Computer Networks xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Data routing strategies in opportunistic mobile social networks: Taxonomy and open challenges

Q1 Konglin Zhu [a,*], Wenzhong Li [b], Xiaoming Fu [c], Lin Zhang [a]

[a] School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China
[b] State Key Laboratory for Novel Software and Technology, Nanjing University, Nanjing, China
[c] Institute of Computer Science, University of Göttingen, Göttingen, Germany

A R T I C L E I N F O

Article history:
Received 14 May 2015
Revised 26 September 2015
Accepted 22 October 2015
Available online xxx

Keywords:
Opportunistic mobile social networks
Social-based routing
Taxonomy
Open challenges

A B S T R A C T

Opportunistic Mobile Social Network (MSN) is a kind of Delay Tolerant Network (DTN) in which nodes are mobile with social characteristics. Users in such network carry data, move and forward it to others for information dissemination. To enable efficient data routing in opportunistic MSNs, the social metrics of users, such as mobility pattern, social centrality, community and etc. are leveraged in context of MSNs. In this paper, we investigate the data routing strategies in opportunistic MSNs in the following aspects: (1) the architecture of MSNs and its routing challenges and (2) routing strategies investigation on the basis of different social metrics. We study opportunistic MSN architecture and investigate the social metrics from encounter, social features and social properties, respectively. We show that encounter information is important exemplification of social metrics in opportunistic MSNs. We present other social metrics such as social features and social properties, including social graph properties and community structure. We then elaborate the routing strategies from different perspectives accordingly: encounter-based routing strategies, routing schemes according to social features and routing strategies based on social properties. We discuss the open issues for data routing in opportunistic MSNs, including limitations of routing metrics, collection of social information, social privacy and security, future applications of opportunistic MSNs, and etc.

1. Introduction

As the proliferation of mobile devices, people intend to share information with wireless network among mobile devices. The mobility of users in the network and the limited coverage of access points and cellular network motivate a new kind of mobile wireless network structure, in which mobile nodes work on ad hoc mode and forward data opportunistically upon contacts. The communication of nodes can only be conducted when they are in the communication range of each other. As mobile devices are portable by human beings, which involve the social features into the network, we call such kind of network that combines both opportunistic and social features as opportunistic Mobile Social Network (MSN) [19,20,25,72,84,92].

Although the end-to-end path rarely exists in opportunistic MSNs, the communication among nodes in such network is still desirable. Therefore, effective and efficient data routing strategies are needed to enable the communication in the intermittent connected mobile social networks. In spite of the fact that there are numerous data routing schemes designed for wireless network, they cannot be directly applied to opportunistic MSNs. In the well-connected wireless network, the data routing relies on end-to-end path. Each node maintains routing table according to specific routing policy for the selection of next data relay. According to a specific routing scheme, the entries in the routing table can be

* Corresponding author. Tel.: +86 1062283147.
Q2 *E-mail addresses:* klzhu@bupt.edu.cn (K. Zhu), lwz@dislab.nju.edu.cn (W. Li), fu@cs.uni-goettingen.de (X. Fu), zhanglin@bupt.edu.cn (L. Zhang).

http://dx.doi.org/10.1016/j.comnet.2015.10.018
1389-1286/© 2015 Published by Elsevier B.V.

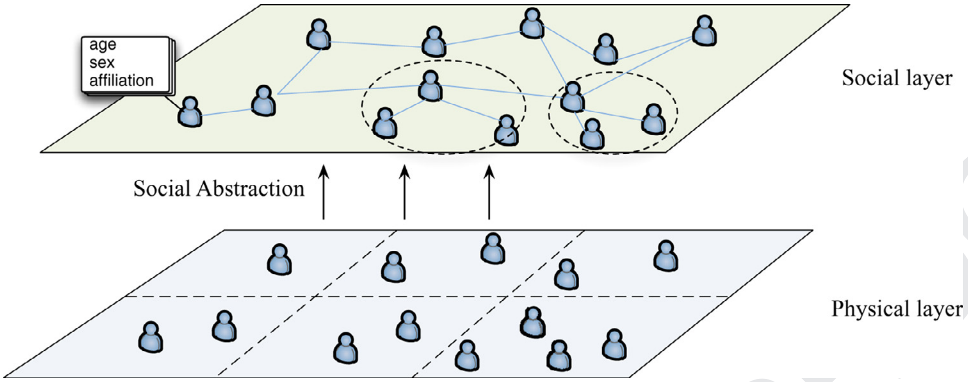


Fig. 1. Two layer presentation of opportunistic MSN.

maintained prior to the arrival of data. Also, since the network is relatively stable, the routing entries are reliable and data routing in such networks can achieve significantly high data delivery ratio. In contrast, the connection in opportunistic MSNs is transient. It is difficult to maintain a complete path during the data forwarding procedure. Therefore, the probability of successful data delivery and time used for data delivery are not guaranteed.

To achieve effective communication without setting up end-to-end communication paths, data transmission in opportunistic MSNs employs the “store-carry-forward” manner, where a node stores and carries data while moving, forwards the data to a relay node on encountering, and propagates the data to further relays until the destination is reached. The main concern of data routing strategies is to decide whether to forward the data to the counterpart when two nodes encounter. Different schemes are devised for data dissemination in opportunistic MSNs.

The most naive strategy such as Epidemic routing [82] was proposed to send data epidemically as long as two nodes encounter until the destination is reached. Based on such epidemic principle, many routing schemes [3,13,48,57,62] using limited copies of messages were developed. Such epidemic based routing strategies suffer from extremely high network cost.

Since nodes in MSNs are mostly controlled by humans, such as mobile phones and vehicles, there are plenty of social relationships and properties which may be used to reveal the network characteristics and facilitate the data routing. The mobility and encounters patterns as well as social features are good contexts to construct social structure in the network and thus can be leveraged for the design of data dissemination strategies. For instance, people encounters frequently are social-connected by encountering events [27]; people with similar social features are socially close [91]; people with similar interest are intend to construct communities [38], and etc. All these social metrics enable to use social analysis methodology to develop data dissemination strategy for efficient data dissemination in opportunistic MSNs.

Although many routing schemes are proposed from different perspectives of social metrics, it lacks surveys that comprehensively summarize the routing schemes from the routing metrics they apply for the design of routing strategies. In this survey, we make taxonomy of the state-of-the-art routing strategies based on routing metrics. We not only focus on the explicit social metrics such as friendship and social features (e.g., gender, age, affiliation, etc.), but also explore the implicit social metrics derived from encounter events. Although they are not the explicit social characteristics represented by the network, they reflect certain level of social interactions. Specifically, people with frequent encounters are likely friends or family members. Many researches consider such information can be used for social analysis [26–28,42,100]. The extraction of social metrics makes the opportunistic MSNs be two layers, as shown in Fig. 1. The lower layer is the physical network status. It mainly infers movements and encounters of nodes in the networks. The upper layer is the abstracted social layer. In this layer, people construct social network, and each node has social features and there are social relationships among different nodes. By applying these social metrics, we can use social analysis methods to build routing utility for data dissemination.

The early stage of the data routing in opportunistic networks mainly focus on using encounter events to construct routing utility for data dissemination. Although they do not explicitly point out the social characteristics in the studies, the metrics such as encounter events are the reflection of social characteristics. Besides, some work directly use social features affiliated with people, such as age, home town, and etc. to study the impact of social features to the data dissemination in opportunistic MSNs. Furthermore, the routing strategies based on social graph are investigated. These schemes use social network properties such as social centrality [27,43], social similarity [27] on the (encounter-based) social graph to seek the appropriate candidate for data forwarding. Also the social structure such as community [43] is studied in opportunistic MSNs and used for data dissemination. In this survey, we will firstly investigate encounter-based routing strategies. Then we will study the routing strategies exploiting social features for data dissemination. Furthermore, we focus on the strategies on the basis of social properties. We will investigate the social graph based MSN routing strategies and community-based routing schemes. We conduct experiments to evaluate the performance of different types of data dissemination strategies.

In the end of the survey, we discuss the several open issues in data routing for opportunistic MSNs. Specifically, We

Please cite this article as: K. Zhu et al., Data routing strategies in opportunistic mobile social networks: Taxonomy and open challenges, Computer Networks (2015), http://dx.doi.org/10.1016/j.comnet.2015.10.018

discuss limitations of routing metrics used for opportunistic MSNs, which mainly focus on several superficial social properties such as social centrality, social similarity, community structure, and etc. whereas more sophisticated social information are not applied. We then focus on the collection of social information. The current collection of social information still mainly relies on the encounter-based graph. The social information from the real life societies is limited for the routing strategies in opportunistic MSNs. We will study privacy and security issue in MSNs, which is ignored by most of data routing strategies but important to people in the network. We also refer the data dissemination to the Information-Centric Networking [2], a new network architecture targeting on the data instead of host, in which the host-to-host data dissemination developed in MSNs cannot be deployed. It needs a new kind of data routing strategy that moves from host-to-host to host-to-data paradigm.

The contributions of this survey are summarized as follows:

- We model three types of information in social perspective, including encounter-based social metrics, social features and social properties like social centrality, social similarity and community.
- We survey data routing strategies in opportunistic MSNs and taxonomy them by routing metrics in social aspects.
- We conduct the experiments to show the performance differences among different types of routing strategies.
- We propose several open challenges in opportunistic MSNs, which cover the composition of routing metrics in opportunistic MSNs, the collection of social metrics, privacy and security issues in MSNs and its application to information-centric networks.

The remainder of the survey is organized as follows. We discuss the architecture and characteristics of opportunistic MSN in Section 2. The motivations and challenges of data routing in opportunistic MSNs are investigated in Section 3. We study encounter-based routing strategies in Section 4. The social feature based routing strategies are surveyed in Section 5. The routing schemes on the basis of social properties are investigated in Section 6. We discuss the challenges and open issues in Section 7. We conclude the survey in Section 8.

2. Architecture and characteristics

In this section, we explain the architecture and characteristics of opportunistic MSNs. We describe the architecture of opportunistic MSNs by its compositions: mobile nodes and their contacts. Then we illustrate characteristics of such network from mobile nodes to their contacts.

