

Social-based Data Routing Strategies in Delay Tolerant Networks

Dissertation
for the award of the degree
“Doctor of Philosophy” Ph.D. Division of Mathematics and Nature Science
of the Georg-August-Universität Göttingen

within the doctoral program in Computer Science (PCS)
of the Georg-August University School of Science (GAUSS)

submitted by
Konglin Zhu

from Shandong, China
Göttingen, 2014

Thesis Committee

Prof. Dr. Xiaoming Fu

(Institute of Computer Science/Computer Networks Group, University of Göttingen)

Prof. Dr. Dieter Hogrefe

(Institute of Computer Science/Telematics Group, University of Goettingen)

Prof. Dr. Wenzhong Li

(Department of Computer Science/State Key Laboratory for Novel Software and Technology, Nanjing University)

Members of the Examination Board

Reviewer: Prof. Dr. Xiaoming Fu

(Institute of Computer Science/Computer Networks Group, University of Göttingen)

Second Reviewer: Prof. Dr. Mario Gerla

(Department of Computer Science/Network Research Lab, UCLA)

Further members of the Examination Board

Prof. Dr. Dieter Hogrefe

(Institute of Computer Science/Telematics Group, University of Göttingen)

Prof. Dr. Jens Grabowski

(Institute of Computer Science/Software Engineering for Distributed Systems, University of Göttingen)

Prof. Dr. Carsten Damm

(Institute of Computer Science/Theory and Algorithmic Methods, University of Göttingen)

Prof. Dr. Stephan Waack

(Institute of Computer Science/Theory and Algorithmic Methods, University of Göttingen)

Date of the oral examination: 25.02.2014

Abstract

Delay Tolerant networks (DTNs) are intermittently connected mobile networks in which the end-to-end paths do not exist. Data delivery in such networks relies upon the contacts that use “store-carry-and-forward” paradigm to forward message from one node to another. However, such intuitive methodology encounters low message delivery ratio and high data transmission delay when applying to different data routing strategies. The design of effective and efficient data routing strategies based on limited knowledge of mobile nodes in DTNs is challenging.

In this dissertation, we explore several aspects of social information that can be applied for data routing in DTNs. We discuss the problems of data routing in DTNs and study the using of different social information and network features to facilitate data routing in DTNs. Specifically, we propose three different data routing strategies relying on different types of social information obtained from mobile nodes: (a) a location-based social routing strategy applying different aspects of location-based social information; (b) an encounter-based social routing strategy relying on several encounter-based social factors of mobile nodes in the network; (c) a community-based routing strategy combining social and mobile factors as well as community structure.

The proposed location-based social routing strategy is motivated by the fact that location information can provide the geographic distance and the direction of information propagation, which can guide the data to the destination effectively. The proposed location-based social strategy considers both geographic distance and user mobility pattern as factors and combines them into one utility function for data forwarding.

We propose the encounter-based social routing strategy based on the fact that users in DTNs are interactively connected by encountering events. The design of encounter-based social strategy involves social centrality and social similarity. Compared to location-based social strategy, the usage of encounter-based social information is much less sensitive than location-based social information. By convoluting two social factors into utility function, the proposed algorithm can achieve competitive performance with location-based routing strategy.

The design of the community-based strategy is motivated by the observation that the mobility of people concentrates on a local area and the communication occurs in the form of communities. To apply such characteristics for elevating data routing performance in DTNs, we propose a Social and Mobile Aware Routing sTrategy (SMART). It exploits a distributed community partitioning algorithm to divide the DTN into communities regarding user locations and interaction routines. For intra-community communication, a decayed routing metric convoluting social similarity and social centrality is calculated, which is used to decide forwarding node efficiently while avoiding the newly identified blind spot and dead end problems. Meanwhile, to enable efficient inter-community communication, we choose the fringe nodes which travel remotely as relays, and propose the node-to-community utilities for routing decision across communities.

The major contribution of the thesis is to compose comprehensive routing metric to overcome the situation that is not addressed by using single routing metric, and then identify and tackle the blind spot and dead end problem, which are severe but not noticed in the existing studies. The proposed location-based strategy and encounter-based strategy are to construct comprehensively routing metric in geographic and encountering perspectives, and the proposed SMART is to tackle the blind spot and dead end problems.

Among all three strategies, the objective is to enhance the data delivery ratio, reduce the average delay and meanwhile maintain the low cost for data delivery. We present the simulation results regarding to the performance of the proposed routing strategies with the state-of-the-art data routing strategies in DTNs. By comprehensively consider multiple aspects of routing metrics, the proposed location-based and encounter-based routing strategies outperforms the previous studies around 10% in terms of different evaluation metrics. Through identifying and solve blind spot and dead end problems, the proposed SMART resolves both of them and thus outperforms previous studies over 20%.

Acknowledgements

This dissertation is the last journey of my Ph.D study in Computer Science at Geoge-August-Universität Göttingen. I appreciate many people who have helped me to go through this colorful and meaningful four years.

First of all, I would like to show my deepest gratitude to my supervisor Professor Xiaoming Fu. I appreciate him for his continuous support and encouragement throughout my research. I admire his open, rigorous and dedicated attitude towards the scientific research and life. At every step of my thesis, he guided me with his profound knowledge and insight. I sincerely thank him.

I also appreciate my co-supervisor Dr. Wenzhong Li. I am grateful for every step that he has taught and guided me in the four years. I cannot accomplish my Ph.D study without his help.

I would also like to show my thanks to Professor Mario Gerla, Professor Dieter Hogrefe, Professor Jens Grabowski, Professor Carsten Damm and Professor Stephan Waack for being members of my committee. Thanks for their advice and comments through forming my thesis work.

I wish to thank all the research staff in Computer Networks (NET) Group. I really appreciate people who have studied or worked in NET group for their help during my Ph.D study. We had wonderful discussions about academics and lives. I learnt a lot from them.

Finally, I show my deep gratitude to my parents. Thanks for their support and give me whatever they can. They always stand by me and encourage me. I appreciate they provide me the life and let me live in a happily family. I love them and would like to dedicate this thesis to them.

Table of Contents

Abstract	i
Acknowledgements	iii
Table of Contents	v
List of Figures	ix
List of Tables	xiii
1 Introduction	1
1.1 Delay Tolerant Network	1
1.2 Data Routing Problem in DTNs: in Social Perspective	2
1.3 Our Contributions	5
1.4 Thesis Structure	6
2 Related Works	7
2.1 Location-based DTN Routing	7
2.1.1 Geographic distance	7
2.1.2 Mobility pattern	9
2.2 Encounter-based DTN Routing	10
2.2.1 Direct encountering	10
2.2.2 Social information derived from encountering	12
2.3 Community-based DTN Routing	14
3 Conceptual Framework	19
3.1 Problem Statement	19
3.1.1 Motivation and challenges	19
3.1.2 Research statement	20
3.2 Network Model	22
3.2.1 Modeling social graph	22

Table of Contents

3.2.2	Data sets	22
3.3	Basic Approach	24
4	Location-based Routing Strategy	27
4.1	Location-based Social Information	27
4.1.1	Location-based graph	27
4.1.2	Geographical distance	28
4.1.3	Similarity of mobility pattern	29
4.2	Strategy Design	30
4.3	Performance Evaluation	33
4.3.1	Experiment setup	33
4.3.2	Strategies in comparison	33
4.3.3	Performance analysis	35
4.4	Summary of Contributions	38
5	Encounter-based Routing Strategy	39
5.1	Encounter-based social information	39
5.1.1	Encounter-based social graph	39
5.1.2	Social similarity	40
5.1.3	Social centrality	40
5.2	Strategy Design	41
5.3	Performance Evaluation	43
5.3.1	Experiment setup	43
5.3.2	Strategies in comparison	44
5.3.3	Performance analysis	46
5.3.4	Compare with the Loc strategy	50
5.4	Summary of Contributions	55
6	Community-based Routing Strategy	57
6.1	Mobile and Social Characteristics of DTN	59
6.1.1	Distributed community partitioning	61
6.1.2	Locality of user contacts	64
6.2	Strategy Design	65
6.2.1	Assumptions	65
6.2.2	Intra-community communication	66
6.2.3	Inter-community communication	69
6.3	Discussion	71

6.3.1	Tackling blind spot and dead end problems	71
6.3.2	Efficiency of inter-community communication	72
6.4	Performance Evaluation	74
6.4.1	Experiment setup	74
6.4.2	Impact of community numbers	74
6.4.3	Impact of community partitioning algorithms	75
6.4.4	Performance analysis	77
6.5	Summary of Contributions	82
7	Discussion and Future Works	85
7.1	A Comparison of Three Strategies	85
7.2	Future Works	88
8	Conclusions	91
	Bibliography	93

List of Figures

1.1	Three snapshots a Delay Tolerant Network. A solid line suggests the connectivity between two nodes.	2
1.2	Location-based routing vs. Encounter-based routing.	4
4.1	An example of visiting locations	30
4.2	Performance comparison of location-based strategies on MIT Reality data trace	34
4.3	Performance comparison of location-based strategies on DieselNet data trace	35
4.4	Performance comparison of location-based strategies on Cabspotting data trace	36
4.5	Performance comparison of location-based strategies on synthetic data trace	37
5.1	Cumulative distribution function of shortest path lengths	44
5.2	Performance of encounter-based social schemes on MIT Reality data trace	45
5.3	Performance of encounter-based social schemes on DieselNet data trace	46
5.4	Performance of encounter-based social schemes on Cabspotting data trace	47
5.5	Performance of encounter-based social schemes on synthetic data trace	48
5.6	Performance comparison of Soc and Loc strategies on MIT Reality data trace	49
5.7	Performance comparison of Soc and Loc strategies on DieselNet data trace	50
5.8	Performance comparison of Soc and Loc strategies on Cabspotting data trace	51

List of Figures

5.9	Performance comparison of Soc and Loc strategies on synthetic data trace	52
5.10	Performance comparison of Soc and Loc strategies on synthetic data traces as a function of node speed	53
5.11	Performance comparison of Soc and Loc on synthetic data traces as a function of network size	54
6.1	Proportion of blinds spots and dead ends in utility-based strategies .	58
6.2	Two taxi trajectories in Cabspotting trace	60
6.3	The CDF and PDF of node movements (Cabspotting).	61
6.4	The number of encounters vs. distance (Cabspotting).	62
6.5	Local contact and remote contact	64
6.6	Percentage of blind spot and dead end in SMART.	72
6.7	Delivery ratio of community-based strategies	73
6.8	The performance metrics as a function of community number and time (MIT Reality).	75
6.9	The performance metrics as a function of community number and time (DieselNet).	76
6.10	The performance metrics as a function of community number and time (Cabspotting).	77
6.11	The performance of SMART under different community partitioning algorithms (MIT Reality).	78
6.12	The performance of SMART under different community partitioning algorithms (DieselNet).	79
6.13	The performance of SMART under different community partitioning algorithms (Cabspotting).	80
6.14	The performance comparison of various strategies on MIT Reality Mining trace	81
6.15	The performance comparison of various strategies on DieselNet trace	82
6.16	The performance comparison of various strategies on Cabspotting trace	83
7.1	Performance comparison of three social-based strategies on MIT Reality data trace	86
7.2	Performance comparison of three social-based strategies on DieselNet data trace	87

7.3	Performance comparison of three social-based strategies on Cab-spotting data trace	88
-----	--	----

List of Tables

2.1	Comparison of three categories routing strategies	17
3.1	Statistics of three real data traces	23
3.2	Parameters of synthetic data traces	23
3.3	Summary of three proposed routing strategies	25
4.1	Table of visiting history	31
5.1	Table of encounter history	41
6.1	Routing utilities in DTNs	57
6.2	Proportion of local and remote contacts	65
6.3	Local contact table	66
6.4	Remote contact table	66

Chapter 1

Introduction

A delay tolerant network (DTN) is a sparse dynamic wireless network where mobile nodes work on ad hoc mode and forward data opportunistically upon contacts [27]. Since the DTN is sparse and nodes in the network are dynamic, the end-to-end path rarely exists. The communication of nodes can only be conducted when they are in the communication range of each other. When a node has a copy of message, it will store the message and carry it until forwarding the message to a node in the communication range which is more appropriate for the message delivery.

Since DTNs allow people to communicate without network infrastructure, they are widely used in battlefield, wildlife tracking, and vehicular communication etc. where setting up network infrastructure is hard and costly [56, 15, 7]. In recent years, with the proliferation of social network applications and mobile devices, people tend to share texts, photos and videos with others via mobile devices in DTNs [86, 62, 55, 101].

1.1 Delay Tolerant Network

Delay Tolerant Networks (DTNs) are described as a kind of network where the nodes in the network are mobile and the connections among nodes are changing over time thus the communication between nodes is opportunistically occurs only when they are in communication range. Due to network structure of DTNs, they are characterized by large delays, frequent disruptions and lack of stationary paths between nodes. Such network can be constructed by human beings [62, 34, 37, 66], wildlife [43, 82], or even vehicles [99, 74, 81].

