1, 2007



(d) Taxi hot spots at 8 a.m. on Feb. 3, 2007 (e) Taxi hot spots at 12 a.m. on Feb. 3, 2007 (f) Taxi hot spots at 22 a.m. on Feb. 3, 2007

Fig. 1. Observation of hot spots in Shanghai

4.2 Constructing Social Structures

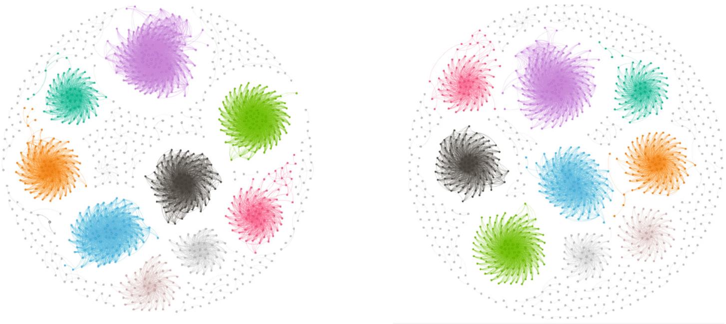
We use the Shanghai taxi data set including GPS data on Feb. 20, 2007. This information includes the following fields: ID, TaxiID, Longitude, Latitude, Speed, Angle, DateTime, status, etc. The granularity of reports is one minute for taxies with passengers and about 15 s for vacant ones.

We first extracted effective V2V communication, as known as contact based on literature [24] and generated the contact graph accordingly. Many literatures have put forward different opinions on the weight of the edge of contact graph. For example, the age of last the contact frequency [25], contact [26], contact ratio [8]. We choose vehicle contact frequency as the weight measurement method of the edge. For the contact graph, each node in the network represents a vehicle, and each edge represents the contact frequency between two vehicles. Then, we get social structure by using fast unfolding algorithm [27] find community and calculate the corresponding modularity [28] defined as

$$Q = \frac{1}{2m}[\sum_{ij} A_{i,j} - \frac{k_i k_j}{2m}]\delta(c_i, c_j) \tag{4}$$

where m is the total number of edges. If there is an edge between node i and j, $A_{i,j}$ is the weight of edge between node i and j. Otherwise, $A_{i,j} = 0$. k_i and k_j are the degrees of node i and j, respectively. c_i and c_j are the community where node i and node j belong, respectively. If c_i is equal to c_j , $\delta(c_i, c_j) = 1$ and zero otherwise.

Figure 2a illustrates a contact graph using Shanghai taxi trace on Feb. 20, 2007, which contains 1127 vehicles. Obviously, there is a clear community structure. The modularity is 0.845. Figure 2b shows the community distribution on Feb. 21, 2007. The modularity is 0.85. Obviously, the relationship of some communities is pretty close. These nodes of close communities should naturally become relay nodes for forwarding messages. Besides, we observe that although different vehicles had different trajectories, vehicles present a stable community relationship.



(a) Community distribution of 1127 taxis on Feb. 20, 2007.

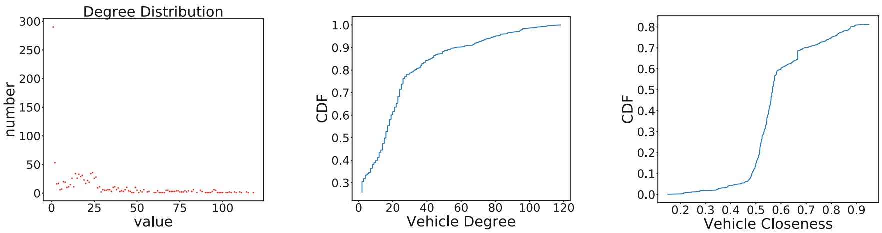
(b) Community distribution of 1152 taxis on Feb. 21, 2007.

Fig. 2. Contact graph extracted from Shanghai taxi trace.

4.3 Analyzing Centrality on Social Structures

Through the collected data and the above community analysis, we obtain the degree distribution of Shanghai taxi in Fig. 3a and calculate the average degree of contact graph. The result is 21.514. Then, we plot the Cumulative Distribution Function (CDF) of degree centrality and closeness centrality in Fig. 3b and c.