2.1. The architecture of opportunistic MSN

Opportunistic mobile social network is a form of sparse dynamic wireless network where mobile nodes work on ad hoc mode. It exploits the human social characteristics, such as daily routines, mobility patterns and interests, to share information and forward data opportunistically upon contacts [83]. As the opportunistic MSN is sparse and nodes in the network are dynamic, the end-to-end path rarely exists. The mobile devices can make contact only when humans encounter each other. The networks are tightly coupled with human social interactions. Therefore, the opportunistic mobile social networks exploit the human behaviors and social relationships to build more efficient and trustworthy message dissemination schemes.

The core of opportunistic MSNs is the mobile nodes and their contacts, which enables the data communication in the network. Nodes in opportunistic MSNs own two dynamic properties: the encountering events between nodes and the mobility of each node. These two properties describe the fundamental channels for communication (encountering) as well as the dynamics of the network (mobility). Additionally, each node is represented by human being or devices controlled by a person. Therefore, each node affiliates social features such as age, professions and etc. Based on the mobility, encounters, and social features, nodes in the network are linked by social relationships, which present the social properties of the network. The major characteristics of mobile nodes in opportunistic MSNs are summarized as follows.

Opportunistic contacts: Mobile users in MSNs intermittently connected and they have opportunity to communicate when two nodes are in communication range of each other. The contact frequency and contact duration are two main factors which affects the probability of data dissemination.

Mobility: People have their own daily routines to commute from home to work place and vice versa or they also visit another places, which form the movement trajectories of nodes in MSNs. Node movements trigger the contacts of people.

Social features: Nodes as individuals which have different social features in the network. For example, node i may be a *female* working as a *faculty* member in a *university*, while node j may be a *male* working as a *manager* in a *company*. Two nodes have different genders, professions and afflictions. These properties compose the social features of nodes.

Social relationship: People in MSN connected by social links construct the social graph. The social relationship represented by social links derives various social properties such as social degree, social centrality and etc. People group together by interest to construct social communities. The community structure makes nodes in one community are highly social related while nodes in different communities are less socially connected.

All these characteristics compose the core social metrics of opportunistic MSNs and are the basis for construction of social graph.

2.2. Social graph and contact-based graph

Opportunistic MSNs can be described as a graph according to different characteristics of nodes and network structure. Integrating social characteristics, an opportunistic MSN can be modeled as a social graph $G = (V, E, W)$ where V is the set of mobile nodes in the network, the set of social links is represented by E and the set of links' weights is depicted by W. The social links indicate the social relations between two nodes and the weight of a link suggests the social strength. Involving encounter events, two nodes have a social link if they have encountered each other. The weight of the edges indicates the frequency of encounters between two nodes. We call such kind of graph that triggered by encounter events as contact-based graph. Social graph can be considered as the network graph with each node associating social features and there are social relationships between nodes. Contact-based graph is the view of the network graph in contact perspective. The nodes are users in the network, and there is an edge if two nodes have contacts. The weight of the edge indicates the frequency of contacts. Since many researches use contacts to measure social relationships between nodes, the contact-based graph is considered as a kind of social graph [5,18,47,59,69].

2.3. Contact-based metrics

The contacts between nodes are opportunistically but compose the most important channel for communication. In this survey, we use contact and encounter interchangeably to present the action that when two nodes are in the communication range.

In the context of opportunistic MSNs, the contact-based metrics are represented as the encounter frequency, contact duration, distribution of inter-contact time and etc. Generally speaking, nodes with higher encounter frequency, longer contact duration are more likely to be friends and their social links are more likely to be stronger. The encounter frequency suggests the frequency that two nodes encounter each other over a period of time. It considers that the more often two people contact with each other, the more chance they have to communicate. Many contact-based data routing strategies for opportunistic MSNs consider encounter frequency as one of the most important metrics to construct the data routing utility function. The contact duration, indicating the time length that two nodes stays in the communication range, is another contact-based metric. It is also important for the data dissemination in opportunistic MSN, especially when the shared data is large which a short contact duration cannot finish the transmission of entire data trunk. Many researchers considers inter-contact time as important metrics for data sharing as it is discovered as a Poisson distribution in many data traces [37,45,51]. Such distribution is used to model different opportunistic MSN scenarios.

2.4. Social features

Different from both geographical and encounter-based metrics, social features are the attributes associating with individuals in the network. For example, in Wu's work [88,91], they considered the social features in two aspects: physical features (i.e. gender, city, profession), and logical features (i.e. a membership in a social group). These features are internal social attributes of nodes. They do not need to use extra equipments or exploit network resource to collect from the networks. We use F_i to denote one social feature. Then each individual is represented by a vector of different features as $(F_1, F_2, \ldots, F_m)$ where each F_i has n_i distinct values for $i = 1, 2, \ldots, m$. For instance, for the social feature $F_x = gender$, where $x \in [1, m]$, it contains two distinct values $\{male, female\}$.

2.5. Social properties

Social properties suggest the social status of a node and its relationships to the social network. Typical social properties utilized for data routing in opportunistic MSNs include: *node degree* [11], known as the degree of a node in social graph, representing the number of friends; *social tie strength* [16], denoted by the weight of an edge between the pair of nodes in social graph; *social centrality*, suggesting the importance of node in the network [63], which is defined by many different ways, such as degree centrality [29], measuring the centrality of a node from the perspective of node degree, closeness centrality [68], evaluating the distance to other nodes in the network, and betweenness centrality [35], measuring the connectivity of the node to the rest of the network; *social similarity* [17], represented by the common friends of two nodes in social graph.

The community is one of the fundamental social structures in social network analysis. People with similar interest or geographical locations compose communities [8,34,65,67]. Individuals in the same community meet each other with high frequency and regularity, while individuals belonging to different communities merely meet each other. The discovery of community structure has been studied for decades. The studies such as modularity methods [9,64,65], label propagation algorithms [77], and etc. A survey of community detection methods for static network is summarized in [34]. In the context of opportunistic MSNs, where nodes are dynamic and distributed, the distributed dynamic community structure discovery is carried out by many studies [10,21,39,44,96]. They apply user interest and proximity to discover the temporal communities in the network in a decentralized manner.

Overall, we discuss the architecture and characteristics of opportunistic MSNs. We consider characteristics of opportunistic MSNs as *social characteristics*, including opportunistic contacts, social features and social relationships. They are fundamentals to compose the social metrics. The social metrics contain contact-based metrics, social features and social properties. In particular, *social features* are the social attributes of individuals in the network, while social properties are the social status of a node and its relationships to the social network. In this paper, we distinguish *social characteristics*, *social features* and *social properties* from each other. They convey different meanings respectively.

3. Motivations and challenges for data routing in opportunistic MSNs

Data routing in opportunistic MSNs is important as it provides the methods for data dissemination. Due to network structure of opportunistic MSNs, it is characterized by large delays, frequent disruptions and lack of stationary paths between nodes. Data dissemination accordingly faces the challenges as follows:

Dynamic network: Nodes in the network are mobile. The movements of nodes are not controlled. Network topology changes from time to time. The continuous changing topology leads to arbitrary disconnections. Thus, the end-to-end path is difficult to be maintained, which results in large delays and unpredictable data dissemination paths. The

Please cite this article as: K. Zhu et al., Data routing strategies in opportunistic mobile social networks: Taxonomy and open challenges, Computer Networks (2015), http://dx.doi.org/10.1016/j.comnet.2015.10.018

successful delivery of data cannot be guaranteed as the nonexistence of end-to-end path.

Limited network information: Due to the fact of dynamic network and unstable connections among nodes, they cannot obtain all network information from opportunistic MSNs. It makes the traditional mobile ad hoc routing protocols (such as AODV [71], DSDV [70] and etc.) cannot adapt to opportunistic MSNs directly. The limited network information leads to the static routes not applicable for dynamic topologies. Besides, the lack of updated and whole information of the network make the calculation of best paths for different destinations become challenging.

Uncertain connection duration and limited resources: Data dissemination in opportunistic network is also related the size of the data. Due to node movements, the connection duration between two nodes is unknown and difficult to be predicted. To enhance the capability of data delivery, node needs to decide how much data will be delivered or which piece of data needs to be delivered when it encounters another peer. In opportunistic MSNs, deciding the number of messages and the size of data for transmission is also affected by the resource of nodes. Nodes in MSNs are portable mobile devices (such as mobile phones), which normally have limited energy supply, storage, CPU and etc. that directly affect the efficiency of data dissemination.

We use an example in a workplace to show these challenges. Consider the opportunistic MSN scenario that people with mobile devices working in the same company. They move from one place to another, which makes the network become dynamic. The connection between two nodes may keep connecting when they stay in the same office while the connection is disrupted when they go to other places, which makes the end-to-end path be difficult to maintain. From the point view of each node, it only has partial information about other peers. Due to the movements of nodes, the changing connection status makes two nodes exist no constant route between them. Any developed routing strategies need to rely on the encountering events. Besides, due to the movement of nodes, the encountering duration is unpredictable. The delivery of data is determined by the size of data and the technology applied for data transmission, as well as the routing policy. Moreover, each mobile device held by people has limited battery, storage and etc. When the energy or the storage is about to run out, people will consider which message should carry for the further data transmission.

By employing different types of social metrics mentioned above, the routing strategies for opportunistic MSNs are divided into encounter-based routing, social feature-based routing and social property-based routing (including social graph-based routing and community-based routing). They either use the historical information to compose the routing metric, which is known as state routing strategies, or apply only the current status of the network for routing utility calculation, known as stateless routing strategies. The taxonomy of the routing schemes is illustrated in Fig. 2. We will explain each of them one by one in the following sections.