We use an example to illustrate the main characteristics of DTNs. Fig. 1.1 shows a sample delay tolerant network. It depicts the network topology snapshots over three different time periods t_1 , t_2 and t_3 ($t_1 < t_2 < t_3$). The movements of nodes lead to the positions of them different from one snapshot to another. Node mobility leads to several pairs of nodes moving into communication range (e.g., node A and B cannot communicate at t_1 , but they run into communication range at

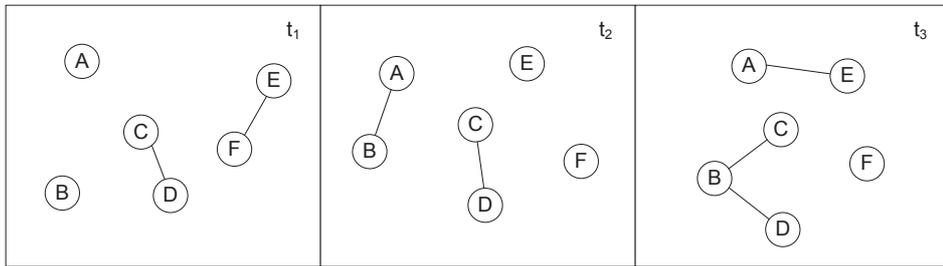


Figure 1.1: Three snapshots a Delay Tolerant Network. A solid line suggests the connectivity between two nodes.

t_2) or moving out of communication range (e.g., node C and D are in communication range at t_1 and t_2 , but they becomes unreachable at t_3). Therefore, the stable end-to-end path does not exist between any couple of nodes. The communications between a pair of nodes are often disrupted due to unstable connections. Besides, if a node wants to send a message to another node, it may suffer from large delay. This is because the data transmission between any pair of nodes needs that they are in communication range. However, delay tolerant network does not guarantee that two nodes are in communication range permanently. It may spend a long time period for two nodes to move into communication range. Thus the communication delay between two nodes is longer than wired networks like Internet. For instance, if node A needs to send a message to node E in the sample DTN, it can only deliver the data to node E at t_3 when they are in communication range at this time period.

1.2 Data Routing Problem in DTNs: in Social Perspective

Although the end-to-end path rarely exists in delay tolerant networks, the communication among nodes in such network is still desirable. Therefore, an effective and efficient data routing strategy is needed to enable the communication in the intermittent connected networks. Although there are numerous data routing schemes designed for wireless network, they cannot be directly applied to DTNs.

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 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 DTN 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 DTNs 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 the relay selection.

The most naive strategy such as Epidemic routing [89] is proposed to send data epidemically as long as two nodes encounter until the destination is reached. Based on such epidemic principle, many routing schemes [83, 59, 64, 14, 43, 3] using limited copies of messages are developed. Such epidemic based routing strategies suffer from extremely high network cost.

Since nodes in DTNs 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. For example, people with similar social properties may spend long time together, and be willing to sharing information and resources with each other [96]. By exploring social features in DTNs, the prediction of encountering opportunities of nodes will be more effective. Therefore, many routing schemes are developed based on social information. Generally speaking, there are two kinds of social information widely used for data routing in DTNs: location-based social information and encounter-based social information. The location-based social information refers to as the geographical related data, including the geographical coordinates [87], the distance between individuals [45] and etc., which represents the physical property of human activity. The encounter-based social information is defined as the inferred human relationship from encountering events. It can be the contacts of individuals [56], social centrality [40], social similarity [24] and etc., which represents the logical information of human interactions. Accordingly, routing strategies in DTNs can be divided into two categories: the location-based routing strategies [18, 49, 45, 87] and the encounter-based routing strategies [7, 40, 33, 24, 94, 83].

There are advantages and disadvantages of both kinds of strategies. On the one hand, location-based strategies forward data to the nodes geographically closer

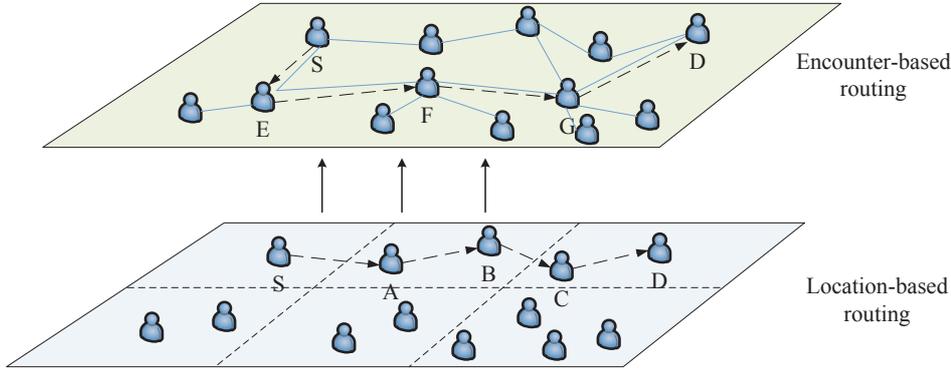


Figure 1.2: Location-based routing vs. Encounter-based routing.

to the destination, which tend to achieve geographical shortest routing path. An example is shown in Fig. 1.2, where node S wants to send a message to node D . The lower layer indicates the physical locations of the mobile devices. Based on the measurement of geographical distance, it tends to choose $S \rightarrow A \rightarrow B \rightarrow C \rightarrow D$ as the shortest routing path. However, since encounter opportunity is not taken into account, such routing path may not be efficient and the delivery ratio is not guaranteed. On the other hand, the encounter-based strategies forward data to the nodes logically closer to the destination, which tend to achieve logical shortest routing path. As in the example of Fig. 1.2, the upper layer indicates the encounter-based social connections (in solid lines) of nodes. Based on the measurement of their connections, S tends to forward data via the path $S \rightarrow E \rightarrow F \rightarrow G \rightarrow D$ (since the number of connections are $S < E < F < G$). Such forwarding strategy seems to enhance the chance of data delivery, but due to unawareness of location, it may also lead to a longer routing path and higher delay.

Besides, the social structure, such as community, is also important for data routing in DTNs. People from the same community may contact each other more frequently. The community-based strategies forward data according community structure based on the fact that people tend to group into communities by their social relationships. By dividing the network into multiple communities, the nodes within a community have strong connections, while their links across communities are weak ties. The community structure favors intra-community communication where nodes are closely connected, but also encounters the difficulty of inter-community communication via weak links.

In this dissertation, we apply different types of social information and structure to devise data routing strategies in DTNs. In each strategy, we exert different routing

mechanisms to enhance data delivery ratio, reduce average delay and meanwhile maintain low cost.

1.3 Our Contributions

In this thesis, we apply social information to enhance performance of data routing efficiency in delay tolerant networks. We firstly propose a location-based routing scheme which comprehensively combines multiple factors in geographic perspective to elevate routing performance. Considering the privacy of location-based social information, we then devise an encounter-based routing strategy. Since the encountering event is the fundamental information for data forwarding in DTN, the leverage of encounter information is less sensitive. The proposed encounter-based routing scheme considers several encounter-based social factors to achieve competitive performance. Finally, to solve several existing problems in current DTN data routing schemes, we propose a new algorithm called SMART, which relies on the social properties and community structure of DTNs to facilitate data routing in such networks.

Contributions of this dissertation are summarized as follows:

- To confront the challenge that routing metrics relying on one aspect of node or network features cannot fully adapt to diverse network situations, we devise a comprehensive location-based data routing scheme in DTNs. We model DTN using a location-based social graph and define geographic distance and similarity of mobility pattern relying on the graph. The proposed location-based social strategy utilizes the combination of similarity of mobility pattern and geographic distance to enhance the data routing efficiency. The simulation results show that the proposed location-based social strategy outperforms other location-based strategies by around 10%.
- Due to the fact that location-based social information is sensitive to users in DTNs, while encounter-based social information is less sensitive, the encounter-based social information is applied for data routing in DTN. We model DTN using an encounter-based social graph and define social similarity and social centrality based on the graph. We propose an encounter-based routing scheme which comprehensively combines social similarity and social centrality by convolution. The proposed strategy outperforms other encounter-based rout-

ing schemes by 15% and has competitive performance with location-based routing strategy.

- We identify blind spot and dead end problems that exist in most of utility-based routing schemes in DTNs. By measuring the blind spot and dead end problems in several existing routing strategies, we discover that the proportion of these two problems reaches about 20%.
- We design a distributed community detection algorithm based on node encounters and carry out a community-based routing strategy called SMART. It divides a DTN into several communities, and exploits different principles for data routing depending on whether the source and the destination are in the same community. Routing utilities integrating different metrics with convolution and decay function are explored to overcome the difficulties of intra- and inter-community communications.

1.4 Thesis Structure

The remainder of the thesis is constructed as follows. We summarize the related works in literature in Chapter 2. We classify data routing schemes into different kinds of groups according to different aspects of social information. Specifically, the routing schemes are categorized into location-based routing strategies, encounter-based routing strategies and community-based routing strategies. In Chapter 3, we discuss the motivation and challenges of this thesis. We also depict the basic approach of this work. In Chapter 4, we devise a location-based routing strategy to enhance the efficiency of data routing. In Chapter 5, we propose an encounter-based routing scheme based on the fact that encounter-based social information is less sensitive compared with location-based social information. We study the problems in existing data routing schemes, carry out a distributed community detection method and propose a community-based routing scheme named as SMART in Chapter 6 to elevate the performance of data routing in DTNs. In Chapter 7, we conduct a comprehensive comparison of three proposed routing strategies and discuss the future works. We conclude the dissertation in Chapter 8.

Chapter 2

Related Works

Delay tolerant networks have been proposed for more than one decade [27]. Researchers focus on the data routing, one primary issue in DTNs, and many studies have been carried out to handle the data delivery in the intermittent connected environment. This thesis mainly focuses on the routing strategies relying on social information. Specifically, we divide social information into three categories: location-based social information, encounter-based social information and social community. Hence, the corresponding routing strategies are divided as location-based routing strategies, encounter-based strategies and community-based strategies.

2.1 Location-based DTN Routing

Geographic information, as one aspect of social information, is well applied for data delivery in DTNs. The derived location-based routing strategies in wireless networks were widely studied in the past decade. They make forwarding decision according to the geographic information. Specifically, the location-based routing strategies are categorized based on the exerted geographic information: (1) geographical distance, and (2) mobility pattern. In the following, we present two groups of location-based routing strategies respectively.

2.1.1 Geographic distance

As one of the earliest works on wireless routing strategies using geographic information, B. Karp proposed Greedy Perimeter Stateless Routing (GPSR) [45] for wireless mobile ad hoc networks. It makes greedy forwarding using the geographic positions of a router's neighbors in the network topology. Specifically, a node obtains its neighbors' positions by information exchange. Then it locally chooses the optimal next hop with the neighbor geographically closest to the packet's destination. Forwarding in this regime follows successively closer geographic hops until

the destination is reached. When a packet reaches a region where greedy forwarding is impossible (i.e. packets reaches the local maxima), the algorithm recovers by routing around the perimeter of the region. It exploits the right-hand rule which traverses the interior of a closed polygonal region in clockwise edge order to seek for the next hop. However, the application of GPSR needs planar graph.

Besides, authors in [57] proposed Greedy Perimeter Coordinator Routing (GPCR) by taking advantage of the fact that streets and junctions form a natural planar graph to handle data routing. It contains a restricted greedy forwarding procedure and a repair strategy. In the greedy mode, the data packets are forwarded to a node at a junction. Then junction node forwards packets by choosing one neighbor which has the shortest distance to destination. In the perimeter mode, it also applies right-hand rule [28] when the greedy forwarding is impossible. GPCR assumes that there is always a node at a junction. When the junction node is missing, it causes routing loops and packet loss.

A work [93] named Mobility-Centric Data Dissemination Algorithm for Vehicle Network (MDDV) applies similar idea for data forwarding. It exploits Global Position System (GPS) embedded in each node and the bundles are forwarded to neighboring nodes that are geographically closest to the region containing the destination. There is no route recovery mechanism in this work. It does not consider the situation that the geographic routing is not applicable.

GeOpps [50] takes the advantages of the suggested routes by navigation systems to seek for the data relay that is likely to move closer to the destination node. It calculates the shortest path between the destination node and the nearest point of the path and estimate the time for arriving at the destination. During the routing process, if a node with shorter estimated arrival time appears, the data packet will be forwarded to the node. The process is conducted repeatedly until the destination is reached. Obviously, GeOpps requires navigation information and the estimation of arrival time needs global view of the network. Both of them are difficult to be accomplished in DTNs.