It can be seen that node degrees are almost concentrated within 30, which is illustrated from CDF of degree in Fig. 3b and c as well. For any one of these vehicles, it is almost impossible to meet more than one third of the vehicles in a day. Besides, we have some significant findings from Fig. 3b and c. First, both the degree centrality and closeness centrality are effective metrics that distinguish part of vehicles with centrality. Second, in vehicle degree centrality graph, the curve that vehicle degree is more than 30 flattens out and grows slowly. That means the proportion of vehicles whose degree is more than 30 is much less than that within 30. A similar case can be confirmed from vehicle closeness centrality graph.



(a) Degree distribution of 1127 taxis. (b) CDF of degree. (c) CDF of closeness.

Fig. 3. Degree distribution and CDFs of degree, closeness in contact graph constructed from GPS data of Shanghai taxis.

5 Community Collaboration Scheme Design

From the above analysis, we can see that the node movement has sociality. A pair of nodes that have met are likely to meet again in the future. In this section, we study the delivery probability between nodes and the contribution of nodes to the community, and predict future delivery probability and contribution changes by using Markov chain models.

Using potential collaborative nodes as the relay nodes of message dissemination has two advantages. First, the high dynamic of vehicle movement makes the selection of relay node flexible and changeable. Although, the average delay by infrastructure is lower than V2V or V2R (if there are enough infrastructures deployed), the total budget will greatly increase. Moreover, since the infrastructure is located in a fixed location, V2R communication should be ensured for the data transmission of the "last kilometer" if there are places where data cannot be delivered directly. Second, the vehicles, as mobile nodes, keep moving and contact with more vehicles, which will naturally spread data more rapidly. In contrast, infrastructure can only provide data with vehicles passing through its range due to its fixed location.

5.1 Delivery Probability and Community Collaboration Degree

Although, there were many studies based on degrees and closeness, these studies are based on a fact that all the nodes in a network is voluntary to open its own on-board equipment for other nodes for forwarding messages. However, it is unrealistic. Many vehicles do not want to participate in forwarding messages for privacy reasons. For dynamic networks, we denote the number of contacts between node a and node b as $E_a(b)$, where a or b represents anyone of nodes. Therefore, we get the total number of contacts as follows:

$$N = \sum_{a \in \Omega} \sum_{b \in \mu_a} E_a(b) \tag{5}$$

where μ_a denotes the set of contact of node a, Ω denotes the all nodes in the network. In the opportunity network, the node carrying the messages has two ways to deliver the messages to the destination, one is direct delivery and another is indirect delivery. For a and b, we have the definition:

Definition 2. ***Direct Delivery Probability****: The probability that node a delivers messages directly to node b, denoted by $DD_a(b)$.*

It is calculated as:

$$DD_a(b) = \frac{E_a(b)}{N} \tag{6}$$

Definition 3. ***Indirect Delivery Probability****: The probability that node a delivers messages by relay nodes to node b, denoted by $ID_a(b)$.*

It is calculated as:

$$ID_a(b) = \sum_{k \in \mu_a, k \neq b} \left(\frac{E_k(b)}{N} \cdot \frac{E_a(k)}{E_a}\right) \tag{7}$$

$$E_a = \sum_{i,j \in \mu_a, i \neq j} E_i(j) \tag{8}$$

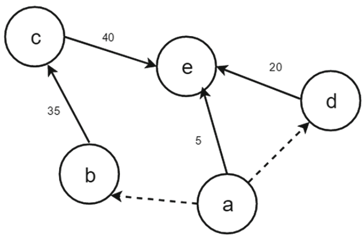


Fig. 4. Messages delivery process.

Here, we give a example. As shown in Fig. 4, we assume that node a encounters node b and node d, node a is going to deliver messages to node e. The total number of encounters between node a and other nodes is $N = E_a(e) + E_b(c) + E_c(e) + E_d(e) = 100$. For node a delivering messages to node e, we have two ways: direct delivery and indirect delivery. If delivered directly, its delivery probability is $DD_a(e) = 0.05$. We observed that if node a delivers the message to b and b indirectly delivers the message to e, its delivery probability $ID_a(e) = max\{ID_b(e) = 0.35 * 0.4, DD_d(e) = 0.2\}$ was higher than $DD_a(e)$. Intuitively, node a is more likely to forward messages to d rather than to b or directly delivering.