4. Encounter-based MSN routing

Encounter-based routing strategies make forwarding decision relying on the encounters of nodes. There are several

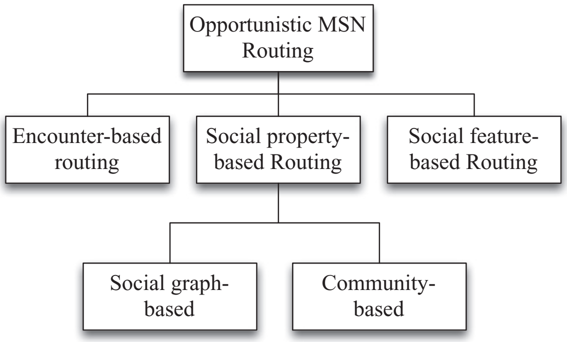


Fig. 2. The taxonomy of opportunistic MSN routing schemes.

strategies directly using encounter information for data routing. For instance, Prophet [56], RAPID [7], MaxProp [14] and etc. were studied in past years. They forward data items according to node contacts, and choose the node with higher contact probability as the relay for data delivery.

The Probabilistic ROuting Protocol using History of Encounters and Transitivity (Prophet) [56] applies the predictability for data delivery as the metric for relay selection. Specifically, the predictability is a probabilistic metric that is calculated by encountering patterns. Each node calculates such predictability for the specified destination. There are three major characteristics of the predictability P. First, the value of P is iteratively determined by the previous value of P, denoted by $P_{(a,b)_{old}}$ for nodes a and b:

$$P_{(a,b)} = P_{(a,b)_{old}} + (1 - P_{(a,b)_{old}}) * P_{init}, \tag{1}$$

where P_{init} is an initialized constant in [0,1]. Second, the value of P decreases if there is no encounter for a certain time interval, which is specified as:

$$P_{(a,b)_{old}} = P_{(a,b)_{old}} * \gamma^{\kappa}, \tag{2}$$

where $\gamma \in [0, 1]$ is a constant and κ is the time interval that have been elapsed from last update. Finally, the transitivity of P is explained as, if a meets b with predictability value $P_{(a,\,b)}$ and b meet c with predictability value $P_{(b,\,c)}$, the predictability value between a and c will be:

$$P_{(a,c)} = P_{(a,c)_{old}} * P_{(a,b)} * P_{(b,c)} * \beta. \tag{3}$$

The scheme works as follows. When two nodes encounter, they exchange predictability values as well as encounter vectors to evaluate the quality of the node. If the predictability value of the counterpart is higher for a destination specified by a piece of data, the data will be transferred to the encountered node.

Jain et al. [47] presented a routing metric named as Minimum Expected Delay (MED) by assuming future contact periods are known. They modify the Dijkstra algorithm [30] to compute the path for DTN with minimum delay. For the improvement, multiple disjoint paths with similar costs are calculated and randomly selects one path from them for data relay. It improves the load balance and reduces the chance of congestions. However, such calculation can only adapt to certain types of opportunistic MSNs that knowing all contact information in advance. To address this limitation, they

Table 1
Social feature based routing strategies.

Routing strategies	Metric	State/stateless	Feature
Social-aware routing [4]	Interest similarity	State	Uses user interest to enhance the utilization of content replication
Homophily-based routing	Homophily	State	Spreads most similar data items among friends and most different data items to strangers
Social greedy [33]	Social distance	State	Makes the data forwarding decisions by comparing the social distance, which is calculated by the similarity of attributes
SANE [58]	Interest similarity	Stateless	Data is only forwarded to the node if the interest similarity between them is larger than a threshold
Social feature-based routing [88,91]	Social attributes	State	Conducts hypercube and calculate feature distance to measure the closeness as routing utility

proposed a new metric, named as Minimum Estimated Expected Delay (MEED). The encounter history is flooded in the whole network. After obtaining the flooded encounter history, it conducts the Dijkstra algorithm to compute the shortest path. Obviously, it introduces too much control overhead.

Spyropoulos et al. proposed a series of multi-copy data delivery schemes, such as Spray and Wait [79] and Spray and Focus [80]. Spray and Wait simply spread the messages to nodes it encounters and each data carrier waits until it meets destination. It is a two-hop data routing scheme with significant waste of data. Besides, no criteria for the data relay selection is carried out in Spray and Wait. To address this issue, the Spray and Focus was proposed to limit the data carriers. The spray phase is similar as Spray and Wait and simply forwards the data to nodes encountered. In the focus phase, a utility function is defined based on the age of the encounter timers to determine whether the node is a good relay for data delivery or not. If a node's utility value is larger than the data carrier, then the data bundles will be forwarded.

The MaxProp [14] was proposed based on prioritizing both the schedule of packets transmitted to other nodes and the schedule of packets that will be deleted from the buffer. Specifically, the packets are transmitted to other nodes when node meetings are addressed by ranking the packets. The packets will be deleted if the buffer is full according to the packet ranking. The ranking mechanism is initialized by a certain value. When two nodes meet, the ranking value will be increased by 1 and it will be exchanged when nodes encounter. Afterwards, a cost for the possible path is calculated, and the path with the lowest cost will be selected for the data delivery.

The Resource Allocation Protocol for Intentional DTN (RAPID) [7] considers the data routing in opportunistic MSN as a resource allocation problem. It proposes an intentional opportunistic MSN routing protocol that can optimize a specific routing metric such as worst-case delivery delay or the fraction of packets that are delivered with a deadline. RAPID translates the routing metric into per-packet utilities which determine how packets should be replicated in the system.

In summary, the encounter-based routing schemes enhance the performance for data dissemination by calculating encounter-based utilities. However, the evaluation of node relationships can only be reflected by the encounter event. Furthermore, it requires exchanging encounter information of nodes in the network, which introduces large amount of control overhead. For showing the characteristics of encounter-based routing strategies, we summarize them in Table 1.

5. Social feature-based MSN routing

Beyond the social properties abstracted from encounter-based routing metrics, the social features, such as user attributes and interest, are also be used for the design of routing strategies for opportunistic MSNs.

An et al. believe people with similar interest have more likelihood to access the same content. Based on this assumption, they proposed a social relation aware routing protocol [4]. It uses the similarity of users' interest as the routing metric and chooses the node with higher similarity of interest as the data relay to increases the utilization of content replication in intermediate nodes.

Zhang et al. proposed a data diffusion strategy based on "homophily" [97]. The "homophily" phenomenon is explained as the trend that two nodes share common characteristics (i.e. interest). It utilizes the friendship and "homophily" to diffuse data pieces. It spreads most similar data items among friends and most different data items to strangers. In this way, data can be diffused in a further wide area, thus achieve shorter data delivery delay.

Social greedy [46] proposed by Jahanbakhsh et al. makes the data forwarding decisions by comparing the social distance between two nodes. The social distance is calculated by the similarity of attributes (such as address, affiliation, school, city, country, etc.) between two nodes. Two nodes with more common attributes, they are closer to each other, and more likely to be chosen as relays for data delivery.

A social-aware and stateless routing (SANE) [58] was proposed by the observation that people with similar interest are more often to meet each other. It uses a k-dimension vector to represent the interests of nodes and calculate the similarity of interest by a cosine function. The cosine similarity calculates the interest similarity between data and node. Data will only be forwarded to the node if the cosine similarity between them is larger than a threshold. Compared with state routing strategies, SANE does not need to store additional information for the calculation of cosine similarity.

Social feature-based algorithm [88,91] takes the multi-dimension social attributes and chooses the node with most similar social features as the destination for data forwarding. Specifically, it conducts hypercube by various social features and uses the feature distance to measure the closeness

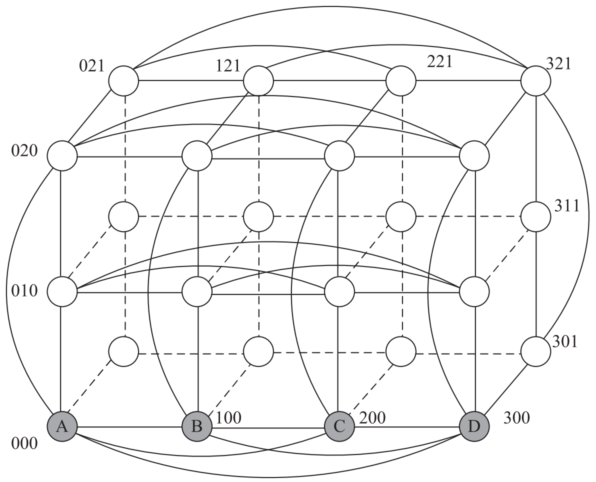


Fig. 3. A 3-dimensional hypercube [91].

between two nodes. As shown in Fig. 3, each digit of the number indicates different social features in one feature space. The node with the closest social features will be selected as the relay for data delivery.

In summary, these social feature based MSN routing schemes still mainly limited in the simple comparison of attributes (e.g., address, and etc.), which lacks the comprehensive social profiles of nodes, thus the improvement of routing performance is restricted by the limited information of social features.

6. MSN routing based on social properties

In this section, we discuss MSN routing strategies on the basis of social properties. Specifically, we will consider the routing strategies based on social graph metrics. Additionally, as community is one of the most important structure in social network and many routing strategies rely on community structure, we separately discuss this group of routing schemes.

6.1. Social graph based MSN routing

SimBet [27] takes the linear combination of social similarity and social centrality as the forwarding utility to construct the data forwarding path. Instead of only considering single social property, the SimBet scheme considers the utility function as the sum of social similarity and social centrality, which measures both the social closeness with destination node and social position of the node in the network. In this work, the social similarity is represented by the number of common friends. The social centrality is calculated by local betweenness. Two separated utility functions are formulated in the following:

$$SimUtil_n(d) = \frac{Sim_n(d)}{Sim_n(d) + Sim_m(d)}, \tag{4}$$

$$BetUtil_n = \frac{Bet_n}{Bet_n + Bet_m}. \tag{5}$$

The overall utility is combined as:

$$SimBetUtil_n(d) = \alpha SimUtil_n(d) + \beta BetUtil_n, \tag{6}$$

where α and β are two parameters defined by authors and $\alpha + \beta = 1$. The scheme chooses the node with higher combination utility value as the relay for data forwarding. The similar idea that uses the concept of social centrality can also be found in [36].