The hybrid routing strategy GeoDTN+Nav [21] is proposed with three different modes: a greedy mode, a perimeter mode, and a DTN mode. It switches between non-DTN mode and DTN mode by evaluating the network connectivity based on the number of hops a packet has traveled, neighbor's delivery quality and neighbor's direction to the destination. To achieve this, it uses Virtual Navigation Interface (VNI) to abstract information from underlying hardware and provides necessary information for the strategy to determine its routing mode. Besides, VNI provides the

option to protect nodes' private information. Such hybrid routing strategy brings the longer delay and higher packet loss issues when it conducts the switching operation.

Locus [87] proposes a location-based overlay for DTNs. It utilizes devices nearby to keep the data in a specific area. To enable the data routing, it defines a utility function based on the geographic distance from a specific location to a data's home location, and finds the node having the closest geographic distance with the data item for data forwarding. Locus requires there always some nodes in the data home location and multiple copies of data are needed.

In summary, the data routing strategies rely on geographical distance mainly initiate their relay selection by greedy forwarding. When greedy forwarding is failed, different repair strategies are applied to reduce the chance of packet loss.

2.1.2 Mobility pattern

Compared with geographic distance, mobility pattern is defined in a more sophisticated manner. It may refer to many different characteristics of user movements, such as the node trajectories, visiting histories and etc.

The Utility-based Distributed routing algorithm with Multi-copies (UDM) [52] defines "home community" where the nodes passed by or stayed close to most frequently. It selects the data relay as the node with the highest utility value to the "home community". Besides, it exploits binary transmission, which means that when it finds a proper relay, the node sends half number of its packet copies to the new node. The process continues until the destination is reached.

A similar work named MV routing was proposed in [18]. It also proposes to forward data packet to a stationary location. The difference is that MV routing uses the meeting frequencies and visits to the stationary location to construct the utility function. Both methods need multiple copies of data packets and the relays used for data delivery are also difficult to be determined.

Besides, both MoVe [48] and VeRo [44] apply movement vectors for data routing schemes design. Specifically, MoVe exerts nodes velocity and direction to estimate the shortest path to the destination. When two nodes encounter, they exchange the trajectory and bundles decide whether to be forwarded by the direction and distance between the candidate node and destination. Similarly, nodes in VeRo records their position and angle changes, and exchange the data to a node that is moving away from it. The limitation of both movement vector based strategies is that they

need to exchange the trajectory of nodes and the load balance is difficult to achieve.

Furthermore, CAR [65] uses the probability that two nodes visit the same location (colocation) combining the degree change of the node to calculate the probability that a node can successfully deliver a message to a destination. However, the colocation information cannot represent the mobility pattern of a node well. For example, a node may visit a location with time duration one hour, while another node only comes to the same location in 10 seconds. Although they are co-located, the probability for successful data delivery is not high.

To represent mobility pattern more accurately in DTNs, Mobyspace [49] calculates the similarity of mobility pattern by Euclidean distance of two nodes' visiting history and chooses the node with shorter distance (i.e., more similar mobility pattern) with the destination node as the relay for data delivery. Specifically, it considers a node's visiting history as a vector. Each value in the vector represents the percentage of time that the node stays at the location. The distance between two nodes is computed by the Euclidean distance between two vectors. Although Mobyspace can represent mobility pattern of nodes more specifically, it ignores the temporal factors that lead to the change of mobility patterns in DTNs.

Overall, the routing strategies based on mobility pattern construct the utility value according to statistical results of the mobility characteristics. It may be more accurate than directly using geographic distance. However, due to its requirement for detailed location information, the privacy concern of users is still a great concern.

2.2 Encounter-based DTN Routing

Generally, encounter-based routing strategies make forwarding decision relying on the encounters of nodes. In this thesis, we investigate the encounter information in social perspective. They are also mainly divided into two types: (1) directly encountering, and (2) social information derived from encountering.

2.2.1 Direct encountering

There are some strategies directly using encounter information for data routing. For instance, Prophet [56], RAPID [7], MaxProp [15] 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.

2.2. Encounter-based DTN Routing

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 encounter 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 node a and b :

$$P_{(a,b)} = P_{(a,b)_{old}} + (1 - P_{(a,b)_{old}}) * P_{init}, \quad (2.1)$$

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:

$$P_{(a,b)_{old}} = P_{(a,b)_{old}} * \gamma^\kappa, \quad (2.2)$$

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:

$$P_{(a,c)} = P_{(a,c)_{old}} * P_{(a,b)} * P_{(b,c)} * \beta. \quad (2.3)$$

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 encounter node.

Jain et al. [42] presented a routing metric named as Minimum Expected Delay (MED) by assuming future contact periods are known. They modify the Dijkstra algorithm to compute the path for DTN with minimum delay. However, such calculation can only adapt to certain types of DTNs. To address this limitation, they propose a new metric, named as Minimum Estimated Expected Delay (MEED), which is calculated by past encounter history and then floods the metric value within the whole network. It introduces too much control overhead.

Spyropoulos et al. proposed a series of multi-copy data delivery schemes, such as Spray and Wait [83] and Spray and Focus [84]. Spray and Wait simply spread the messages to nodes it encounters and each data carrier waits until it meets destination. Obviously, it has significant waste of data and also there is no any

criteria for the selection of data relays. To address this issue, the Spray and Focus is 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 value is used to determine whether the node is a good relay for data delivery. If its utility value is larger than the data carrier, then the bundles will be forwarded.

The MaxProp [15] is 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] is proposed by taking the constraint resource into account. It calculates utility functions according to different resource constraints. The bundles are forwarded to nodes with higher utility.

In summary, the directly encounter based routing schemes enhance the performance for data delivery by calculating encounter-based utilities. However it requires exchanging encounter information of nodes in the network, which introduces large amount of control overhead.

2.2.2 Social information derived from encountering

Another group of routing strategies rely on the social information derived from encounter-based social graph. Although they do not directly use encounter information, most of these works are based on the encounter-based social graph.

SimBet [24] 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:

$$SimUtil_n(d) = \frac{Sim_n(d)}{Sim_n(d) + Sim_m(d)}, \quad (2.4)$$

$$BetUtil_n = \frac{Bet_n}{Bet_n + Bet_m}. \quad (2.5)$$

The overall utility is combined as:

$$SimBetUtil_n(d) = \alpha SimUtil_n(d) + \beta BetUtil_n, \quad (2.6)$$

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 [32].

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 increase the utilization of content replication in intermediate nodes.

Zhang et al. proposed a data diffusion strategy based on "homophily" [100]. 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. Specifically, 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 [41] 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.

Social feature-based algorithm [94] takes the multi-dimension social properties and chooses the node with most similar social features as the destination for data forwarding. Specifically, it conducts hypercube by various social properties and uses the feature distance to measure the closeness between two nodes. The node with the closest social features will be selected as the relay for data delivery.

Alternatively, PeopleRank [63] tries to rank nodes in a social graph in a dis-

tributed manner. It measures the relative importance of a node in the network and the message is forwarded to nodes with same or higher rankings.

Fabbri and Verdone proposed a sociability-based routing strategy in [26]. It exerts the nodes with high degrees of sociability 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.

A social-aware and stateless routing (SANE) [60] is proposed by the observation that people with similar interest are more often to meet each other. It uses 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 the 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.

Li and Shen proposed a duration utility-based social routing scheme named SEDUM [53]. 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.

In summary, social routing based on different kinds of social properties derived from encounter-based graph. It enhances the performance from social perspective. However, the enhancement is still limited only based on these social properties. In the last section of this chapter, we will review routing schemes relying on another important structure feature in social networks: community-based routing strategies.

2.3 Community-based DTN Routing

Community as a very important social structure is applied to enhance the performance of data routing in DTNs. 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

2.3. Community-based DTN Routing

communities are weak. There are several routing strategies [39, 40, 33, 51, 13, 1] exploiting community structure for data routing in DTNs.

One of the earliest works named label routing using community structure for DTN routing is proposed by Hui and Crowcroft [39]. The data routing mechanism is built on Pocket Switched Networks (PSNs) [38], 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. 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 [40] considers the data routing in PSN which consists of several communities and there are social relationships among users. It uses k-clique percolation as the basic community detection method. 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. [33]. It presents multicasting path selection based on social centrality and social community. In the case of single data multicasting, it measures the social centrality, and chooses the node with higher centrality value as the successor for data forwarding. 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 social properties to find the destination.

LocalCom [51] 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. Then the marking and pruning schemes are used: static pre-pruning schemes and dynamic pruning. In case that the source and destination nodes of a packet reside in different communities, 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) [13]. 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 [95] is a zero-knowledge multi-copy routing algorithms. It defines community homes, which are considered as the common locations visited by a group of people with same interest. 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 [97] 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 expected 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 DTNs [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.

In summary, community-based routing strategies try to improve data forwarding

2.3. Community-based DTN Routing

efficiency by community structure. However, most existing community partitioning is 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.

Overall, the major characteristics of three categories are summarized in Table 2.1.

Table 2.1: Comparison of three categories routing strategies

Metrics	Routing Strategies	Encounter Information	Location Information	Feature
Location-based: geographic distance	GPSR [45], GPCR [57], MDDV [93], GeOpps [50], GeoDTN+Nav [21], Locus [87]	Yes	Yes	Forward data to node with closer distance to destination node or location
Location-based: mobility pattern	UDM [52], MV [18], MoVe [48], VeRo [44], CAR [65], Mobospace [49]	Yes	Yes	Forward data to node with more similar mobility pattern with destination
Encounter-based: direct encounter	Prophet [56], RAPID [7], MaxProp [15], MED [42], Spray and Focus [84]	Yes	No	Forward data to node with higher encounter frequency or duration with destination
Encounter-based: derived social information	SimBet [24], social relation aware routing protocol [4], SDM [33], Social greedy [41], PeopleRank [63], SANE [60], SEDUM [53]	Yes	No	Forward data to node more socially similar with destination
Community structure	LABEL [39], Bubble Rap [40], MDM [33], LocalCom [51], FBR [13], Homing spread [95], Community-aware opportunistic routing [97], SGBR [1]	Yes	No	Forward data according to community structure

Chapter 3

Conceptual Framework

In this chapter, we describe the motivation and challenges for data routing in delay tolerant networks. We present the research statement of the thesis and give an overview of the network model and basic approach that we use for routing strategies designing.

3.1 Problem Statement

We discuss the motivation and challenges for data dissemination in DTNs and outline the research statement in this section.

3.1.1 Motivation and challenges

A key problem in DTN is data dissemination. The accomplishment of data dissemination requires effective data routing strategies that can address the following challenges in DTN:

- *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.
- *limited network information*. Due to the fact of dynamic network and unstable connections among nodes, they cannot obtain all network information from DTN. It makes the traditional mobile ad hoc routing protocols (such as AODV [76], DSDV [75] and etc.) cannot adapt to DTN 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 makes the calculation of best paths for different destinations become challenging.

- *uncertain connection duration and limited resources.* Data dissemination in DTNs also refers to the size of the data. Due to node movements, the connection duration between two nodes is unknown and difficult to be predicted. Long connection duration can help to transmit a large number of messages or messages with large sizes. Therefore, 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 delay tolerant networks, deciding the number of messages and the size of data for transmission is also affected by the resource of nodes. Nodes in DTNs 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 workplace to show these challenges in DTNs. Consider the DTN 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.

3.1.2 Research statement

In this dissemination, we investigate data routing strategies for data dissemination in DTNs from social perspective. The previous proposed routing strategies are proposed to address the above-mentioned challenges in DTNs. They are divided into three main categories based on the social information they used for data routing, as explained in Chapter 2: location-based, encounter-based and community-based. Location-based routing schemes make the forwarding decision based on the

3.1. Problem Statement

location-based social information, such as geographic distance, colocation and etc. Encounter-based routing schemes construct the social graph based on encounter information and the derived social properties for data routing. Community-based strategies exert community structure to facilitate data routing in DTNs.

The key issue of existing social-based routing schemes is lack of researches that utilize comprehensive social information for data routing. Thus, the performance needs further enhancement. Specifically, current researches construct the utility metric based on one aspect of information. It may not adapt to different situations of network topologies and dynamics. For example, the location-based routing schemes relying on geographic distance (e.g., GPSR [45]) only consider the distance between nodes temporally, which is not efficient when the update of distance information is not frequent, while location-based routing strategies relying on mobility pattern (e.g., Mobyspace [49]) take statistical mobility patterns as the major concern. Due to lack of distance information, the delivery ratio cannot be guaranteed. Similar situation can also be found in encounter-based social routing schemes that only consider one aspect of encounter information thus cannot fully represent the situation of the network. In this thesis, we propose two comprehensive routing strategies from the geographical and encountering social perspective.