Definition 4. ***Delivery Probability****: The ratio of the number of contacts between a and b to N, denoted by $P_a(b)$.*

Given any two nodes, they are affected by both the direct delivery probability and the indirect delivery probability. Here we define the delivery probability of node a to node b:

$$P_a(b) = \alpha DD_a(b) + (1-\alpha)ID_a(b), 0 < \alpha < 1 \tag{9}$$

This probability determines whether the node is significantly affected by the direct delivery probability or the indirect delivery probability. When a node carrying messages encounters multiple nodes and these nodes are qualified for forwarding messages, it is natural to select a node with a high delivery probability as the collaborative nodes. However, the following scenario may occur: We assume that node a encounters node b, c and d, and these three encountered nodes are qualified for forwarding. If the delivery probabilities of the three nodes are $E_a(b) > E_a(c) > E_a(d)$. We should give priority to node b as the collaborative node based on delivery probability. Nevertheless, if node b does not forward or rarely forward messages from other nodes before compared with node c and d, the number of messages forwarded by node b is relatively small compared with the total number of messages forwarded by the whole community. In other words, node b contributes less to community in forwarding messages. Besides, the vehicle will forward messages from multiple nodes on the way, it may be better to select c or d from the perspective of the whole community. In order to study the above situation, we have the following definitions.

Definition 5. ***Node Request Numbers****: The number of messages that node a wants node b to forward, denoted by $Q_a(b)$.*

Definition 6. ***Node Forwarding Numbers****: The number of messages that node a forwards messages from request of node b, denoted by $R_a(b)$.*

Based on Definitions 5 and 6, we have a conclusion that if two nodes forward messages for each other, we think that these two nodes are collaborative. If anyone of them only sends the request and does not forward messages from another node, we think such selfish node should not have priority in forwarding data. Further, we get define the community collaboration degree to measure the importance that a node to the whole community.

Definition 7. ***Community Collaboration Degree****: The total node forwarding numbers of a node a to any node b in a community divided by the sum of the node request numbers and node forwarding numbers, denoted by CS_a.*

$$CS_a = \frac{\sum_{b \in \mu_a} Q_a(b)}{\sum_{b \in \mu_a} (R_a(b) + Q_a(b))} \tag{10}$$

Intuitively, if a node has a high community collaboration degree, this node forward more messages from other nodes than. We should regard it as a collaborative node.

5.2 Inferring Future Delivery Probability and Community Collaboration Degree

As we study how to select the optimal collaborative node in potential collaborative nodes set, we prefer to estimate the delivery probability and community collaboration degree utilizing Markov chain model. In Markov chain model, the current state of the process only depends on a certain number of previous values of the process, which is the order of the process. In order to capture the community collaboration degree dynamics in the network, we divide time into slot of equal length δ. It is great important that measuring the length of δ. If δ is relatively short, we can observe the community collaboration degree dynamics between consecutive time slots but at the same time more random factors would involve in the observations which makes it hard to capture the correlations of community collaboration degree. If δ is relatively long, we find it stable, but lose the dynamics.

For each node i, we examine the number of contacts of each node in a series of contact graphs like Fig. 2 and get a sequence of the number of contacts. After discretizing continuous measures, we get a finite state space of contacts named as Θ_c. Let any state $s \in \Theta_c$ and $m \in \Theta_c^k$, where k is the number of order and $m = \{m_1, m_2, ..., m_k\}$. We denote n_{ms}^c as the number of times that state m equals to state s in a given sequence and n_m^c as the number of times that state m is observed. Therefore, we get the estimation of the state transition probability of contacts when state $m = \{m_1, m_2, ..., m_k\}$ transfers to state $\{m_2, m_3, ..., m_k, s\}$ as follows:

$$p_{ms}^c = \frac{n_{ms}^c}{n_m^c}, n_m^c > 0 \tag{11}$$

Besides, we get a series of requests and forwarding and conduct the state transition probability of requests and forwarding. We denote any state $s' \in \Theta_q$ and $m' \in \Theta_q^k$, where Θ_q is the finite state space of requests. Also, Θ_r is the finite state space of forwarding, any state $s'' \in \Theta_r$ and $m'' \in \Theta_r^k$. $n_{m's'}^q$ is the number of times that state m' equals to state s' in a given sequence of requests and $n_{m'}^q$ is the number of times that state m is observed. Also, $n_{m''s''}^r$ is the number of times that state m'' equals to state s'' in a given sequence of forwarding and $n_{m''}^r$ is the number of times that state m'' is observed in the sequence.

$$p_{m's'}^q = \frac{n_{m's'}^q}{n_{m'}^q}, n_{m'}^q > 0 \tag{12}$$

$$p_{m''s''}^r = \frac{n_{m''s''}^r}{n_{m''}^r}, n_{m''}^r > 0 \tag{13}$$

For any node i in the network, we have the estimated number of contacts in the next slot as follows:

$$E_i'(j) = \sum_{s \in \Theta} p_{ms}^c \cdot s \tag{14}$$

Besides, the estimated number of requests and forwarding calculate as follows:

$$Q_i'(j) = \sum_{s' \in \Theta_q} p^q_{m's'} \cdot s' \tag{15}$$

$$R_i'(j) = \sum_{s'' \in \Theta_r} p^r_{m''s''} \cdot s'' \tag{16}$$

5.3 Choosing the Collaborative Nodes

The collaborative message dissemination problem has been proven to be NP-hard in Sect. 3. Therefore, we give our community collaboration (CC) algorithm based on greedy heuristics. Algorithm 1 shows our Community Collaboration (CC) Algorithm. Given the information, our scheme greedily selects the collaborative nodes with the higher delivery probability and the community collaboration degree. Specifically, when a node carrying messages meets one or more nodes, we compare the delivery probability between it and the destination and the delivery probability between the encountered node and the destination. If the delivery probability of the encountered node is higher than itself, then we further compare their community collaboration degree. Here, we consider two types of collaborative nodes. For potential collaborative nodes, we believe that they will provide forwarding services by giving certain rewards.

Algorithm 1. Community Collaboration (CC) Algorithm

Input: All nodes set Ω, Request list R_i of node i, Contact set C_i of node i
Output: Forwarding set $G = \{G_1, G_2, ..., G_n\}$, $n = |\Omega|$

Begin
$G = \emptyset$
while $\Omega \neq \emptyset$ **do**
    node i transmit NCT_i to nodes it encounters and request NCT
    select node i from Ω and node j from C_i
    for any destination node k in request list R_i **do**
        if $P_i(k) < P_j(k)$ **then**
            if $CS_i < CS_j$ **then**
                Inform node i to add j to candidate set $G_i(k)$
    if node i received NCT_j from j **then**
        $G_i(k) = G_i(k) \bigcup j$
        $\Omega = \Omega \setminus \{j\}$
for destination k in request R_i **do**
    select $\max(P_i(k)/CS_i)$ node in $G_i(k)$
    $G_i(k) = G_i(k) \bigcap max(P_i(k)/CS_i)node$
    $G_i = G_i \cup G_i(k)$
return G_i
End

For high quality collaborative nodes, we believe that these public vehicles cannot refuse to provide forwarding services for privacy reasons. Therefore, we do not pay more attention to high quality collaborative nodes.

To implement our algorithm, we use two tables, node contact table (NCT) and request list table. Node contact table saves updateId field, the list of contact nodes, contact numbers, forwarding numbers, request numbers and request list table saves destinations, data, isForwarded field and keepingTime field. When the node a carrying message meets node b, a sends its contact table and request list table to b. After receiving the contact table of a, b will calculate which messages are appropriate to forward to itself and inform node a. Node a will update the records of corresponding node after receiving it. When the node contact table changes, the updateId field for the node that corresponds to the changed node is updated. New record will override the previous record. When two nodes exchange table information, if the updateId of the same node is different, the record behind the timestamp should be adopted.