SDM [37] captures the encounter frequency in several data sets and uses Poisson process to formulate the inter-contact time as the basis of the social network modeling. Then it defines a centrality value by the probability that a node is encountered by others in a certain time period as the routing utility, and data is forwarded to nodes with higher centrality. Similarly, PeopleRank [61] is inspired by the PageRank algorithm used in Google's search engine to measure the relative importance of a Web page within a set of pages. Analog to the PageRank, PeopleRank identifies the most popular nodes in a social graph first in a central manner and then deriving to a distributed manner. The message is forwarded to nodes with same or higher rankings, given that popular nodes are more likely to meet others in the network.

Fabbri and Verdone proposed a sociability-based routing strategy in [33]. It assigns to each network node a time-varying scalar parameter which captures its social behavior in terms of frequency and types of encounters. Specifically, it exerts the nodes with high degrees of sociability (i.e., frequently encounter many different nodes) as data relays. The sociability indicator is defined by counting the number of encounters with other nodes in the network. The message will be forwarded to the node with higher sociability.

Moreira et al. proposed an opportunistic routing strategy based on daily routines [60]. It considers the dynamics of social properties in opportunistic MSNs. They evaluate the dynamism of users's behavior by considering user daily routines. In particular, they presented dLife, a routing algorithm able to capture the dynamics of the network represented by time-evolving social times between pair of nodes. dLife represents the dynamics of social structures as a weighted contact graph, where the weights (i.e., social strengths) express how long a pair of nodes is in contact over different period of times. It considers two complementary utility functions: Time-Evolving Contact Duration (TECD) that captures the evolution of social interaction among pairs of users in the same daily period of time, over consecutive days; and TECD Importance (TECDi) that captures the evolution of users importance, based on its node degree and the social strength towards its neighbors, in different periods of time.

Li and Shen proposed a duration utility-based social routing scheme named SEDUM [54]. It exploits both contact frequency and duration in node mobility patterns of social networks to define the duration utility. It increases routing throughput and reduces routing delay by building an effective buffer scheme which maintains the messages by their life time. Those messages with longer lifetime have higher priority to be sent out from buffers. In this scheme, it discovers the minimum number copies of messages to achieve a desired routing delay by using an optimal tree replication algorithm.

The social-tie-based information dissemination in mobile opportunistic social networks [87] applies the strength of

Please cite this article as: K. Zhu et al., Data routing strategies in opportunistic mobile social networks: Taxonomy and open challenges, Computer Networks (2015), http://dx.doi.org/10.1016/j.comnet.2015.10.018

Table 2
Social graph based routing strategies.

Routing strategies	Metric	State/stateless	Feature
SimBet [27]	Centrality and similarity	State	Linearly combines of social similarity and social centrality as the routing utility
SDM [37], PeopleRank [61]	Centrality	State	Measures the relative importance of a node in the network and use ranking for data forwarding
Sociability-based routing [33]	Sociality	State	Counts the number of encounters with other nodes as sociability and forward data to node with higher sociablilty
SEDUM [54]	Encounter frequency and duration	State	Exploits both contact frequency and duration in node mobility patterns of social networks to define the duration utility
Social-tie-based routing [87]	Social tie	State	Weak ties are assigned with more tokens when remote routing and influential nodes are considered more important when local routing

the social ties to disseminate messages. By identifying the strength of social ties in MSNs, it assigns weak-tie nodes with more tokens for future forwarding when two nodes locates in different areas, and after the information is spread to one area, the strong-tie forwarding scheme is applied, in which the influential nodes hold more tokens and considered more important for data forwarding.

SPRINT [23] combines both online social information and contact information to predict the data routing in opportunistic MSNs. It proposes ON algorithm that introduces an additional social routing criterion: online social information about nodes. Furthermore, the ON algorithm also adds the predictable contacts as into the proposed routing algorithm. It not only uses social information about the ON participants learned from the history of contacts, but also from social networks. addition, it includes the possible Poisson-based prediction of a nodes future behavior in the routing decisions

ONSIDE [24] assumes that nodes with common interests tend to meet each other more often than nodes that do not, and connections from online social networks are respected in nodes' encounters in opportunistic MSNs. It leverages information about node's social connections, interests and contact history, in order to decrease network overhead and congestion, while not affecting the network's hit rate and delivery latency.

We summarize the social graph-based routing strategies as shown in Table 2. Data routing schemes based on different kinds of social properties derived from social graph, enhance the performance from social perspective. However, the enhancement is still limited only based on these social properties, without involving structure of social graph, such as community structure.

6.2. Community-based MSN routing

Community as a very important social structure is applied to enhance the performance of data routing in opportunistic MSNs. Community-based strategies make data forwarding decision according to the community structure of the network. By dividing the network into multiple communities, they use different routing strategies to handle the intra-community and inter-community data delivery due to the fact that the connections within a community are rich while the connections between different communities are weak.

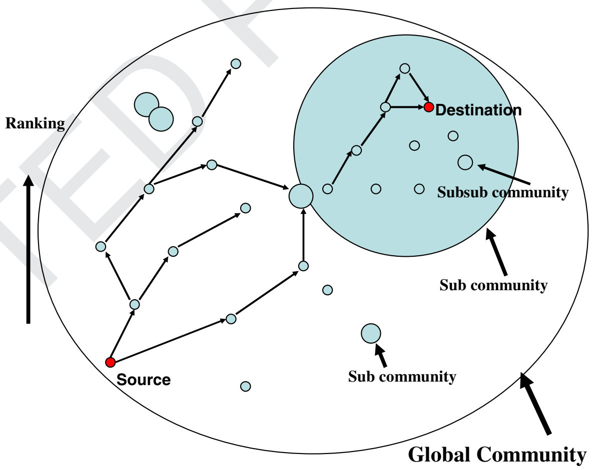


Fig. 4. Illustration of the BUBBLE Rap algorithm [37].

There are several routing strategies exploiting community structure for data routing in DTNs.

One of the earliest works named label routing using community structure for DTN routing was proposed by Hui and Crowcroft [41]. The data routing mechanism is built on Pocket Switched Networks (PSNs) [40], a type of DTN in which the mobile devices are portable by human beings and two devices can communicate when the carriers meet each other. The proposed routing strategy exploits the label affiliated to people to select forwarding relay. The label is assigned according to the community where a person belongs to. The general idea of the label routing works as follows. Each person in the network is assigned with a label based on community structure. When people meet, they exchange the label information. For the selection of the relay, it chooses the node with the same label as the destination node until the destination is reached.

Later, they devised the Bubble Rap algorithm. Bubble Rap [43] considers the data routing in PSN which consists of several communities and there are social relationships among users. It uses distributed version of k-clique percolation [44] as the basic community detection method. Fig. 4 shows the algorithm of Bubble Rap. There are two steps of routing in Bubble Rap. The first step is to forward data to the destination community. It delivers data items from outside of the destination's community according to a node's global social centrality. A node with higher global social centrality will be selected as the relay for data forwarding. Within the destination's community, the forwarding utility is based on a node's local social centrality. The data item will be forwarded to a node with higher local social centrality.

A work related to social-based data multicasting was proposed by Gao et al. [37]. It presents multicasting path selection based on social centrality and social community. In the case of multiple data multicasting, it takes the community structure into consideration. It finds the nodes with destination awareness and forwards the data to the node with highest delivery probability within the community. It continues the forwarding procedure by the constructed social forwarding path to find the destination.

LocalCom [52] uses the degree sum of a node and its neighbors as the metric for community detection. It considers that nodes with high degree sum should belong to the same community. The intra-community routing takes the single hop source routing to forward data. The packet will be directly forwarded along a proposed virtual link. This scheme is based on the high similarity and short hop-count distance within the community. For inter-community data routing, it defines nodes can reach other communities as bridges. The source first forwards the packet to the bridges of the current community by intra-community forwarding mechanism. Each bridge is decided by the pre-pruning process and then further forwards the packet based on the dynamic information. It needs multiple replicas for the inter-community data forwarding.

A work taking the friendship community for information propagation was proposed as Friendship-Based Routing (FBR) [12]. It clusters the nodes which can communicate with short delays as one community. FBR considers the friendship community of varied periods of time. For intra-community communication, it sprays several copies of messages to a number of nodes in the community. For inter-community communication, the data is forwarded only when the destination is in the same periodical community as the relay, which uses the temporal direct connection between communities to tackle the relay selection issue.

Homing spread [93] is a zero-knowledge multi-copy routing algorithms. It assumes mobile nodes in the networks generally visit some locations frequently, which is defined as community homes, while other locations are visited less frequently. In homing spread, people with the same interest are considered to share the same common locations. The messages are spread to community homes at the first place. Then the copies of messages are spread to other homes and mobile nodes. The destination fetches the message when it meets any message holder.

Community-aware opportunistic routing [94] uses similar community home concept for single-copy routing algorithm design. It chooses the community home by calculating the centrality of nodes. The node with the highest centrality is considered as the community home. The messages then are forwarded to those homes. By maintaining an optimal set of relays, each home determines the best relay and meanwhile computes the minimum excepted delivery delay. Afterwards, the home nodes send the messages to the optimal selected relays until the destination home is reached.

Abdelkader et al. proposed a routing protocol named as SGBR using social grouping for opportunistic MSNs [1]. It assumes that there is a global observer which can collect the information from the entire network. SGBR uses social relations to build groups and spreads message copies to those nodes with higher metric values to the message carrier. By this manner, it reduces the need of collecting network wide information, maximizes the delivery ratio and meanwhile minimizes the overhead.

SMART [99] uses a heuristic method for the community detection and then applies the convolution of social centrality and social similarity as the utility to reduce the chance of dead end and blind spot. For intra-community communication, it defines social centrality and social similarity locally within the community. For inter-community communication, it extends the concept of centrality and similarity to a community level. By constructing efficient and effective inter-communication routing strategy, it improves the routing performance of inter-community communication. For both intra and inter-community communication, nodes with higher utility values are selected for data forwarding.

In summary, community-based routing strategies try to improve data forwarding efficiency by community structure. However, most existing community partitioning methods are complicated and static when applied to DTNs. Furthermore, data transmission between communities is difficult task due to rare efficient routing schemes are proposed for inter-community communication. A summary of community-based routing strategies is presented in Table 3.