Furthermore, the existing routing schemes rely on utility to make forwarding decision. That is, for a certain destination, a node calculates a utility value based on network structures or node features. When the node encounters another node in DTN, it compares the utility value with the encountering node. If the encountering node has higher utility value, data will be forwarded to it. However, such greedy forwarding schemes run into two issues: blind spot and dead end. Blind spot results from the utility value of a node so close to utility values of its neighbors that the node cannot find the next data relay. Dead end is because of the highest local utility value that the data is stuck into the node until it expires. By dividing the network into multiple communities, the nodes within a community have strong connections, while their links across communities are weak ties [35]. The community structure favors intra-community communication where nodes are closely connected. Although community structure is applied to reduce the chance of blind spot and dead end, it brings new issue that the communication among communities becomes difficult. We propose a social and mobile aware routing strategy that addresses both blind spot and dead end problem, and meanwhile achieves efficient inter-community communication.

3.2 Network Model

In this section, we provide the basic network model for design of various social-based routing schemes. We also describe the data sets we use throughout the thesis.

3.2.1 Modeling social graph

Delay tolerant networks can be described as a graph according to different characteristics of nodes and network structure [61, 42, 19, 73, 5]. In this dissertation, we use social features to characterize the network graph. We model a DTN 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.

Delay tolerant networks possess two basic elements: the encountering events between nodes and the geographic information of each node. These two elements describe the fundamental channel for communication as well as the dynamics of the network. In social perspective, people moves in the network leading to encounters. Both location information and encountering events are characterized as social information. According to the location-based and encounter-based social information, nodes in the network are grouped into different communities. This community structure makes nodes in one community are highly social related while nodes in different communities are less socially connected. In the rest of the thesis, we discuss the social graph in aspects of geographical locations and encountering events, as well as the community structure in the network.

3.2.2 Data sets

We use two types of data sets for evaluating the proposed routing strategies: real data traces and synthetic data traces.

Real data traces

We use the MIT Reality [25], DieselNet [16] and Cabspotting [77] three real data traces to characterize delay tolerant networks. The MIT Reality data trace consists of 97 users equipped with smart phones at MIT over the course of the 2004-2005 academic year. It records the information such as call logs, Bluetooth devices in proximity, cell tower IDs, application usage and phone status. Over the whole

3.2. Network Model

experimental period, it covers more than 30 thousand cellular towers. DieselNet logs mobility traces of 34 buses in Amherst, covering an area of more than 350 km^2 . Each bus is equipped with a computer and a GPS. It records the GPS logs for all the buses during the 20 days experiment time period 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, which covers over 2,000 km^2 . The statistics of the three mobility traces are summarized in Table 3.1. 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. All three data sets present the human being activity, including user location information and encountering events between nodes. We consider them as representatives of delay tolerant networks.

Table 3.1: Statistics of three real data traces

Traces	MIT Reality	DieselNet	Cabspotting
No. of devices	97	34	536
No. of contacts	54,667	2,284	111,153
Duration (days)	246	20	30
Contact rate	0.024	0.10	0.013
Field size (km^2)	N/A	358	2,367

Synthetic data traces

To provide general assessing of routing strategies, we produce a group of synthetic data sets to conduct the comprehensive comparisons.

Table 3.2: Parameters of synthetic data traces

No. of nodes	20 to 100
Node speed	0.5m/s to 2.5m/s
Duration	14 days
Field size	48 km^2

We use SUMO simulator [9] to mimic nodes' movements by generating random trips during a period of two weeks. The experiment area of the synthetic trace is chosen as MIT campus and its surroundings with a rectangle covering 48 km^2 ($6km * 8km$). The node speed (ns) (by walking) in one trace is constant and starts

from 0.5 m/s with a 0.5 m/s increment for each trace. Therefore, we generate 5 synthetic data traces with different node speeds. Meanwhile, we generate 5 synthetic data traces with different number of nodes (nn) ranging from 20 to 100. For each synthetic trace, we record the movements of each node in the experiment area and extract the encounter-based and location-based social information. The parameters of synthetic data traces are summarized as Table 3.2.

3.3 Basic Approach

In order to cope with intermittent connectivity, and use predicted and opportunistic connectivity to serve for data routing in Delay Tolerant Networks, known as Bundle Protocol [17], we proposed three social-based data routing strategies. Specifically, to enhance the performance of DTN routing, we propose two comprehensive routing strategies with one exploiting location-based social information and the other one exerting encounter-based social information. Then we propose a third data routing strategy using community structure to solve blind spot and dead end problems as mentioned in the above section, and meanwhile it achieves efficient inter-community communication.

The location-based social routing strategy [103], named as Loc, combines two metrics, similarity of mobility pattern and geographic distance to construct the comprehensive location-based data routing scheme. For the design of Loc, we assume that each node needs to know the realtime position of its own, which means that every node in the network is equipped with additional devices (such as Geographic Positioning System (GPS)) to aware of its position. They exchange location information when they encounter. The efficient manner for location information exchange can be found in [36].

However, the utilization of location-based social information must be very careful since it is much sensitive and private concern to users in the network. Using location information may violate user privacy. Malicious node can apply the collected location information to realize the mobility patterns of others in the network, which may be used for node tracking. Meanwhile, collecting location information needs additional equipments, such as GPS. Compared with location-based social information, encounter-based social information is less sensitive and easy to obtain. Therefore, we propose the encounter-based social routing strategy [103], called Soc, consisting of two social properties, social centrality and social similarity as basic factors for data routing.

3.3. Basic Approach

Additionally, blind spot and dead end problems lead to the message expired before reaching the destination, which significantly decreases the delivery ratio of routing strategies. To reduce the chance of blind spot and dead end and meanwhile achieve efficient inter-community communication. We devise a community-based social routing strategy named as SMART [102]. The proposed community-based strategy works as follows. It first introduces a distributed community partitioning method based on the observation that movements of DTN nodes are regular and restricted in local areas where more encounters occur than that in remote areas. With distributed community partitioning, mobile nodes can flexibly adjust their community IDs to assign with the group they most frequently encounter, and the community structure is formed by exchanging only local information, which is easy to be implemented in DTNs. For intra-community communications, the routing utility is calculated by integrating the convolution of social similarity and social centrality with a decay function. For inter-community communications, nodes frequently traveling across communities are chosen as “fringe nodes”, and the utilities of communicating between fringe nodes and communities are measured for routing decision, which enhances the delivery ratio effectively. Table 3.3

Table 3.3: Summary of three proposed routing strategies

Strategy	Metric	Location or Encounter	Blind Spot & Dead End	Remarks
Loc	Geographic distance + Mobility pattern	Location	No	Achieves higher performance than single location-based utility metric
Soc	Social centrality + Social similarity	Encounter	Partially	Achieves better performance than single encounter-based utility metric and while less sensitive and easier for collection than location information
SMART	Social information + Community	Encounter	Significantly	Resolves blind spot and dead end problems by efficient intra and inter community communication strategy

We use the following metrics to evaluate the performance of various data routing strategies [33, 13].

- **Delivery ratio:** the ratio of the number of destinations that have received the delivered data to the total number of destinations.

- **Average delay:** the average time delay used for destinations to receive the data.
- **Average cost:** the average number of relays used for the messages successfully delivered to destinations.

A good routing strategy is supposed to have high delivery ratio, low average delay and low average cost. The objective of our devised strategies is to enhance data delivery ratio, reduce average delay and meanwhile maintain low cost.

Chapter 4

Location-based Routing Strategy

Location-based social information provides accurate position of nodes in the network. If such data is used properly, it can be applied to enhance data routing performance in DTN. In this chapter, we propose a comprehensive routing metric that combines several aspects of location-based features. The previous literatures consider only one aspect of node characteristics or network features for constructing routing metric, which cannot fully describe the network situation along the entire period of routing process. For example, routing strategies based on geographic distance cannot make accurate forwarding decision in the case that the change of the network topology is frequent and unpredicted, or it cannot forward data farther when nodes are with similar geographic distance with destinations. Similarly, routing strategies relying on mobility pattern also need to consider the situation that the data forwarding choice when nodes with similar mobility pattern. For instance, two nodes often stay in the same location do not suggest that they meet each other frequently. The sole mobility pattern metric cannot represent the routing criteria in such situation.

To overcome above-mentioned problems in location-based routing, the proposed routing scheme combines both geographic distance and mobility pattern to comprehensively select data relay for further data delivery. The detailed design of the routing scheme is represented in the following sections.

4.1 Location-based Social Information

We consider a network consisting of multiple mobile nodes that may travel in different locations. Each node is equipped with GPS device for awareness of geographical coordinates.

4.1.1 Location-based graph

We have presented the basic social graph model in Chapter 3. The nodes in the graph represent the mobile nodes in DTNs. The link between two nodes suggests

the social relationship of them and the weight on the link indicates the strength of the relationship. The involving of location information enriches the graph in spatial aspect. Specifically, the network area is with geographic coordinates and we convert it to a grid with squares. Each square is considered as a location and is assigned with a unique location ID. The network area is represented by $L = l_1, l_2, \dots, l_n$ where l_i means location i .

To describe location information and mobility of nodes, we consider their movements as discrete time-varying events. Each event suggests the location of a node with time label. In particular, an event is described by four elements: node ID, location ID, start time and time duration. The start time is the time when the node enters the location and the time duration is the time length that the node stays at the location. Based on the event sequences, we can obtain the statistical location information such as the similarity of mobility pattern which is described as time proportion that a node stays at a location, and temporal location information like the distance between two nodes at a certain time. The statistical location information suggests the user mobility pattern. For example, if a user stays at a location for a large proportion of time, he will likely to visit the same location in the future. The temporal location information, on the other hand, represents the instant user behavior at a certain time. We describe two types of location information in the following.

4.1.2 Geographical distance

The *geographical distance* between two nodes measures their physical separation. Several schemes [87, 45] take the distance as utility for data forwarding based on the fact that the closer geographical distance two nodes have, the more likely they will meet with each other. We calculate the distance of a pair of locations (x, y) where n_i and n_j have visited as:

$$g_{xy} = ||s_x - s_y||,$$

where s_x is the GPS coordinates of location x , and g_{xy} is the distance between x and y . In line with the characteristics of utility value that is the larger the better, we make a conversion as:

$$d_{xy} = \frac{1}{1 + g_{xy}},$$

The larger g_{xy} , the smaller d_{xy} is. Each node visits several locations. To measure the distance between two nodes, we use a distance matrix to represent it. Suppose node n_i has visited a series of locations $L_i = \{l_{i1}, l_{i2}, \dots, l_{im}\}$, and node n_j has visited $L_j = \{l_{j1}, l_{j2}, \dots, l_{jn}\}$. The distance matrix between i and j therefore is written as:

$$D_{ij} = \begin{bmatrix} d_{l_{i1}, l_{j1}} & \cdots & d_{l_{i1}, l_{jn}} \\ \vdots & \ddots & \vdots \\ d_{l_{im}, l_{j1}} & \cdots & d_{l_{im}, l_{jn}} \end{bmatrix}.$$

4.1.3 Similarity of mobility pattern

The *similarity of mobility pattern* measures the extent that different nodes visiting the same places. It is calculated by the time proportion that two nodes spend in the same locations [49]. The larger time proportion that nodes stay at the same places, the more similar their mobility patterns are. We show the measurement of similarity of mobility pattern as follows.

For each node n_i , it spends different proportions of time at various locations over a defined time interval. Similar to [49], we use a vector m_i to present its time proportions of different locations:

$$m_i = (c_{1i}, c_{2i}, \dots, c_{ni}), \text{ with } \sum_{k=1}^n c_{ki} = 1,$$

where c_{ki} is the time proportion that node n_i at location k . The product of time proportion between node n_i and n_j in a pair of locations (x, y) is:

$$s_{x_i, y_j} = c_{x_i} * c_{y_j}.$$

It suggests that the probability that node n_i stays at location x and node n_j stays at location y . Suppose node n_i has visited a series of locations $L_i = \{l_{i1}, l_{i2}, \dots, l_{im}\}$, and node n_j has visited $L_j = \{l_{j1}, l_{j2}, \dots, l_{jn}\}$, we use a matrix to denote the product of time proportion between n_i and n_j in different locations.

$$S_{ij} = \begin{bmatrix} s_{l_{i1}, l_{j1}} & \cdots & s_{l_{i1}, l_{jn}} \\ \vdots & \ddots & \vdots \\ s_{l_{im}, l_{j1}} & \cdots & s_{l_{im}, l_{jn}} \end{bmatrix}.$$

The matrix depicts the probability that node n_i and node n_j stay at various locations.

It reveals the similarity of their mobility patterns.

4.2 Strategy Design

Location-based data forwarding strategies are usually determined by the location information among different nodes. In this section, we propose a comprehensive location-based social routing strategy called *Loc* in delay tolerant networks.