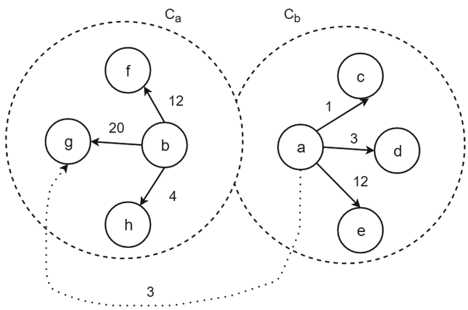


Fig. 5. Cross-community forwarding.

Our algorithm starts with the meeting of two nodes. As shown in Fig. 5, C_a and C_b denotes two communities. Assuming that there are only two communities and node a encounters node b, $\alpha = 0.5$. it starts timing when sending its own NCT to node b and waits for receiving NCT from node b. The total number of contacts $N = E_a(c)+E_a(d)+E_a(e)+E_a(g)+E_b(f)+E_b(g)+E_b(h) = 55$. If node a will deliver messages to node g, the direct delivery probability between node a and node g is $DD_a(g) = E_a(g)/N = 0.055$ and indirect delivery probability $ID_a(g) = E_b(g)/N = 0.364$. $P_a(g) = 0.055 * 0.5 + 0.364 * 0.5 = 0.21$ and $P_b(g) = 0.5 * DD_b(g) + 0.5 * ID_b(g) = 0.5 * 0.36 + 0 = 0.18$. Therefore, node a has a higher delivery probability than node a, and node a should deliver directly it to node g instead of forwarding the data to node b. If $P_a(g) < P_b(g)$, we check how many source nodes of messages in the request list of nodes a and b are themselves, which is the number of requests. And the number of messages forwarded by the node itself, that is, the forwarding amount. If the community collaboration degree of node b is higher than that of node a, then node b will send its own NCT. It will stop the calculation if the update of its own NCT

has not finished within the contact threshold. If node a receives NCT_b from the node b, it will update its own NCT. Then, each destination in the request list is forwarded accordingly.

6 Performance Evaluation

We use the Opportunistic Network Environment (ONE) simulator [29], an effective tool developed by Java. This simulator can simulate the movement of nodes and can evaluate various network parameters. It is designed for Delay Tolerant Networks (DTNs), and also suitable for VANETs. The performance of our scheme is measured and compared with other two schemes: Epidemic [5] and Prophet [22].

6.1 Simulation Setup

In our experiment, we adopt the city of Helsinki as simulation scenario, which is a rectangle of size $5 * 4\,\text{km}^2$. Nodes move according to the Shortest Path Map Based Movement model, map-based movement model that uses Dijkstra's algorithm to find shortest paths between two random map nodes. Though the working day movement model [30] is best for simulating the movement pattern of vehicles, since it simulates the real activities of human being such as working at office, sleeping at home and visiting some places. In fact, compared with the working day movement model, the Shortest Path Map Based Movement model dilute the influence of node movement pattern with time. In turn, we can observe the nodes forwarding in a long time. Considering the simulation time, we have omitted the differences of day and night. The simulation run lasts for 10 h and messages are created after the first 2 h. In order to avoid accidental factors, each simulation is repeated 5 times with random nodes. Other parameters are specified in the Table 1. Here, the participation rate represents the maximum proportion of collaborative nodes in the network. It means that the ratio of the number of collaborative nodes and the total number of nodes should not exceed that value.

We compare our CC algorithm with Epidemic [5], Prophet [22]. The epidemic scheme indicates that random pair-wise exchanges of messages ensure eventual message delivery among mobile nodes in the range of communication. The goal of epidemic is to maximize message delivery ratio and minimize message delay. The prophet generates a delivery predictability sequence from history traces. That can calculate how likely it can relay a message to the destination. The metrics used to compare above schemes are listed as follows:

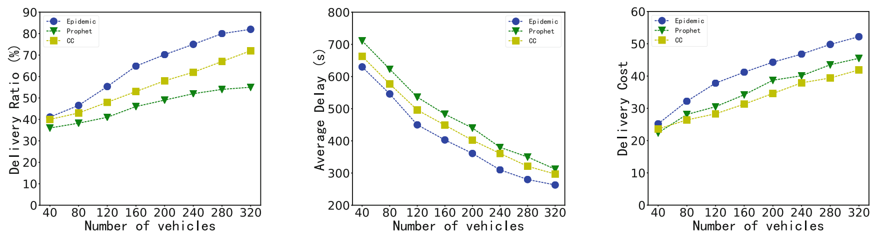
- **Delivery Ratio**: The ratio of messages successfully delivered to destination to all messages generated.
- **Average Delay**: The average duration that each message is successfully delivered.
- **Delivery Cost**: The ratio of the total number of message packets in a network to the number of source packets created.

Table 1. Simulation setting

Network area	$5 * 4$ km^2
Simulation time	10 h
Transmit speed	2 Mbps
Transmission range	200 m
Number of node	240
Package size	50 KB ∼ 1 MB
Messages TTL	1 h
Participation rate	0.8
α	0.5

6.2 Impact of the Number of Vehicles

We first show that how the number of vehicles impacts the performance of the different schemes. As shown in Fig. 6a, as the number of vehicles increases, the delivery ratio of three schemes grows smoothly. The epidemic has the highest delivery ratio in three schemes. This is because the main idea of epidemic is that a node will forward messages from them when it meets other nodes. This scheme can get the higher delivery ratio than other schemes, but the loss packet rate will increase due to network congestion, resource exhaustion and other reasons when the number of copies reaches a certain number. Our scheme is higher than prophet in delivery ratio. Figure 6b show that the average delay lowers as the number of vehicles. The CC algorithm is 6% lower than prophet in delivery delay. Compared with prophet, our scheme not only considers the delivery probability, but also considers the contribution of a node to the community. As for Fig. 6c, we can see that a message which is successfully forwarded requires multiple copies. Our scheme is 7% lower than prophet in delivery cost.

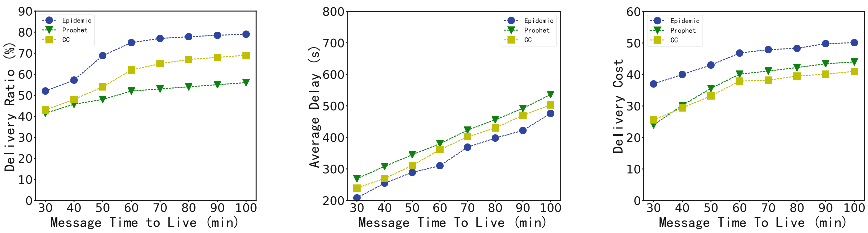


(a) Delivery ratio with different number of vehicles. (b) Average delay with different number of vehicles. (c) Delivery cost with different number of vehicles.

Fig. 6. Impact of the number of vehicles.

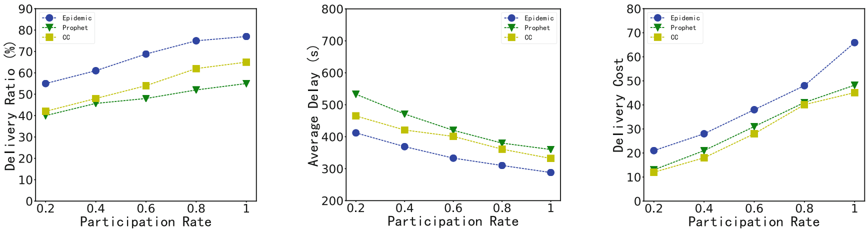
6.3 Impact of Message Time to Live

Figure 7 shows that how the message time to live impacts the performance of the different schemes. As shown in Fig. 7a, the epidemic has the highest delivery ratio in three schemes. Compared with prophet, our algorithm is higher than prophet in delivery ratio. This is because the effective preservation period of messages becomes longer as the increase of TTL, so that there are more opportunities to contact other nodes, which naturally increases the chance of forwarding. In Fig. 7b, our algorithm is 7% lower than prophet in delivery delay. Although our algorithm is less than epidemic, which takes the method of forwarding messages as soon as a node encounters another. As a result, there exist a large number of copies in the network and they waste cache resources. That may cause network congestion. As for delivery cost, our algorithm is 8% lower than prophet and also lower than epidemic in Fig. 7c.