7. Performance comparison

To show the performance of different kinds of routing strategies in opportunistic MSNs, we conduct a comprehensive performance comparison by including several typical routing strategies such as PROPHET, SimBet, BubbleRap, FBR and SMART, which represent encounter-based and social-based routing schemes, respectively.

7.1. Data traces

Our experiment is conducted on three public available opportunistic MSN traces: MIT Reality [31], DieselNet [15] and Cabspotting [73]. The MIT Reality data set consists of the location traces of 97 users with Nokia 6600 smart phones at MIT during the 2004–2005 academic year. DieselNet logs mobility traces of 34 buses in Amherst. Each bus is equipped with a computer and a GPS. It records the GPS locations of all the buses during the 20 days from October to November in 2007. Cabspotting is a mobility trace of taxi cabs in San Francisco. Each taxi is outfitted with a GPS tracking device. It contains GPS coordinates of 536 taxis collected over 30 days in San Francisco Bay Area. The statistics of the three data sets are summarized in Table 4. The three traces cover a large diversity of mobility patterns and environment, from human movements on campus (MIT Reality) to vehicles mobility in cities (DieselNet and Cabspotting), with experimental periods from a few days to several months.

Table 3
Community-based routing strategies.

Routing strategies	Community detection	Metric	Feature
LABEL [41]	Academic affiliation	N/A	Chooses node with same label (in the same group) for data forwarding
Bubble Rap [43]	K-Clique [44]	Centrality	Forwards messages by global and local centrality for outside and inside of destination community
MDM [37]	K-Clique [44]	Centrality	Finds the nodes with destination awareness and forwards the data to the node with highest delivery probability
LocalCom [52]	Neighbor graph and NCuts [85]	N/A	Use single hop source routing and bridges for intra and inter-community routing, respectively
FBR [12]	Two-hop friends	Inter-contact time	Sprays several copies of messages in the community, and the data is forwarded only when the destination is in the same periodical community
Homing spread [93]	Colocation/interest	Visiting frequency	The messages are spread to community homes at the first place. Destination fetches the message when it meets any message holder
Community-aware routing [94]	Colocation/interest	Centrality	The messages are forwarded to community homes. Afterwards, the home nodes send the messages to the optimal selected relays until the destination home is reached
SGBR [1]	Encounter frequency	Inter-contact time	Uses encounter frequency to build groups and spreads message copies to those nodes with higher metric values to the message carrier
SMART [99]	m-partition [99]	Centrality and similarity	Applies the convolution of social centrality and social similarity as the utility value to reduce the chance of dead end and blind spot

Table 4
Statistics of the DTN data sets.

Traces	*MIT reality*	*DieselNet*	*Cabspotting*
Network type	Bluetooth	802.11b	None
No. devices	97	34	536
No. contacts	54,667	2,284	111,153
Duration (days)	246	20	30

7.2. Experiment setup

We launch the experiment on the HaggleSim simulator [41]. It takes the discrete sequential encounter events and the corresponding social graph as the inputs and makes data forwarding decision using various routing algorithms. For each experiment, we emulate 1000 messages sent from a random selected source to destination. We run every experiment 20 times for statistical convergence.

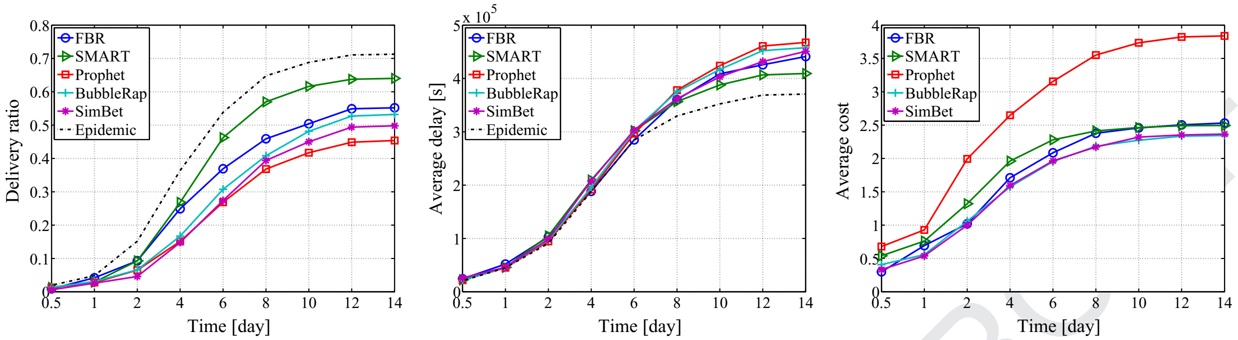
The following performance metrics are used to evaluate the performance of DTN routing algorithms.

- *Delivery ratio:* the ratio of the number of destinations having received the data to the total number of destinations.
- *Average delay:* the average time delay for each data item delivered from the source to the destination.
- *Average cost:* the average number of relays used for data delivery from the source to the destination.

We extract a 2-week session from MIT Reality, DieselNet and Cabspotting respectively and run the simulator over the selected sessions with uniformly generated traffic.

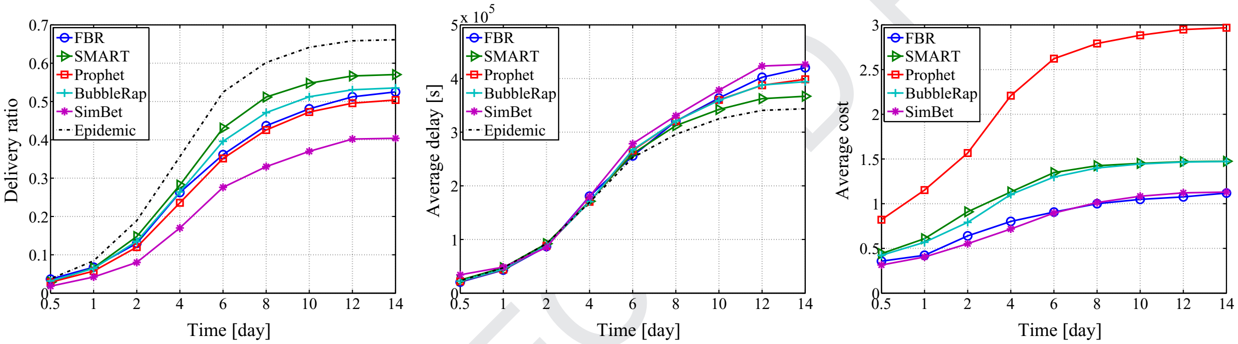
The performance comparison in three data sets is presented in Figs. 5–7. Fig. 5 shows the performance of various algorithms as a function of time on MIT Reality trace. The delivery ratio is compared in Fig. 5a. The results of Epidemic provide the upper bound for the delivery ratio, and it reaches 70% in the end of the experiment period. On the contradictory, the delivery ratio of PROPHET only has 40%. The reason PROPHET performs the worst is due to the strong community structure of MIT Reality trace. When source and destination are inter-connected by a long path, the performance of PROPHET degrades. SimBet exploits social properties to enhance the delivery ratio which fits of such kind of human network thus the delivery ratio reaches to 45%. Bubble Rap, FBR and SMART take advantages of both social properties and community structure, so they perform even better, which achieve approximate 48%, 50% and 60% of delivery ratio respectively. Average delay is compared in Fig. 5b. In this case, most of the time their performance are very close. The Epidemic achieves the lower bound of average delay. Those strategies rely on social properties and community structure (e.g., Bubble Rap, FBR and SMART) can locate the destination in the community rapidly, thus they have lower average delay than both SimBet and PROPHET. Average cost is compared in Fig. 5c. Epidemic shows the largest cost which is much larger than others so that cannot represent in the figure. Besides, the cost of PROPHET is the highest. This indicates that “transitivity” in PROPHET is not accurate enough to predict the relay selection. SMART also has a high cost due to its decay function which may choose a longer path for packet delivery.

Fig. 6 presents the performance results of various algorithms as a function of time on DieselNet data set. The delivery ratio is depicted in Fig. 6a. Epidemic again provides the upper bound of the delivery ratio. In contrast, SimBet in this case performs the worst, which achieves less than 40% of



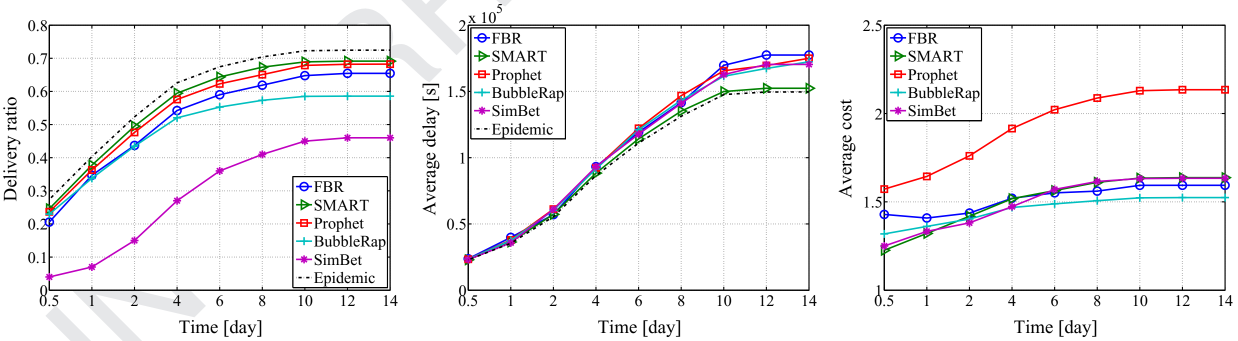
(a) Delivery ratio (b) Average delay (c) Average cost

Fig. 5. The performance comparison of various strategies on MIT Reality Mining trace [99].



(a) Delivery ratio (b) Average delay (c) Average cost

Fig. 6. The performance comparison of various strategies on DieselNet trace [99].