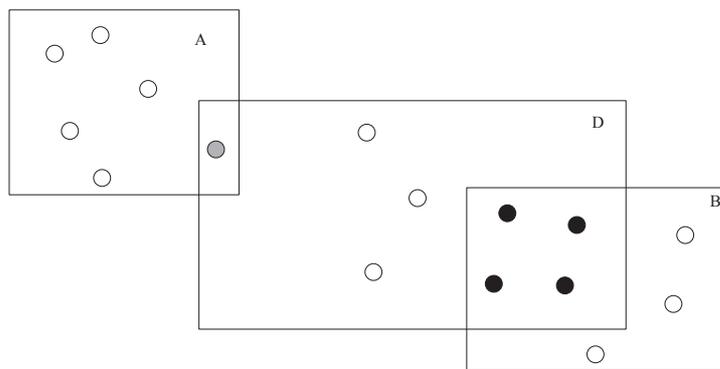


Figure 4.1: An example of visiting locations

The design of *Loc* is inspired by the observation that two nodes physically stay close to each other and commonly visit same locations are more likely to meet each other. Such situation occurs between colleagues, neighbors and etc. That is, two nodes with similar mobility pattern and close in geographical locations (i.e. they share many common visited places and their visited places are close) are more likely to meet each other. For example, as shown in Fig. 4.1, suppose node *A* and *B* are possible relays to deliver data to node *D*. Three squares represent the visiting areas of three nodes respectively. The circles inside of squares are several locations they visit. Overall, the average distance between visiting locations of node *A* (circles in the left square) and visiting locations of node *D* (circles in the middle square) is longer than the distance between node *B*'s visiting locations (circles in the right square) and node *D*'s visiting locations. Besides, node *A* shares one location with node *D* as shown in grey circle, whereas node *B* shares 4 locations with node *D* as shown in black circles. Based on the similarity of mobility pattern and geographical distance, node *B* will most likely be selected for data delivery to node *D*. According to this observation, we model *Loc* scheme by incorporating the similarity of mobility pattern and geographical distance to the destination node.

To measure the similarity of mobility pattern and the geographical distance,

4.2. Strategy Design

Table 4.1: Table of visiting history

<i>Location</i>	<i>Visiting time</i>	<i>Duration</i>	<i>Time propagation</i>
x	t	Δt	c_x

each node needs to maintain a list of locations they have visited as shown in Table 4.1, which contains the time of visiting the location as well as the time proportion of staying at it. When two nodes encounter, they exchange the list of locations and calculate corresponding metrics. From the point view of location information, on the one hand, the similarity of mobility pattern only measures the extent that different nodes staying at the common places. It cannot reflect the spatial distance of nodes. The geographical distance, on the other hand, only shows the temporal value of distance, which cannot present the geographical closeness of two nodes. Therefore, we propose Loc scheme by combining similarity of mobility pattern and geographical distance. To present this compound measurement, we take the Hadamard product of the two matrices. The Hadamard product produces a matrix H_{ij} that each element pq is the product of pq elements in S_{ij} and D_{ij} . The operation is depicted in the following:

$$H_{ij} = S_{ij} \circ D_{ij} = \begin{bmatrix} s_{l_{i1},l_{j1}} d_{l_{i1},l_{j1}} & \cdots & s_{l_{i1},l_{jn}} d_{l_{i1},l_{jn}} \\ \vdots & \ddots & \vdots \\ s_{l_{im},l_{j1}} d_{l_{im},l_{j1}} & \cdots & s_{l_{im},l_{jn}} d_{l_{im},l_{jn}} \end{bmatrix}.$$

The matrix presents both similarity of mobility pattern and the geographical distance of two nodes. We take the average of the sum of all elements in the matrix as the geographical metric between node n_i and n_j . The larger of the average, the closer geographical relation two nodes have, and therefore the more chance they will encounter. When two nodes (i.e. n_i and n_j) encounter, for messages carried by n_i (n_j), they decide whether to take n_j (n_i) as the next relay by comparing their utilities to the destination, which is the average of the sum of elements in H_{id} (H_{jd}). If the average value of n_i (n_j) is smaller than that of n_j (n_i), the message will be forwarded from n_i (n_j) to n_j (n_i). The detail of Loc algorithm is illustrated in Algorithm 1.

Algorithm 1: Loc algorithm

input : visiting location information and encounter events in DTNs

output: utility and data routing decision

begin

 assume the initial time is t_0 ;

if n_i encounter with n_j at time t **then**

foreach data x in node i **do**

 update S_{id_x} ;

 update D_{id_x} ;

$H_{id_x} = S_{id_x} \circ D_{id_x}$;

 update S_{jd_x} ;

 update D_{jd_x} ;

$H_{jd_x} = S_{jd_x} \circ D_{jd_x}$;

if $avg(H_{id_x}(t)) < avg(H_{jd_x}(t))$ **then**

 | forward data to n_j ;

foreach data y in node j **do**

 update S_{id_y} ;

 update D_{id_y} ;

$H_{id_y} = S_{id_y} \circ D_{id_y}$;

 update S_{jd_y} ;

 update D_{jd_y} ;

$H_{jd_y} = S_{jd_y} \circ D_{jd_y}$;

if $avg(H_{jd_y}(t)) < avg(H_{id_y}(t))$ **then**

 | forward data to n_i ;

4.3 Performance Evaluation

We conduct experiments to study the performance of location-based social routing strategies.

4.3.1 Experiment setup

We use HuggleSim simulator [39] to launch our experiments, which uses encounter entries as inputs to estimate data delivery path according to different data routing strategies. We extract 14-day session from three real traces and synthetic traces. We utilize the GPS location information as the basic location-based social information as the input to construct corresponding location-based social graph and related compound location information. The simulator generates 1,000 messages for each round of simulations. Each message is assigned with a random source and destination. For each message, we keep three copies in the network to make the data delivery with higher chance reach destinations. The message keeps alive until the experimental session (14 days) is end. We do each simulation 20 times and take the average value of results for statistical convergence.

To measure the location information of nodes in the network, we divide the experiment area of data traces into discrete locations. The network area is marked with geographic coordinates. According to the coordinates, we calculate the size of network area and convert it to a grid with multiple blocks. Each block in the grid is defined as a location. It is distinguished by a unique ID. Specifically, we divide the MIT Reality experiment area by cellular towers. Since MIT Reality data trace does not contain the GPS coordinates of node mobility and a location is marked by detected cellular tower ID, we consider the area of a cellular tower is a location. In contrast, we convert the experiment area of the other two data sets to a grid with adjacent squares. The size of each square is 1 km^2 and each of them is assigned with a unique ID. If a node's mobility range falls into a location, we record the visiting history of the node. We measure the distance of two locations by the coordinates of their centers.

4.3.2 Strategies in comparison

We compare the Loc scheme with other two location-based data routing strategies: Mobyspace [49], and GPSR [45].

Mobyspace calculates the similarity of mobility pattern by Euclidean distance

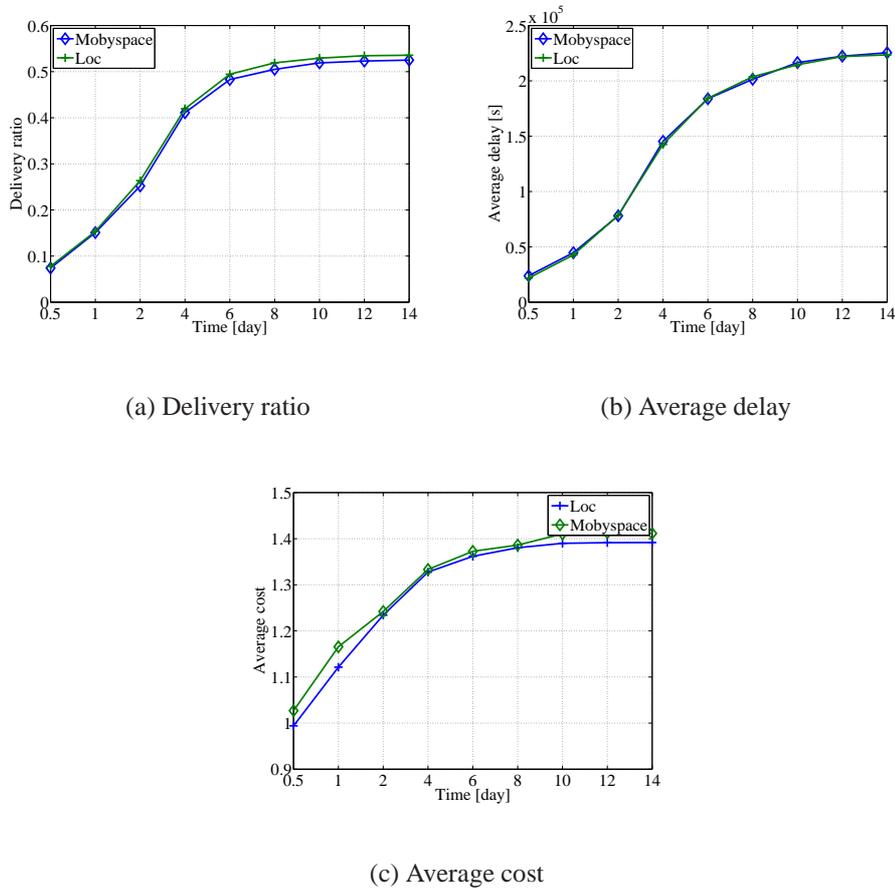
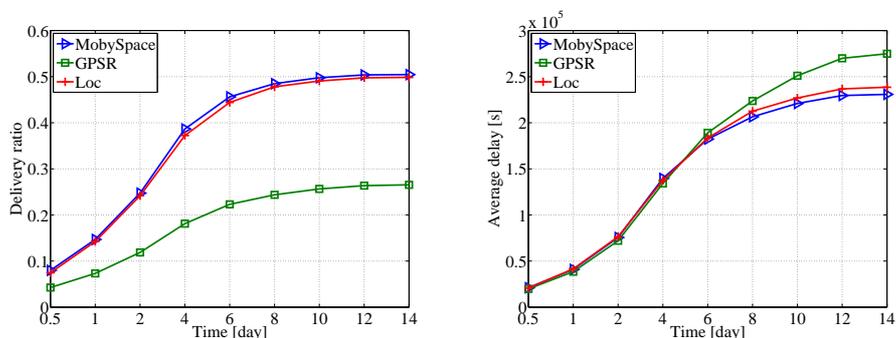


Figure 4.2: Performance comparison of location-based strategies on MIT Reality data trace

of two nodes' visiting history and chooses the node with shorter distance (i.e., more similar mobility pattern) with the destination node as the relay for data delivery. Specifically, it considers a node's visiting history as a vector. Each value in the vector represents the percentage of time that the node stays at the location. The distance between two nodes is computed by the Euclidean distance between two vectors.

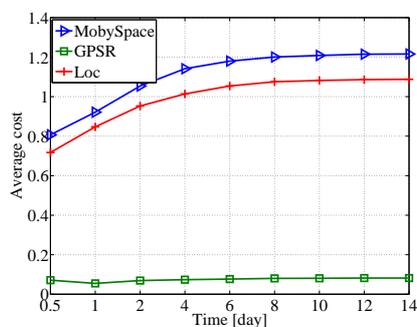
GPSR routes data based on geographical distance. It makes greedy forwarding using the geographic positions of a router's neighbors in the network topology. Specifically, a node obtains its neighbors' positions by information exchange. Then it locally chooses the optimal next hop with the neighbor geographically closest to the packet's destination. Forwarding in this regime follows successively closer geographic hops until the destination is reached. When a packet reaches a region where greedy forwarding is impossible (i.e. packets reaches the local maxima), the

4.3. Performance Evaluation



(a) Delivery ratio

(b) Average delay



(c) Average cost

Figure 4.3: Performance comparison of location-based strategies on DieselNet data trace

algorithm recovers by routing around the perimeter of the region. It exploits the right-hand rule which traverses the interior of a closed polygonal region in clockwise edge order to seek for the next hop.

4.3.3 Performance analysis

To implement the Loc scheme, we set the period for GPS coordinates refreshment as one day. We use the same settings as the original paper for the implementation of MobySpace and GPSR.