(a) Delivery ratio with different TTLs. (b) Average delay with different TTLs. (c) Delivery cost with different TTLs.

Fig. 7. Impact of the message time to live.



(a) Delivery ratio with different participation ratio. (b) Average delay with different participation ratio. (c) Delivery Cost with different participation ratio.

Fig. 8. Impact of the participation ratio.

6.4 Impact of Participation Ratio

Figure 8 shows the participation rate to change the number of collaborative nodes in the network and compare the changes of three metrics. In Fig. 8a, we can see that the delivery ratio increases as the participation rate increases. When the participation rate reach 1.0, the delivery ratio reaches a maximum. Because there is no privacy node in the network at maximal participation rate. All nodes voluntarily forward messages from other nodes with rewards. Our algorithm has higher delivery ratio compared with prophet, although it is lower than epidemic. Figure 8b show the average delay of three schemes. Our algorithm is 8% higher than prophet. The epidemic has the lowest delay, because its forwarding strategy like flooding that sacrifices space for time. Also, Fig. 8c show that delivery cost of three schemes increase as the participation rate increases. Our algorithm is 9% higher than prophet. If there are many potential collaborative nodes, the delivery ratio of the whole network is relatively low, and there are fewer forwarding nodes, so delivery cost is relatively low. When the participation rate increases, there are fewer potential collaborative nodes and higher forwarding frequency, so the delivery cost increases.

7 Conclusion

In this paper, we study the message forwarding based on community collaboration degree in VANETs. We find that there is a clear social structure within the network by analyzing the real traces of taxis. That inspires us to combine the community into scheme. Firstly, we define the delivery probability including direct delivery probability and indirect delivery probability and community collaborative nodes. Secondly, we purpose our CC algorithm to improve the delivery ratio and lower the delivery delay and delivery cost. The results show effectiveness of our methods under different environments. As a future work, we will study how to choose collaborative nodes in a hybrid vehicular networks.

References

1. Altayeb, M., Mahgoub, I.: A survey of vehicular ad hoc networks routing protocols. Int. J. Innov. Appl. Stud. **3**(3), 829–846 (2013)
2. Ramanathan, R., Hansen, R., Basu, P., Rosales-Hain, R., Krishnan, R.: Prioritized epidemic routing for opportunistic networks. In: International MobiSys Workshop on Mobile Opportunistic Networking, pp. 62–66 (2007)
3. Cao, Y., Sun, Z.: Routing in delay disruption tolerant networks: a taxonomy, survey and challenges. IEEE Commun. Surv. Tutorials **15**(2), 654–677 (2013)
4. Baiocchi, A., Salvo, P., Cuomo, F., Rubin, I.: Understanding spurious message forwarding in vanet beaconless dissemination protocols: an analytical approach. IEEE Trans. Veh.Technol. **65**(4), 2243–2258 (2016)
5. Vahdat, A., Becker, D.: Epidemic routing for partially-connected ad hoc networks. Master thesis (2000)
6. Boldrini, C., Conti, M., Passarella, A.: Social-based autonomic routing in opportunistic networks. Auton. Commun. **15**(1), 31–67 (2009)