(a) Delivery ratio (b) Average delay (c) Average cost

Fig. 7. The performance comparison of various strategies on Cabspotting trace [99].

delivery ratio. PROPHET has better delivery ratio which reaches 47% in the end of the experiment period. The rest routing strategies with social properties and community structure have higher delivery ratios. Compared with MIT Reality which is a human networking, DieselNet is a bus network. The contact routine of DieselNet is predictable, thus the encounter-based PROPHET performs better than SimBet. Furthermore, the community structure helps to enhance the delivery ratio. This is the reason that FBR, Bubble Rap and SMART have better performance than others. Regarding the average delay and the average cost of each strategy as shown in Fig. 6b and Fig. 6c, The average cost of SMART is about 50% of that of PROPHET and higher than FBR and SimBet. Due to the regular and repetition routine of buses in DieselNet, it makes the SimBet take more time to wait until destinations. Therefore, it has lower delivery ratio and higher average cost. Since DieselNet has more tight clustering structure, it makes Bubble Rap, FBR and SMART performs better than others. SMART has similar cost with social-related strategies but much lower cost than PROPHET.

Comparison of different algorithms' performance on Cabspotting trace is shown in Fig. 7. Compared with MIT Reality and DieselNet, Cabspotting is a taxi mobile social network. Taxies are driven for different destinations which are less clustered. Fig. 7a depicts the delivery ratio of varied algorithms as a function of time. The SMART has very similar performance as PROPHET. It has 5% higher delivery ratio than FBR. Bubble Rap algorithm is impacted by weak community structure, which lowers down its delivery ratio around 10% compared to SMART. SimBet has the lowest delivery ratio, which is much lower than other strategies. In terms of average delay as shown in Fig. 7b, SMART costs as low as Epidemic algorithm delay, which is much lower than others. The delay for FBR is the highest, following PROPHET, BubbleRap and SimBet. The average costs of various algorithms are similar as shown in Fig. 7c. PROPHET is the highest among the evaluated strategies. The rest of them perform similar in terms of average cost.

Overall, the evaluation results indicate that encounter-based routing strategies (e.g., PROPHET) are less sensitive to the social characteristics in the network compared with other types of routing strategies. It has lower performance in the network where social characteristics are explicit (e.g., MIT Reality). In contrast, it performs better than others in the bus (e.g., DieselNet) and vehicle networks (e.g., Cabspotting). Social property based routing strategies (e.g., SimBet) is highly associated with the social properties in the network. In networks with rich social properties (e.g., MIT Reality), social property based routing strategies have much better performance, whereas their performance is unconspicuous in the networks with less social properties (e.g., Cabspotting). Community based routing strategies usually combines social properties and community structure. It has even better performance than social property based routing strategies in networks with social properties. Although some community based strategies have humble performance in the vehicular networks, they are still competitive with others.

8. Open issues

Opportunistic MSN as a new kind of MSN, its research is still in an early stage. Many studies targeting on the routing of opportunistic MSNs have been summarized in the above sections. However, the study in this area still faces different types of challenges and open issues, which refers the characteristics of dynamics, delay and social aspects of opportunistic MSN. In this section, we try to point out these open issues and discussion them in detail.

8.1. Limitation of routing metrics

Different types of routing metrics are used for data forwarding in opportunistic MSNs. They have various limitations. For instance, the geographical based information has significant privacy issue and needs additional devices or modules for data collection, while encounter based information normally needs much storage spaces and computation resource. Although the social features and social properties derived from both geographical and encounter based graph can be used to enhance the routing performance, the exploration of social features and properties are still limited. So far, the social features used for data routing mainly focus on the interest of users and their basic affiliated attributes (e.g., age, affiliation, and etc.). Social properties are also limited in similarity, centrality, social ties and community structure. It leaves a large space to investigate profound social routing metrics. It faces two major challenges for deeper exploration of social features and properties in opportunistic MSNs. First of all, the more sophisticated social properties and structures such as clustering coefficient, triadic closure [32], small world [89] may also be attempted for data routing in MSNs. More importantly, due to different MSN scenarios, it would be interesting to study the representative social features and properties for different scenarios. For example, in a university scenario, the community structure may be more important than the scenarios of shop districts or subways. Second, the deeper exploration of different social features and properties relying the calculation of those features and properties. As opportunistic MSNs are distributed and intermittently connected, whereas majority of social features and properties are computed centrally, thus how to calculate the value of sophisticated social metrics in opportunistic MSN becomes challenging. Although there are some distributed algorithms are developed for the calculation of community [44]. The development of new routing protocols is still limited by the calculation of social features and properties in opportunistic MSNs are difficult to be measured.

8.2. Collection of social features and social properties

The social properties derived from geographical and encounter based graph have been well used for data routing in MSNs. The performance of these routing schemes is similar with the encounter-based routing schemes [98]. One question is what the performance of routing if real social properties are used for data dissemination. Besides, only limited number of social features has been utilized for opportunistic MSNs. One very important reason is because the collection and updates of detailed social properties and social features in opportunistic MSNs is challenging. Although there are data traces with social information, such as Facebook dataset [90], or Twitter dataset [50], and meanwhile DTN traces such as Reality Mining [31], Infocom series [78], data traces combining both social and mobile aspects are limited. Collection of such data traces needs both the encounter patterns of nodes, and meanwhile the social information of them. Some researches, such as SPRINT and ONSIDE, also try to exert real social information for data routing. However, due to the lack of data traces, they cannot conduct the experiments on the real data traces. Thus it is hard to evaluate the effectiveness of the proposed routing strategy.

8.3. Privacy issue of routing metrics

Geographical, encounter information, and social features used for data routing in opportunistic MSN are all sensitive to users in the network. Exposing geographical information (e.g., trajectory) among the network may bring personal security concern. Meanwhile, different types of social information will also refer privacy issue. Malicious node in the network may apply the social information to conduct harmful behaviors to innocent node. The study of privacy concern in social networks [6,55,75,76] has been prevailing recently. Besides, people with close friendship may have less privacy issue since people are willing to share part of his privacy information among friends. However, the privacy preserving bound between strangers is much higher. A person may not share any personal information with a stranger. Therefore, one future focus is that how to preserve user privacy during the routing process in opportunistic MSNs from social perspective.

8.4. Routing security in MSNs

The security issue is another alternative concern for routing in opportunistic MSNs. Most of studies consider nodes in the network are innocent, which are not harmful to others. However, this is not true in the real scenarios. Therefore, the secure communication between nodes and avoiding the malicious nodes in the routing of opportunistic MSN are also challenges. The most commonly used mechanisms for secure communication is to encrypt data for dissemination. However, different from the centralized communication where a server can be setup to enable the secure key distribution and user authentication, in opportunistic MSN environment, where the centralized communication is impossible, makes the secure communication becomes an open issue. The malicious behavior such as Denial of Service [95], and Sybil attack [74] and etc. in opportunistic MSNs needs more attention. Although some work has been initiated for the study of Sybil attack in opportunistic MSNs [22], more deeply investigation are still desirable.

8.5. Social-aware resource allocation

In an opportunistic MSN, with human involved, the portable mobile devices have limited energy, storage and computing capability and etc. All these resources may run out during the routing process. One question is how to allocate different resources to make the resource utilized efficiently. Social relationship in such case takes a very important role. People may be willing to share more resource with his friends, but unwilling to serve for strangers. For instance, if the battery of a node is running out, a node may first give up the data delivery for those nodes that it is not familiar with. The similar scenarios can also be found for storage, computation and other resources. The other one is how to preserve the resource for the consideration of further usage of device and environment concern. People with friendship may be cooperative for data storage and computation sharing. For example, a node with very limited storage may apply his friend's storage for those data not urgently used. Similar scenario can also be exploited for energy consumption and computing capability. Although this concern may be similar with the study of social selfish routing [53], most of the work only considers storage of mobile nodes as limitation. The more comprehensive study is still needed, especially when considering the various types of resource allocation of mobile devices.

8.6. Application to information-centric network

As the proliferation of content centric network in opportunistic mobile social networks [49,66,81,86], data dissemination needs to adapt to such network structure. In content-centric mobile social networks, nodes do not need to care where the date is stored. Data is cached on the path for data transmission. Therefore, the source-destination data routing schemes cannot adapt to such kind of networks. The main challenges of routing in content-centric opportunistic MSNs become to address the following two questions: where is the data and how to cache data? Thus, the routing for content-centric MSNs needs to be in end-to-data diagram. As the opportunistic MSN is opportunistically connected, optimal caching mechanism for content-centric MSNs also needs to be carried out.

9. Conclusion

We make a taxonomy of routing strategies in opportunistic mobile social networks according to social metrics in this survey. We discuss the architecture of opportunistic MSNs and study four types of social metrics, including geo-based metrics, encounter-based metrics, social feature and social properties. Accordingly, the routing strategies are divided into geo-based, encounter-based, social feature-based and social properties-based and we discuss each class of routing schemes in social perspective. We also investigate the open issues of the routing in opportunistic MSNs, including routing metrics, collection of social features, privacy and security, resource allocation and future application of opportunistic MSN to information-centric network.