Fig. 4.2 shows the results of delivery ratio, average delay and cost of MobySpace and Loc on MIT Reality data trace. Since MIT Reality trace does not provide coordinates information, the GPSR scheme cannot be evaluated and the utility of Loc can only reflect the similarity of mobility pattern on the data trace. The presentation of Loc expresses the same meaning as MobySpace, though their calculation is

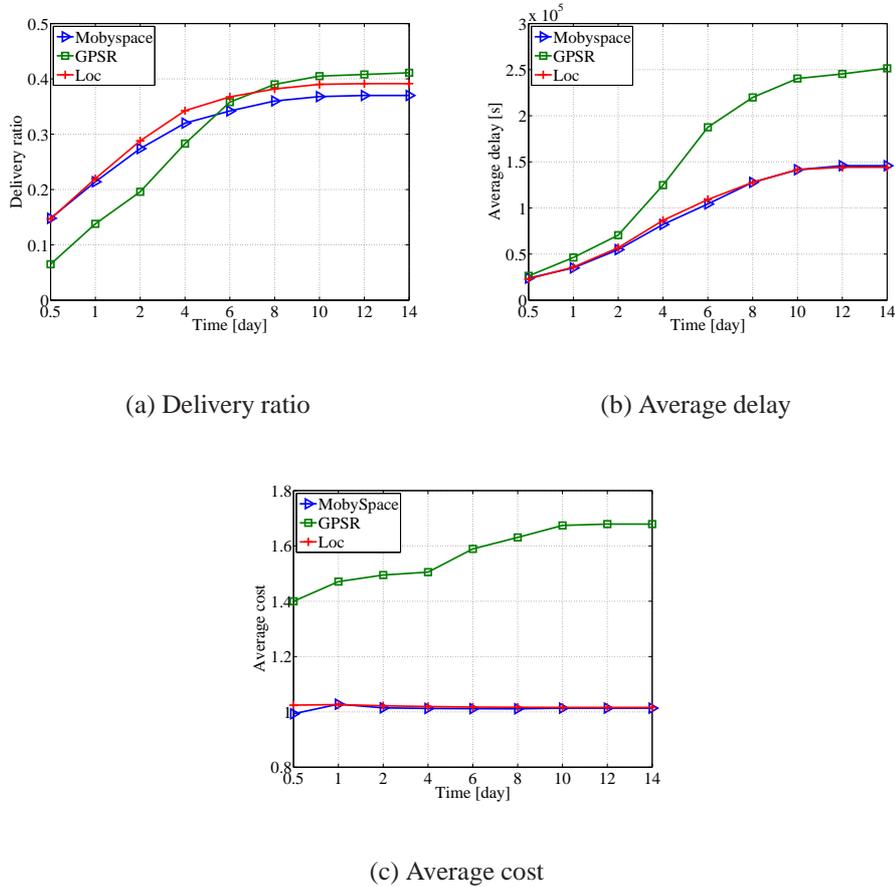


Figure 4.4: Performance comparison of location-based strategies on Cabspotting data trace

slightly different. Therefore, they get very similar performance in terms of various aspects.

The performance on DieselNet data trace is shown in Fig. 4.3. Fig. 4.3a shows that the Loc scheme has similar delivery ratio with Mobyspace on DieselNet data trace throughout the experiment period. Nevertheless, GPSR has 20% degradation compared with the Loc scheme in terms of delivery ratio since the 4th day of the experiment. This degradation attributes to the distance latency of GPSR algorithm, which does not show the real time distance between current node and the destination, thus makes the relay selection inefficient. Similar results as shown in Fig. 4.3b, the average delays of three schemes are similar before 6 days. Afterwards, the average delay of the Loc scheme and Mobyspace is around 80% of that of GPSR. The results of average cost as shown in Fig. 4.3c depict that the Loc scheme has 0.1 hops lower cost than Mobyspace. The cost of GPSR is the lowest, which only

4.3. Performance Evaluation

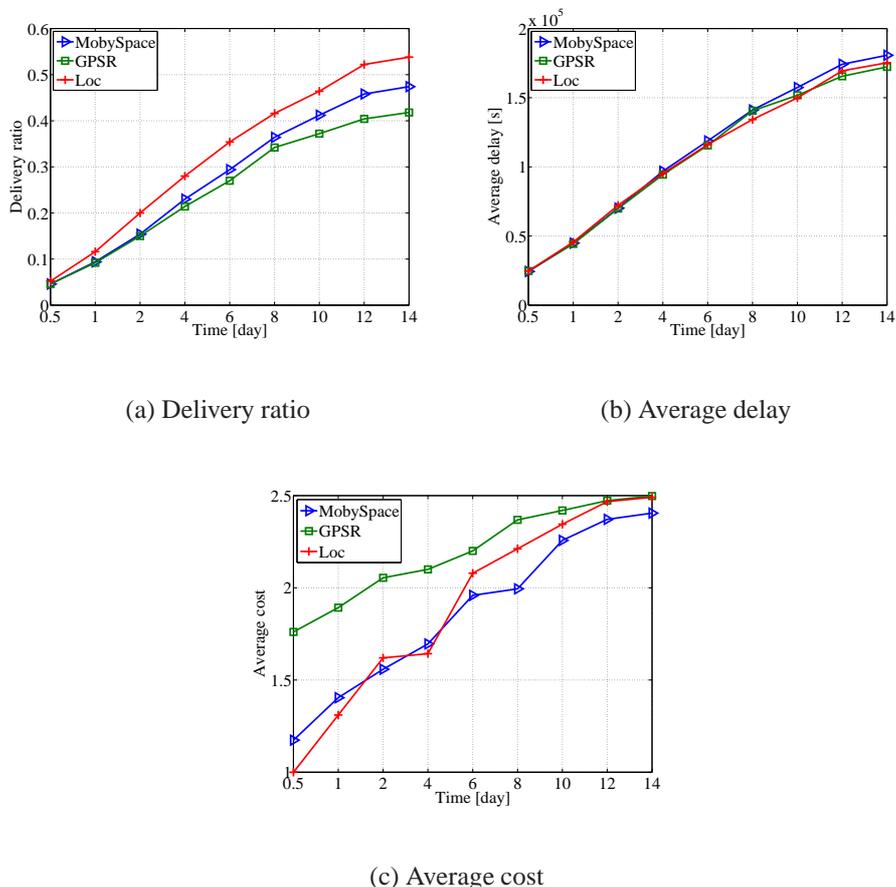


Figure 4.5: Performance comparison of location-based strategies on synthetic data trace

takes 0.1 hops of average cost.

The performance on Cabspotting data trace is depicted as Fig. 4.4. The result of delivery ratio is shown in Fig. 4.4a. It shows that Loc outperforms MobySpace by 2% throughout the experiment. The delivery ratio of GPSR has 10% lower delivery ratio in the first half experiment period and then 5% higher delivery ratio than that of the Loc scheme. Fig. 4.4b presents the average delay of MobySpace and Loc is 60% of that of GPSR from 6th day of the experiment. Similar results as shown in Fig. 4.4c, the average cost of the Loc scheme is very similar with MobySpace, and it is around 0.6 hops lower than that of GPSR throughout the experiment.

In a summary, the experiment results show that Loc outperforms the other location-based routing strategies in most cases on three real world data traces.

Similarly, we conduct the comparison of location-based routing strategies on synthetic traces and show the comparison results on one data trace ($ns = 1.5m/s$,

$nn = 100$). The evaluation results are shown in Fig. 4.5. The delivery ratio of Loc is 8% higher than that of Mobyspace and 15% higher than that of GPSR during the entire experiment period, as depicted in Fig. 4.5a. Three location-based routing strategies have similar delay for data delivery as shown in Fig. 4.5b. Additionally, as shown in Fig. 4.5c the Loc scheme takes 0.1 hops more than Mobyspace at the end of the experiment. It costs 1 hop less than GPS in the beginning of the experiment, and reaches similar cost in the end. The Loc scheme performs not worse than other location-based routing strategies. The performance on synthetic traces confirms our results on the real traces.

4.4 Summary of Contributions

In this chapter, we propose a location-based routing strategy incorporating geographic distance and mobility pattern as two aspects of location information to conduct routing utility. The comprehensive metric is adaptable for more network situations compared with the metric relying on sole aspect of location information. We conduct the performance evaluation on both real and synthetic data traces. The results show that the proposed comprehensive location-based social routing strategy outperforms other location-based strategy around 10% in terms of delivery ratio. It takes less decay and cost for data delivery.

Chapter 5

Encounter-based Routing Strategy

We discuss the location-based routing strategy in the last chapter. However, the utilization of location-based social information is very sensitive and needs more privacy concern [10, 11, 72, 98, 47]. In contrast, encounter-based social information refers to the fundamental element of DTNs. That is, the data delivery in DTNs relies on the encountering event of nodes. Applying such encounter-based information reveals much less sensitive than location information.

Compared with location-based routing, encounter-based routing has different characteristics. Location-based strategies forward data to the nodes geographically closer to the destination, which tend to achieve geographical shortest routing path. In contrast, encounter-based social strategies forward data to the nodes logically closer to the destination. It looks for the logical shortest routing path.

Inspired by the motivation of privacy preserving as well as the characteristics of encounter-based routing strategies, we propose a comprehensive encounter-based social routing scheme.

5.1 Encounter-based social information

We consider two nodes have an encounter if they are in the communication range of each other. The encounter-based social graph is modeled by the following method.

5.1.1 Encounter-based social graph

As we described in Chapter 3, we model DTN 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 weights of links 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 encountering events, two nodes have a social link if the number of their encounters exceeds a threshold. The weight of the edges is the number of encounters between two nodes.

Social information in an encounter-based social graph refers as the information that either directly obtained from encounter-based social graph, or social analysis results of contact graph. It indicates the structural status of a node in the constructed weighted graph. Typical encounter-based social information includes: social degree of a node, which is known as the degree of a node in encounter-based social graph, representing the number of friends (encountering nodes), social strength between two nodes, denoted by the weight of an edge between the pair of nodes in encounter-based social graph, and social similarity, which represented by the common friends of two nodes in encounter-based social graph and etc.

Specifically, we address two widely used social properties inferred from encounter-based social graph: social similarity and social centrality.

5.1.2 Social similarity

Social similarity evaluates the number of common friends of two nodes, which indicates the trustiness and cohesive of social links [23, 22]. We define the social similarity as follows:

$$\mathcal{S}_{i,j}(\tau) = 1 + |F_i(\tau) \cap F_j(\tau)|, \quad (5.1)$$

where $F_i(\tau)$ ($F_j(\tau)$) is the set of friends of node n_i (n_j) at time τ and plus 1 is to avoid 0 values. Thus, the social similarity between n_i and the destination n_d is $\mathcal{S}_{i,d}(\tau)$. Intuitively, if a node has higher $\mathcal{S}_{i,d}(\tau)$ value, it shares more common friends with the destination, thus more likely to transmit the message successfully.

5.1.3 Social centrality

Social centrality is the quantification of the relative importance of nodes in the social network. There are various definitions of centrality, such as edge betweenness [31, 30] or closeness centrality [30], which cannot very easily be exploited in DTNs since they need global information to estimate the centrality value. Therefore, we use the Freeman's degree centrality [30], which only needs the neighbor information, to define social centrality in the context of DTN. For a node n_i , its centrality is defined as follows:

$$\mathcal{C}_i(\tau) = \frac{\sum_{k=1}^N d_{ik}(\tau)}{N},$$

where $d_{ik}(\tau) = 1$ if a direct link exists between n_i and n_k at time τ and N is the number of nodes in the network. Structurally, a central node has strong connection with other nodes, and it is suitable to serve as a hub for information exchange. Previous research [80] shows that there is a strong statistical dependency between delivery ratio and social centrality.

5.2 Strategy Design

Encounter-based data routing schemes handle data based on either direct encounter information or various social information derived from encounter-based social graph, such as social degree and the strength of social ties. We propose a social-based representative data routing scheme in this section, which is named as *Soc*.

Previous research has shown that the status of nodes in a social network is uneven: some nodes are in the central positions of the network while the others are in the edges [92]. An example is that a small fraction of nodes occupy most of degrees in the social graph structure. Generally speaking, forwarding data to the node who is more social active will increase the probability of data delivery. Based on this consideration, we propose the Soc scheme combining social centrality and social similarity to find the most feasible social path for data forwarding.