7. Conti, M., Kumar, M.: Opportunities in opportunistic computing. Computer **43**(1), 42–50 (2010)
8. Zhu, H., Dong, M., Chang, S., Zhu, Y., Li, M., Shen, X.S.: ZOOM: Scaling the mobility for fast opportunistic forwarding in vehicular networks. In: 2013 Proceedings IEEE INFOCOM, pp. 2832–2840. IEEE (2013)
9. Pujol, J.M., Toledo, A.L., Rodriguez, P.: Fair routing in delay tolerant networks. In: INFOCOM, pp. 837–845 (2009)
10. Fraire, J., Finochietto, J.M.: Routing-aware fair contact plan design for predictable delay tolerant networks. Ad Hoc Netw. **25**, 303–313 (2015)
11. Daly, E.M., Haahr, M.: Social network analysis for routing in disconnected delay-tolerant MANETs. In: ACM Interational Symposium on Mobile Ad Hoc Networking and Computing, pp. 32–40 (2007)
12. Liu, Y., Wu, H., Xia, Y., Wang, Y., Li, F., Yang, P.: Optimal online data dissemination for resource constrained mobile opportunistic networks. IEEE Trans. Veh. Technol. **66**(6), 5301–5315 (2017)
13. Hsu, Y.F., Hu, C.L.: Enhanced buffer management for data delivery to multiple destinations in DTNs. IEEE Trans. Veh. Technol. **65**(10), 8735–8739 (2016)
14. Abdelkader, T., Naik, K., Gad, W.: A game-theoretic approach to supporting fair cooperation in delay tolerant networks. In: Vehicular Technology Conference (2015)
15. Cai, Y., Fan, Y., Wen, D.: An incentive-compatible routing protocol for two-hop delay-tolerant networks. IEEE Trans. Veh. Technol. **65**(1), 266–277 (2016)
16. Liu, B., et al.: Infrastructure-assisted message dissemination for supporting heterogeneous driving patterns. IEEE Trans. Intell. Transp. Syst. **18**(10), 2865–2876 (2017)
17. Lin, Y.Y., Rubin, I.: Integrated message dissemination and traffic regulation for autonomous VANETs. IEEE Trans. Veh. Technol. **66**(10), 8644–8658 (2017)
18. Liu, B., Jia, D., Wang, J., Lu, K., Wu, L.: Cloud-assisted safety message dissemination in VANET-cellular heterogeneous wireless network. IEEE Syst. J. **11**(1), 128–139 (2017)
19. He, J., Cai, L., Cheng, P., Pan, J.: Delay minimization for data dissemination in large-scale vanets with buses and taxis. IEEE Trans. Mobile Comput. **15**(8), 1939–1950 (2016)
20. Li, P., Zhang, T., Huang, C., Chen, X., Fu, B.: RSU-assisted geocast in vehicular ad hoc networks. IEEE Wireless Commun. **24**(1), 53–59 (2017)
21. Bi, Y., Shan, H., Shen, X.S., Wang, N., Zhao, H.: A multi-hop broadcast protocol for emergency message dissemination in urban vehicular ad hoc networks. IEEE Trans. Intell. Transp. Syst. **17**(3), 736–750 (2016)
22. Lindgren, A., Doria, A., Schelén, O.: Probabilistic routing in intermittently connected networks. ACM SIGMOBILE Mobile Comput. Commun. Rev. **7**(3), 19–20 (2004)
23. Khuller, S., Moss, A., Naor, J.: The budgeted maximum coverage problem. Inf. Process. Lett. **70**(1), 39–45 (1999)
24. Zhu, H., Li, M., Fu, L., Xue, G., Zhu, Y., Ni, L.M.: Impact of traffic influxes: revealing exponential intercontact time in urban VANETs. IEEE Trans. Parallel Distrib. Syst. **22**(8), 1258–1266 (2011)
25. Li, Z., Wang, C., Yang, S., Jiang, C., Stojmenovic, I.: Space-crossing: community-based data forwarding in mobile social networks under the hybrid communication architecture. IEEE Trans. Wireless Commun. **14**(9), 4720–4727 (2015)

26. Dubois-Ferriere, H., Grossglauser, M., Vetterli, M.: Age matters: efficient route discovery in mobile ad hoc networks using encounter ages. In: Proceedings of the 4th ACM International Symposium on Mobile Ad Hoc Networking and Computing, pp. 257–266. ACM (2003)
27. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. J. Stat. Mech. **2008**(10), 155–168 (2012)
28. Newman, M.E.J.: Modularity and community structure in networks. Proc. Natl. Acad. Sci. **103**(23), 8577–8582 (2006)
29. Keränen, A., Ott, J., Kärkkäinen, T.: The one simulator for DTN protocol evaluation. In: International Conference on Simulation Tools and Techniques, p. 55 (2009)
30. Ekman, F., Karvo, J.: Working day movement model. In: ACM SIGMOBILE Workshop on Mobility Models, pp. 33–40 (2008)