References

[1] T. Abdelkader, K. Naik, A. Nayak, N. Goel, V. Srivastava, SGBR: a routing protocol for delay tolerant networks using social grouping, IEEE Trans. Parallel Distrib. Syst. 24 (12) (2013) 2472–2481.
[2] B. Ahlgren, C. Dannewitz, C. Imbrenda, D. Kutscher, B. Ohlman, A survey of information-centric networking, IEEE Commun. Mag. 50 (7) (2012) 26–36.
[3] E. Altman, G. Neglia, F. De Pellegrini, D. Miorandi, Decentralized stochastic control of delay tolerant networks, in: Proceedings of INFOCOM '09, 2009, pp. 1134–1142.
[4] J. An, Y. Ko, D. Lee, A social relation aware routing protocol for mobile ad hoc networks, in: Proceedings of IEEE International Conference on Pervasive Computing and Communications, PerCom '09, 2009, pp. 1–6.
[5] L. Arantes, A. Goldman, M.V. dos Santos, Using evolving graphs to evaluate DTN routing protocols, in: Proceedings of ExtremeCom 2009, 2009.
[6] R. Baden, A. Bender, N. Spring, B. Bhattacharjee, D. Starin, Persona: an online social network with user-defined privacy, Proceedings of the ACM SIGCOMM 2009 Conference on Data Communication, SIGCOMM '09, ACM, New York, NY, USA, 2009, pp. 135–146.
[7] A. Balasubramanian, B. Levine, A. Venkataramani, DTN routing as a resource allocation problem, Proceedings of the 2007 conference on Applications, technologies, architectures, and protocols for computer communications (SIGCOMM '07), ACM, New York, NY, USA, 2007, pp. 373–384.
[8] B. Ball, B. Karrer, M. Newman, Efficient and principled method for detecting communities in networks, Phys. Rev. E 84 (3) (2011) 036103.
[9] V.D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, J. Stat. Mech.: Theory Exp. 2008 (10) (2008) P10008.
[10] E.G. Boix, A.L. Carreton, C. Scholliers, T.V. Cutsem, W. De Meuter, T. D'Hondt, Flocks: enabling dynamic group interactions in mobile social networking applications, Proceedings of the 2011 ACM Symposium on Applied Computing, ACM, 2011, pp. 425–432.
[11] J.A. Bondy, U.S.R. Murty, Graph Theory with Applications, vol. 6, Macmillan, London, 1976.
[12] E. Bulut, B. Szymanski, Exploiting friendship relations for efficient routing in mobile social networks, IEEE Trans. Parallel Distrib. Syst. 1 (2012) 99.
[13] E. Bulut, Z. Wang, B. Szymanski, Cost-effective multiperiod spraying for routing in delay-tolerant networks, IEEE/ACM Trans. Netw. 18 (5) (2010) 1530–1543.
[14] J. Burgess, B. Gallagher, D. Jensen, B. Levine, Maxprop: routing for vehicle-based disruption-tolerant networks, in: Proceedings of 25th IEEE International Conference on Computer Communications (INFOCOM '06), 2006, pp. 1–11.

Please cite this article as: K. Zhu et al., Data routing strategies in opportunistic mobile social networks: Taxonomy and open challenges, Computer Networks (2015), http://dx.doi.org/10.1016/j.comnet.2015.10.018

[15] J. Burgess, B.N. Levine, R. Mahajan, J. Zahorjan, A. Balasubramanian, A. Venkataramani, Y. Zhou, B. Croft, N. Banerjee, M. Corner, D. Towsley, CRAWDAD Data Set Umass/diesel (v. 2008-09-14). http://crawdad.cs.dartmouth.edu/umass/diesel, 2008.
[16] R.S. Burt, Structural Holes: The Social Structure of Competition, Harvard University Press, 2009.
[17] D.T. Campbell, Common fate, similarity, and other indices of the status of aggregates of persons as social entities, Behav. Sci. 3 (1) (1958) 14–25.
[18] V. Cerf, S. Burleigh, A. Hooke, L. Torgerson, R. Durst, K. Scott, K. Fall, H. Weiss, Delay-tolerant Networking Architecture. RFC4838, April, 2007.
[19] A. Chaintreau, P. Fraigniaud, E. Lebhar, Opportunistic spatial gossip over mobile social networks, Proceedings of the First Workshop on Online Social Networks, ACM, 2008, pp. 73–78.
[20] A. Chaintreau, A. Mtibaa, L. Massoulie, C. Diot, The diameter of opportunistic mobile networks, Proceedings of the 2007 ACM CoNEXT Conference, ACM, 2007, p. 12.
[21] S.-Y. Chan, P. Hui, K. Xu, Community detection of time-varying mobile social networks, Complex Sciences, Springer, 2009, pp. 1154–1159.
[22] W. Chang, J. Wu, C.C. Tan, F. Li, Sybil defenses in mobile social networks, in: Proceedings of Global Telecommunications, GLOBECOM, 2013, pp. 641–646.
[23] R. Ciobanu, C. Dobre, V. Cristea, Sprint: social prediction-based opportunistic routing, in: Proceedings of World of Wireless, Mobile and Multimedia Networks, WoWMoM, 2013 IEEE 14th International Symposium and Workshops on a, 2013, pp. 1–7.
[24] R.-I. Ciobanu, R.-C. Marin, C. Dobre, V. Cristea, C. Mavromoustakis, Onside: socially-aware and interest-based dissemination in opportunistic networks, in: Proceedings of Network Operations and Management Symposium, NOMS, IEEE, 2014, pp. 1–6.
[25] M. Conti, S. Giordano, M. May, A. Passarella, From opportunistic networks to opportunistic computing, IEEE Commun. Mag. 48 (9) (2010) 126–139.
[26] E. Cozzo, R.A. Banos, S. Meloni, Y. Moreno, Contact-based social contagion in multiplex networks, Phys. Rev. E 88 (5) (2013) 050801.
[27] E.M. Daly, M. Haahr, Social network analysis for routing in disconnected delay-tolerant manets, Proceedings of the 8th ACM International Symposium on Mobile Ad hoc Networking and Computing (MobiHoc '07), ACM, New York, NY, USA, 2007, pp. 32–40.
[28] E.M. Daly, M. Haahr, Social network analysis for information flow in disconnected delay-tolerant manets, IEEE Trans. Mob. Comput. 8 (5) (2009) 606–621.
[29] R. Diestel, Graph Theory. 2005. Grad. Texts in Math, 2005
[30] E.W. Dijkstra, A note on two problems in connexion with graphs, Numer. Math. 1 (1) (1959) 269–271.
[31] N. Eagle, A.S. Pentland, CRAWDAD Data Set Mit/reality (v. 2005-07-01). http://crawdad.cs.dartmouth.edu/mit/reality, 2005.
[32] D. Easley, J. Kleinberg, Networks, Crowds, and Markets: Reasoning About a Highly Connected World, Cambridge University Press, 2010.
[33] F. Fabbri, R. Verdone, A sociability-based routing scheme for delay-tolerant networks, EURASIP J. Wirel. Commun. Netw. 2011 (1) (2011).
[34] S. Fortunato, Community detection in graphs, Phys. Rep. 486 (3C5) (2010) 75–174.
[35] L.C. Freeman, A set of measures of centrality based on betweenness, Sociometry (1977) 35–41.
[36] W. Gao, G. Cao, On exploiting transient contact patterns for data forwarding in delay tolerant networks, in: Proceedings of the 18th IEEE International Conference on Network Protocols (ICNP '10), 2010, pp. 193–202.
[37] 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 (MobiHoc '09), ACM, New York, NY, USA, 2009, pp. 299–308.
[38] M. Girvan, M.E. Newman, Community structure in social and biological networks, Proc. Natl. Acad. Sci. 99 (12) (2002) 7821–7826.
[39] R. Grob, M. Kuhn, R. Wattenhofer, M. Wirz, Cluestr: mobile social networking for enhanced group communication, Proceedings of the ACM 2009 International Conference on Supporting Group Work, ACM, 2009, pp. 81–90.
[40] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, C. Diot, Pocket switched networks and human mobility in conference environments, Proceedings of the 2005 ACM SIGCOMM Workshop on Delay-tolerant Networking, WDTN '05, ACM, New York, NY, USA, 2005, pp. 244–251.
[41] P. Hui, J. Crowcroft, How small labels create big improvements, in: Proceedings of the Fifth Annual IEEE International Conference on Pervasive Computing and Communications Workshops, 2007 (PerCom Workshops '07), 2007, pp. 65–70.
[42] P. Hui, J. Crowcroft, Human mobility models and opportunistic communications system design, Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci. 366 (1872) (2008) 2005–2016.
[43] P. Hui, J. Crowcroft, E. Yoneki, Bubble rap: social-based forwarding in delay tolerant networks, Proceedings of the 9th ACM International Symposium on Mobile Ad hoc Networking and Computing (MobiHoc '08), ACM, New York, NY, USA, 2008, pp. 241–250.
[44] P. Hui, E. Yoneki, S.y. Chan, J. Crowcroft, Distributed community detection in delay tolerant networks, Proceedings of Mobility in the Evolving Internet Architecture, Mobiarch '07, 2007,
[45] S. Ioannidis, A. Chaintreau, L. Massoulie, Optimal and scalable distribution of content updates over a mobile social network, in: Proceedings of INFOCOM 2009, IEEE, 2009, pp. 1422–1430.
[46] K. Jahanbakhsh, G.C. Shoja, V. King, Social-greedy: a socially-based greedy routing algorithm for delay tolerant networks, Proceedings of the Second International Workshop on Mobile Opportunistic Networking (MobiOpp '10), ACM, New York, NY, USA, 2010, pp. 159–162.
[47] S. Jain, K. Fall, R. Patra, Routing in a delay tolerant network, Proceedings of the 2004 Conference on Applications, Technologies, aAchitectures, and Protocols for Computer Communications (SIGCOMM '04), ACM, New York, NY, USA, 2004, pp. 145–158.
[48] P. Juang, H. Oki, Y. Wang, M. Martonosi, L.S. Peh, D. Rubenstein, Energy-efficient computing for wildlife tracking: Design tradeoffs and early experiences with zebranet, SIGARCH Comput. Archit. News 30 (5) (2002) 96–107.
[49] V. Kawadia, N. Riga, J. Opper, D. Sampath, Slinky: an adaptive protocol for content access in disruption-tolerant ad hoc networks, in: Proceedings of ACM MobiHoc 2011 International Workshop on Tactical Mobile Ad Hoc Networking, 2011.
[50] H. Kwak, C. Lee, H. Park, S. Moon, What is twitter, a social network or a news media?, Proceedings of the 19th International Conference on World Wide Web, ACM, 2010, pp. 591–600.
[51] U. Lee, S.Y. Oh, K.-W. Lee, M. Gerla, Relaycast: scalable multicast routing in delay tolerant networks, Proceedings of International Conference on Network Protocols, ICNP 2008, IEEE, 2008, pp. 218–227.
[52] F. Li, J. Wu, Localcom: a community-based epidemic forwarding scheme in disruption-tolerant networks, Proceedings of the 6th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON '09), IEEE Press, Piscataway, NJ, USA, 2009, pp. 574–582.
[53] Q. Li, S. Zhu, G. Cao, Routing in socially selfish delay tolerant networks, Proceedings of IEEE INFOCOM, 2010, IEEE, 2010, pp. 1–9.
[54] Z. Li, H. Shen, Sedum: Exploiting social networks in utility–based distributed routing for DTNs, IEEE Trans. Comput. 62 (1) (2013) 83–97.
[55] X. Liang, X. Li, K. Zhang, R. Lu, X. Lin, X. Shen, Fully anonymous profile matching in mobile social networks, IEEE J. Sel. Areas Commun. (2013).(in press)
[56] A. Lindgren, A. Doria, O. Schelén, Probabilistic Routing in Intermittently Connected Networks, volume 3126 of Lecture Notes in Computer Science, Springer-Verlag GmbH, 2004, pp. 239–254.
[57] T. Matsuda, T. Takine, (p,q)-epidemic routing for sparsely populated mobile ad hoc networks, IEEE J. Sel. Areas Commun. 26 (5) (2008) 783–793.
[58] A. Mei, G. Morabito, P. Santi, J. Stefa, Social-aware stateless forwarding in pocket switched networks, in: Proceedings of INFOCOM '11, 2011, pp. 251–255.
[59] A.G. Miklas, K.K. Gollu, K.K. Chan, S. Saroiu, K.P. Gummadi, E. De Lara, Exploiting social interactions in mobile systems, UbiComp 2007: Ubiquitous Computing, Springer, 2007, pp. 409–428.
[60] W. Moreira, P. Mendes, S. Sargento, Opportunistic routing based on daily routines, in: Proceedings of IEEE International Symposium on World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2012, 2012, pp. 1–6.
[61] A. Mtibaa, M. May, C. Diot, M. Ammar, Peoplerank: social opportunistic forwarding, Proceedings of the 29th Conference on Information Communications (INFOCOM '10), IEEE Press, Piscataway, NJ, USA, 2010, pp. 111–115.
[62] P. Mundur, M. Seligman, J.N. Lee, Immunity-based epidemic routing in intermittent networks, in: Proceedings of the 5th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, SECON '08, 2008, pp. 609–611.
[63] M. Newman, Networks: An Introduction, Oxford University Press, 2010.
[64] M. Newman, M. Girvan, Finding and evaluating community structure in networks, Phys. Rev. E 69 (2004) 026113.
[65] M.E. Newman, Fast algorithm for detecting community structure in networks, Phys. Rev. E 69 (6) (2004) 066133.
[66] A.D. Nguyen, P. Senac, V. Ramiro, M. Diaz, Pervasive intelligent routing in content centric delay tolerant networks, in: Proceedings of the IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing (DASC '11), 2011a, pp. 178–185.