Table 5.1: Table of encounter history

<i>Encounters</i>	<i>Time</i>	<i>Similarity</i>	<i>Centrality</i>
n_j	t_j	$\mathcal{S}_{i,j}(t_j)$	$\mathcal{C}_i(t_j)$

To determine the value of social centrality and social similarity, for a node n_i , it records the encounter history for a period of ΔT as shown in Table 5.1, which contains the encounter time and its social properties accordingly. Meanwhile, it also maintains a list of friends. When two nodes encounter, they exchange their friend lists to compute the social similarity. Noticing that social similarity and social centrality only reflect the features of network structure, we also need to consider the dynamics of social networks. Since the encounters of mobile nodes change dynamically, the comprehensive utility should be a time-varying function. To address the dynamic feature and avoid the accumulative effects, we define the comprehensive utility as the convolution of social similarity and social centrality with a factor

Algorithm 2: Soc algorithm

input : encounter events in DTNs and encounter-based social graph

output: utility and data routing decision

begin

 assume the initial time is t_0 ;

if n_i encounter with n_j at time t **then**

foreach data x in n_i **do**

$Y_{i,d_x}(t) = C_i(t_0)$;

$Y_{j,d_x}(t) = C_j(t_0)$;

foreach encounter between n_i and n_{d_x} at t_{i,d_x} ($t_{i,d_x} < t$) **do**

$Y_{i,d_x}(t) = Y_{i,d_x}(t) + \mathcal{S}_{i,d_x}(t_{i,d_x}) * C_i(t_{i,d_x}) / (t - t_{i,d_x})$;

foreach encounter between n_j and n_{d_x} at t_{j,d_x} ($t_{j,d_x} < t$) **do**

$Y_{j,d_x}(t) = Y_{j,d_x}(t) + \mathcal{S}_{j,d_x}(t_{j,d_x}) * C_j(t_{j,d_x}) / (t - t_{j,d_x})$;

if $Y_{i,d_x}(t) < Y_{j,d_x}(t)$ **then**

 forward data to n_j ;

foreach data y in n_j **do**

$Y_{j,d_y}(t) = C_j(t_0)$;

$Y_{i,d_y}(t) = C_i(t_0)$;

foreach encounter between n_j and n_{d_y} at t_{j,d_y} ($t_{j,d_y} < t$) **do**

$Y_{j,d_y}(t) = Y_{j,d_y}(t) + \mathcal{S}_{j,d_y}(t_{j,d_y}) * C_j(t_{j,d_y}) / (t - t_{j,d_y})$;

foreach encounter between n_i and n_{d_y} at t_{i,d_y} ($t_{i,d_y} < t$) **do**

$Y_{i,d_y}(t) = Y_{i,d_y}(t) + \mathcal{S}_{i,d_y}(t_{i,d_y}) * C_i(t_{i,d_y}) / (t - t_{i,d_y})$;

if $Y_{i,d_y}(t) > Y_{j,d_y}(t)$ **then**

 forward data to n_i ;

decaying as time.

$$\begin{aligned} Y_{i,d}(T) &= \mathcal{S}_{i,d}(T) \otimes (\mathcal{C}_i(\tau)/T). \\ &= \int_{\tau=0}^T \mathcal{S}_{i,d}(\tau) * \frac{\mathcal{C}_i(\tau)}{T - \tau}. \end{aligned}$$

The convolution operation provides a time-decaying description of all prior values of social similarity and social centrality. The utility is updated each time by accumulation of social similarity when a new encounter occurs. The decay function suggests that the most recent encounters typically have the more influence, and the impacts of previous encounters decrease as elapsed time and social centrality. When two nodes (i.e. node n_i and n_j) encounter at time t , for messages carried by node n_i and messages carried by node n_j , they are determined whether to be transmitted to n_j (n_i) by comparing the utilities of n_i and n_j to destinations. If $Y_{i,d}(t)$ ($Y_{j,d}(t)$) is smaller than $Y_{j,d}(t)$ ($Y_{i,d}(t)$), the message will be transmitted from n_i (n_j) to n_j (n_i). The algorithm of Soc is outlined in Algorithm 2.

5.3 Performance Evaluation

We conduct experiments to study the performance of the encounter-based social strategies.

5.3.1 Experiment setup

We use HuggleSim simulator [39] to launch our experiments, which uses encounter entries as inputs to estimate data delivery path according to different data routing strategies. We extract 14-day session from three data traces and synthetic data traces. We utilize the encountering events as the input to construct corresponding encounter-based social graph and related social information. The simulator generates 1,000 messages for each round of simulations. Each message is assigned with a random source and destination. For each message, we keep three copies in the network to make the data delivery with higher chance reaching destinations. The message keeps alive until the experimental session (14 days) is end. We do each simulation 20 times and take the average value of results for statistical convergence.

To show the social features of composed encounter-based social graph, we investigate the structural properties of the graph from a random selected periodical social graph. We first measure the cluster coefficient of three data traces. The clus-

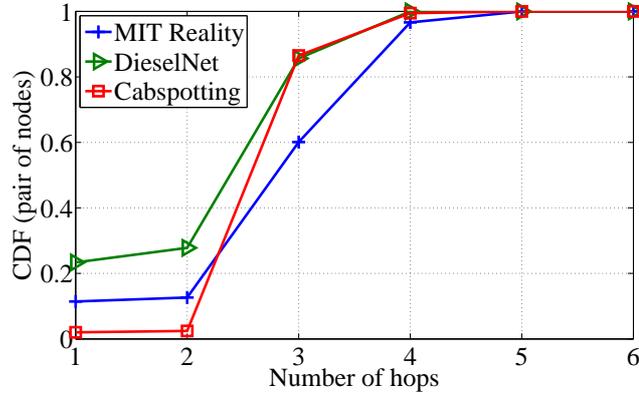


Figure 5.1: Cumulative distribution function of shortest path lengths

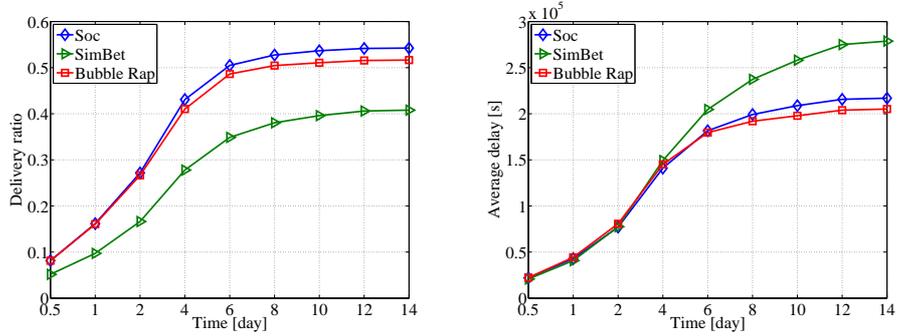
tering coefficient for MIT Reality is 0.5137, DieselNet has the coefficient value 0.7365 and the clustering coefficient of Cabspotting is 0.4544. We also measure the shortest path length of three composed encounter-based social graphs. By randomly sampling pairs of nodes in the social graph and calculating their average shortest path lengths, we draw the cumulative distribution function (CDF) in Fig. 5.1. As shown in the figure, over 95% of the shortest paths in the three data sets are below 5 hops. And almost 100% of the shortest paths are below 6 hops. Thus for a random pair of nodes in the DTNs, they are connected by a shortest path lower than 5 hops with high probability. The two measurements suggest that the encounter-based social graphs are all small world networks [91] and they are as clustered as social networks [92], which suggests the application of social-based properties, such as social centrality and social similarity are feasible for encounter-based social routing strategies.

5.3.2 Strategies in comparison

We compare Soc with two other encounter-based social routing strategies: Bubble Rap [40] and SimBet [24].

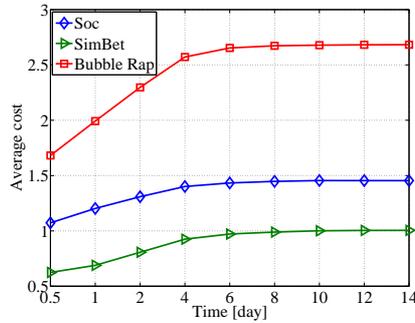
Bubble Rap considers the data routing in pocket switched network (PSN) which consists of several communities and there are social relationships among users. It uses k-clique percolation as the basic community detection method. There are two steps of routing in Bubble Rap. The first step is to forward data to the destination community. It delivers data items from outside of the destination's community according to a node's global social centrality. If a node with higher global social

5.3. Performance Evaluation



(a) Delivery ratio

(b) Average delay

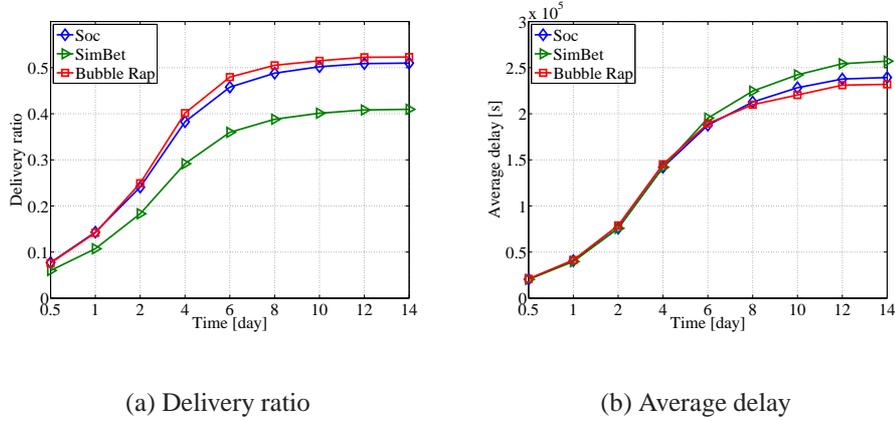


(c) Average cost

Figure 5.2: Performance of encounter-based social schemes on MIT Reality data trace

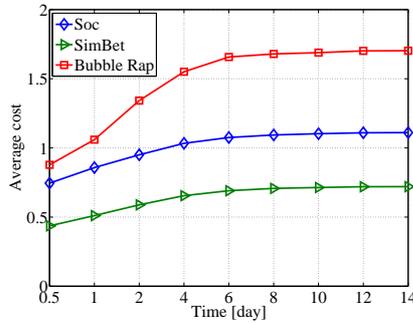
centrality, it will be selected as the relay for data forwarding. Within the destination's community, the forwarding utility is based on a node's local social centrality. The data item will be forwarded to a node with higher local social centrality.

SimBet 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. The scheme chooses the node with higher combination utility value as the relay for data forwarding.



(a) Delivery ratio

(b) Average delay



(c) Average cost

Figure 5.3: Performance of encounter-based social schemes on DieselNet data trace

5.3.3 Performance analysis

To implement the Loc scheme, we set the rate for refreshment of social links as once per day. We use the same settings as the original paper for the implementation of Bubble Rap and SimBet.

The performance of encounter-based social schemes on MIT Reality is shown in Fig. 5.2. It shows the dynamics of delivery ratio, average delay and average cost as a function of time. Due to the low contact rate in DTNs, the node may not reach the destination if the time that the data is sent out from the source is short. Thus, the delivery ratio shows the tight relation with time and increases as time passing. The average delay changes accordingly. Regarding the performance of delivery ratio on MIT Reality, as shown in Fig. 5.2a, due to the social characteristics in MIT Reality trace is apparent, the Soc schemes performs similar as Bubble Rap. Specifically, the Soc scheme outperforms Bubble Rap by 4%, but SimBet performs not as well as the others. The Soc scheme achieves more than 15% higher that SimBet in the

5.3. Performance Evaluation

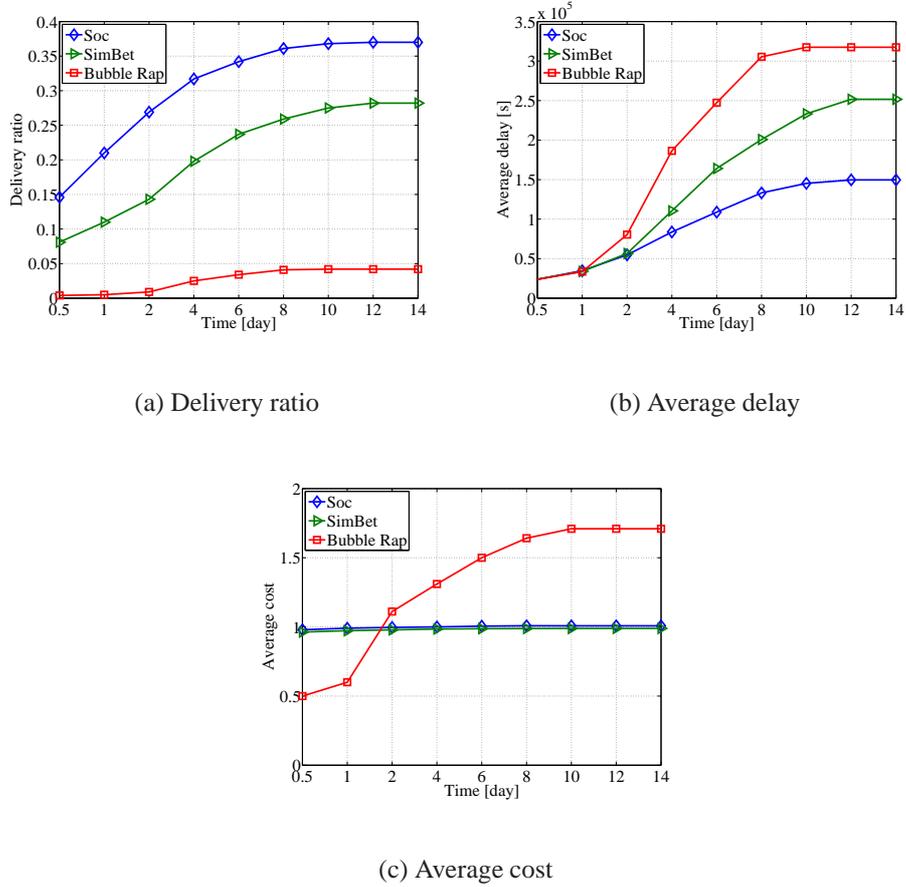
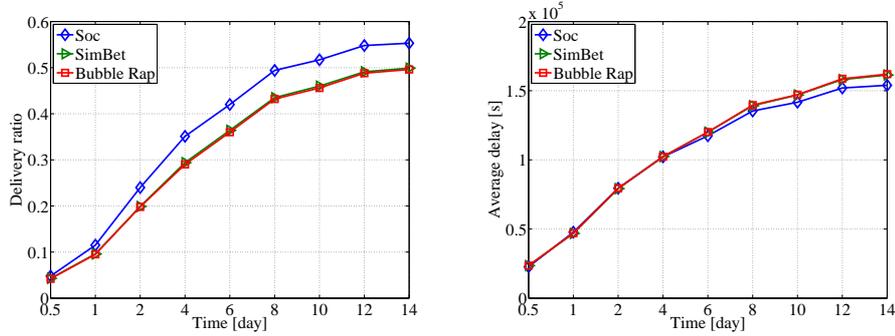


Figure 5.4: Performance of encounter-based social schemes on Cabspotting data trace

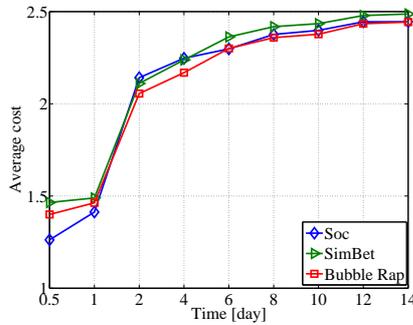
end of the experiment period. Similar results are shown in Fig. 5.2b in terms of average delay. The average delay of Soc is a little longer than that of Bubble Rap, but 40% shorter than that of SimBet in the end of the experiment. Accordingly, in the end of the experiment, Bubble Rap takes the longest average cost, which is 1.2 hops longer than that of Soc, while SimBet takes 0.5 hops shorter average cost than the Soc scheme.