[67] N.P. Nguyen, T.N. Dinh, Y. Xuan, M.T. Thai, Adaptive algorithms for detecting community structure in dynamic social networks, in: Proceedings of 30th Annual IEEE International Conference on Computer Communications (INFOCOM '11), 2011, pp. 2282–2290.
[68] J.D. Noh, H. Rieger, Random walks on complex networks, Phys. Rev. Lett. 92 (11) (2004) 118701.
[69] F.D. Pellegrini, D. Miorandi, I. Carreras, I. Chlamtac, A graph-based model for disconnected ad hoc networks, Proceedings of INFOCOM, IEEE, 2007, pp. 373–381.
[70] C.E. Perkins, P. Bhagwat, Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers, Proceedings of ACM SIGCOMM Computer Communication Review, 24, ACM, 1994, pp. 234–244.
[71] C.E. Perkins, E.M. Royer, Ad-hoc on-demand distance vector routing, Proceedings of Second IEEE Workshop on Mobile Computing Systems and Applications (WMCSA'99), IEEE, 1999, pp. 90–100.
[72] A.-K. Pietiläinen, E. Oliver, J. LeBrun, G. Varghese, C. Diot, Mobiclique: middleware for mobile social networking, Proceedings of the 2nd ACM Workshop on Online Social Networks, ACM, 2009, pp. 49–54.
[73] M. Piorkowski, N. Sarafijanovic-Djukic, M. Grossglauser, CRAWDAD Data Set epfl/mobility (v. 2009-02-24). http://crawdad.cs.dartmouth.edu/epfl/mobility, 2009.
[74] C. Piro, C. Shields, B.N. Levine, Detecting the sybil attack in mobile ad hoc networks, in: Proceedings of SecureComm, 6, 2006, pp. 1–11.
[75] K.P. Puttaswamy, A. Sala, B.Y. Zhao, Starclique: guaranteeing user privacy in social networks against intersection attacks, Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies, CoNEXT '09, ACM, New York, NY, USA, 2009, pp. 157–168.
[76] K.P. Puttaswamy, B.Y. Zhao, Preserving privacy in location-based mobile social applications, Proceedings of the Eleventh Workshop on Mobile Computing Systems & Applications, ACM, 2010, pp. 1–6.
[77] U.N. Raghavan, R. Albert, S. Kumara, Near linear time algorithm to detect community structures in large-scale networks, Phys. Rev. E 76 (3) (2007) 036106.
[78] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, A. Chaintreau, CRAWDAD Data Set Cambridge/haggle (v. 2006-01-31). Downloaded from http://crawdad.org/cambridge/haggle/, 2006.
[79] T. Spyropoulos, K. Psounis, C.S. Raghavendra, Spray and wait: an efficient routing scheme for intermittently Connected mobile networks, in: Proceedings of Workshop on Delay-Tolerant Networking, WDTN '05, ACM, New York, NY, USA, 2005, pp. 252–259.
[80] T. Spyropoulos, K. Psounis, C.S. Raghavendra, Spray and focus: efficient mobility-assisted routing for heterogeneous and correlated mobility, Proceedings of the Fifth Annual IEEE International Conference on Pervasive Computing and Communications Workshops, (PerCom Workshops' 07), IEEE, 2007, pp. 79–85.
[81] G. Tyson, J. Bigham, E. Bodanese, Towards an information-centric delay-tolerant network, in: Proceedings of IEEE INFOCOM Workshop on Emerging Design Choices in Name-Oriented Networking (NOMEN), 2013.
[82] A. Vahdat, D. Becker, et al., Epidemic Routing for Partially Connected Ad hoc Networks, Technical report, cs-200006, Duke University, 2000.
[83] N. Vastardis, K. Yang, Mobile social networks: architectures, social properties, and key research challenges, IEEE Commun. Surv. Tutor. 15 (3) (2013) 1355–1371.
[84] A.M. Vegni, T.D. Little, Vanets as an opportunistic mobile social network, in: Proceedings of Opportunistic Mobile Social Networks, 2014, p. 437.
[85] U.V. Luxburg, A tutorial on spectral clustering, Stat. Computing 17 (4) (2007) 395–416.
[86] L. Wang, R. Wakikawa, R. Kuntz, R. Vuyyuru, L. Zhang, Data naming in vehicle-to-vehicle communications, Proceedings of IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), IEEE, 2012, pp. 328–333.
[87] Y. Wang, J. Wu, Social-tie-based information dissemination in mobile opportunistic social networks, Proceedings of IEEE 14th International Symposium and Workshops on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), IEEE, 2013, pp. 1–6.
[88] Y. Wang, W.-S. Yang, J. Wu, Analysis of a hypercube-based social feature multipath routing in delay tolerant networks, IEEE Trans. Parallel Distrib. Syst. 24 (9) (2013) 1706–1716.
[89] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (6684) (1998) 440–442.
[90] C. Wilson, B. Boe, A. Sala, K.P. Puttaswamy, B.Y. Zhao, User interactions in social networks and their implications, in: Proceedings of European Conference on Computer Systems, EuroSys '09, ACM, New York, NY, USA, 2009, pp. 205–218.
[91] J. Wu, Y. Wang, Social feature-based multi-path routing in delay tolerant networks, in: Proceedings of the 31st Annual IEEE International Conference on Computer Communications (INFOCOM '12), 2012.
[92] J. Wu, Y. Wang, Opportunistic Mobile Social Networks, CRC Press, 2014.
[93] J. Wu, M. Xiao, L. Huang, Homing spread: Community home-based multi-copy routing in mobile social networks, Proceedings of INFOCOM, IEEE, 2013, pp. 2319–2327.
[94] M. Xiao, J. Wu, L. Huang, Community-aware opportunistic routing in mobile social networks, IEEE Trans. Comput. 1 (2013) 99. Q7
[95] F. Xing, W. Wang, Understanding dynamic denial of service attacks in mobile ad hoc networks, in: Proceedings of Military Communications Conference, MILCOM 2006, IEEE, 2006, pp. 1–7.
[96] D. Zhang, Z. Wang, B. Guo, X. Zhou, V. Raychoudhury, A dynamic community creation mechanism in opportunistic mobile social networks, Proceedings of IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and IEEE Third International Conference on Social Computing (Socialcom), IEEE, 2011, pp. 509–514.
[97] Y. Zhang, J. Zhao, Social network analysis on data diffusion in delay tolerant networks, Proceedings of the Tenth ACM International Symposium on Mobile Ad hoc Networking and Computing (MobiHoc '09), ACM, 2009, pp. 345–346.
[98] K. Zhu, W. Li, X. Fu, Rethinking routing information in mobile social networks: location-based or social-based? Comput. Commun. 42 (2014) 24–37.
[99] K. Zhu, W. Li, X. Fu, Smart: a social and mobile aware routing strategy for disruption tolerant networks, in: Proceedings of IEEE Transactions on Vehicular Technology, 2014.
[100] M. Zignani, Geo-comm: a geo-community based mobility model, Proceedings of Wireless On-demand Network systems and Services, WONS '12, IEEE, 2012, pp. 143–150.

Konglin Zhu received his master degree in Computer Science from UCLA and his PhD degree in University of Goettingen, 2014. He is now an assistant professor at Beijing University of Posts and Telecommunications. His research interest includes: Mobile social networking and online social networking data analysis. Q8