Fig. 5.3 is the performance on DieselNet trace. Although DieselNet is a bus trace, it still shows strong social characteristics in its encounter-based social graph. As shown in Fig. 5.3a, Soc has slightly lower delivery ratio than Bubble Rap throughout the experiment, which is about 1%. In contrast, the delivery ratio of Soc is 10% higher than that of SimBet in the end of the experimental period. The comparison of average delay is shown in Fig. 5.3b, where the three schemes take similar delays before 6 days experiment time. Afterwards, the average delay of Soc



(a) Delivery ratio

(b) Average delay



(c) Average cost

Figure 5.5: Performance of encounter-based social schemes on synthetic data trace

is 2% longer than that of Bubble Rap, but 10% shorter than that of SimBet. Additionally, Bubble Rap takes the highest average cost while Soc has the lowest, as shown in Fig. 5.3c. The average cost of Soc is 0.8 hops lower than that of Bubble Rap and 0.3 hops higher than that of SimBet in the end of the experiment.

The performance of encounter-based social schemes on Cabspotting is shown in Fig 5.4. Fig. 5.4a shows that throughout the entire experiment, the Soc scheme outperforms SimBet by 10% in terms of delivery ratio, while Bubble Rap achieves very low delivery ratio. The poor performance of Bubble Rap suggests that the social centrality metric does not opt to Cabspotting data trace for data routing. Similar results as shown in Fig. 5.4b and Fig. 5.4c. The average delay of the Soc scheme is 50% of that of SimBet and 25% of that of Bubble Rap respectively in the end point of the experiment. The average cost of Soc is almost the same as the cost of SimBet throughout the experiment. It has 0.5 hops higher than Bubble Rap in the first two days of the experiment. The cost of Bubble Rap keeps growing and it has 0.7 hops

5.3. Performance Evaluation

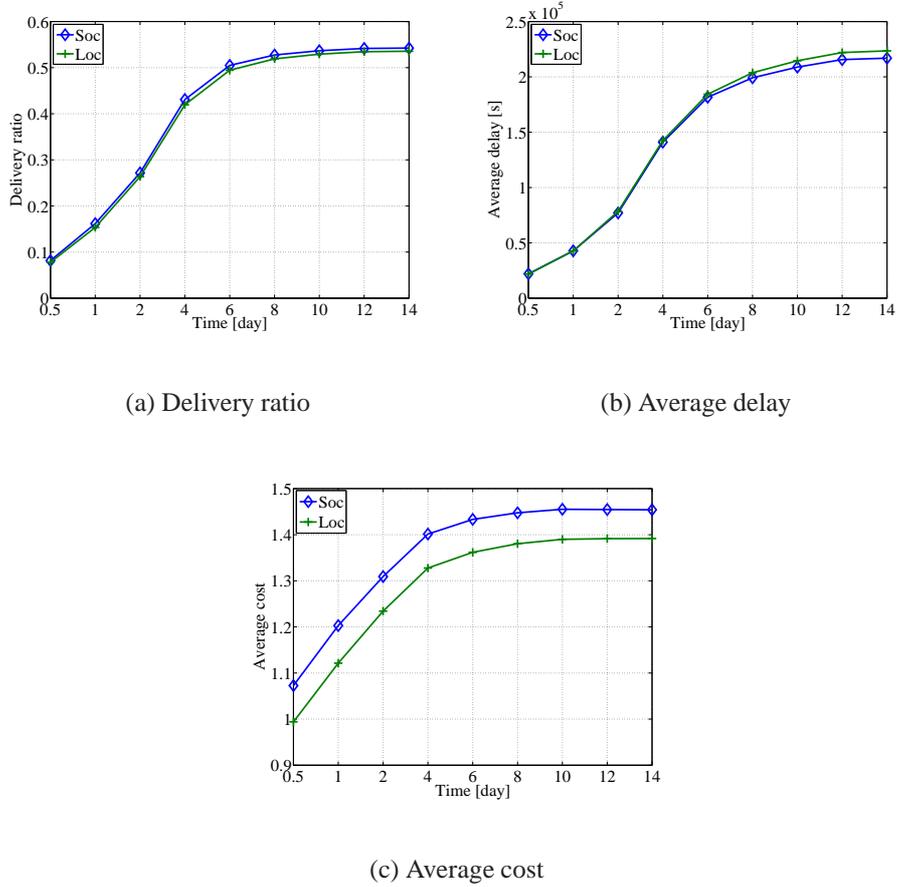


Figure 5.6: Performance comparison of Soc and Loc strategies on MIT Reality data trace

higher than others in the end of the experiment.

Overall, the experiment results show that the Soc scheme has competitive performance compared with other encounter-based social schemes on three real data traces.

To provide general assessing for encounter-based social routing schemes, we compare the performance of encounter-based strategies on synthetic data traces.

Here we show the comparison results on one instance of traces ($ns = 1.5m/s$, $nn = 100$), as presented in Fig. 5.5. Fig. 5.5a suggests that the Soc scheme has 5% higher delivery ratio than that of Bubble Rap and SimBet during the entire experiment period. Similarly, the Soc scheme has outstanding performance in terms of delay and average cost as shown in Fig. 5.5b and Fig. 5.5c. In particular, Soc takes 6% lower delay than both Bubble Rap and SimBet in the end of the experiment. At the same time, throughout the experiment, Soc uses slightly higher cost

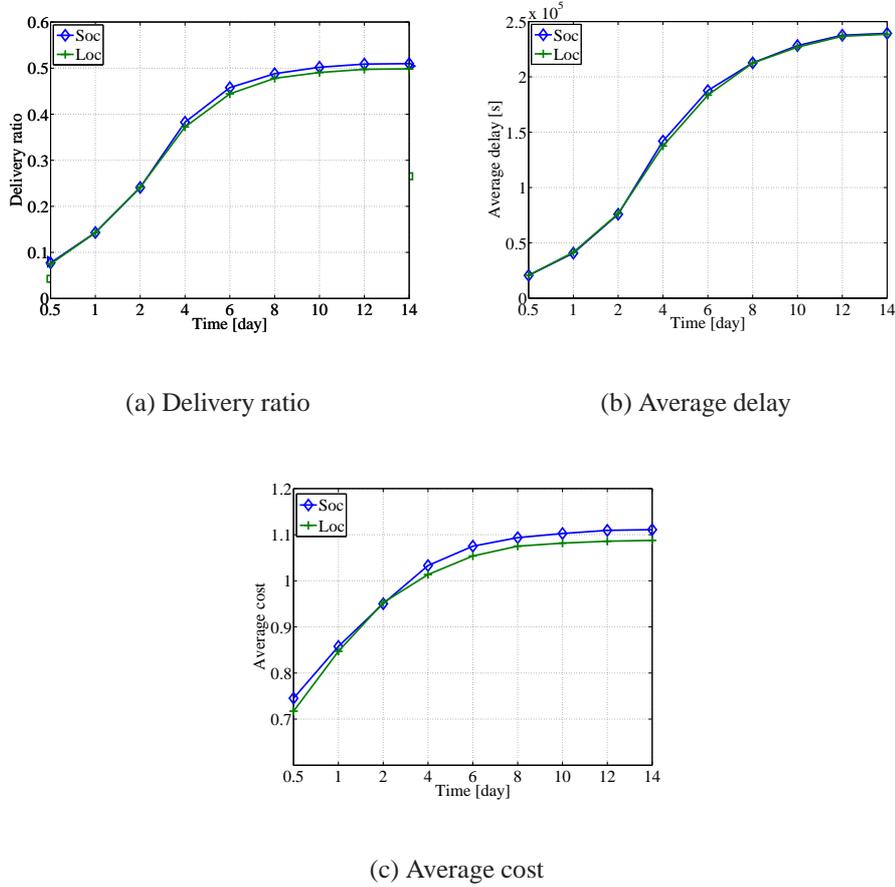


Figure 5.7: Performance comparison of Soc and Loc strategies on DieselNet data trace

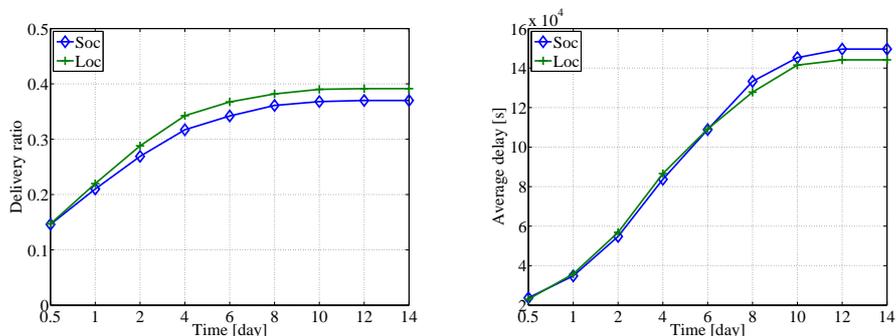
(less than 0.1 hops) than Bubble Rap. It uses 0.1 hops less cost than SimBet. The evaluation of encounter-based social routing strategies on synthetic traces is in line with the results on the real traces that the Soc scheme has competitive performance with other encounter-based social schemes.

5.3.4 Compare with the Loc strategy

In this subsection, we compare Soc to Loc in terms of various performance metrics on three real data traces and synthetic data trace.

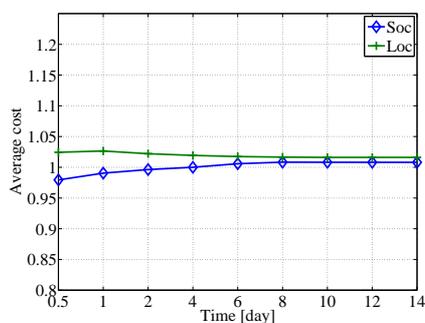
The performance comparison of Soc and Loc on MIT Reality is shown in Fig. 5.6. Fig. 5.6a presents the comparison of delivery ratio. It reveals that both Soc and Loc have very close achievement in terms of delivery ratio, albeit the delivery ratio of Soc is slightly higher (about 1%) than that of Loc throughout the experiment. Two schemes also perform similarly regarding average delay and average cost, as

5.3. Performance Evaluation



(a) Delivery ratio

(b) Average delay



(c) Average cost

Figure 5.8: Performance comparison of Soc and Loc strategies on Cabspotting data trace

shown in Fig. 5.6b and Fig. 5.6c, respectively. Loc takes a little higher average delay than Soc. Loc take 0.05 hops less cost than Soc.

The delivery ratio on DieselNet as shown in Fig. 5.7a presents that the Loc scheme has 1% degradation compared with the Soc scheme during the entire experiment period. Meanwhile, the Soc scheme outperforms the Loc scheme slightly in terms of average cost as shown in Fig. 5.7c. They have very similar average delay as shown in Fig. 5.7b.

The performance comparison on Cabspotting is shown in Fig. 5.8. The Loc scheme outperforms the Soc scheme by 2% in terms of delivery ratio as shown in Fig. 5.8a throughout the experiment. At the same time, the Soc scheme takes 2% higher average delay than that of the Loc scheme as shown in Fig. 5.8b. Regarding the average cost, as shown in Fig. 5.8c, the Soc scheme costs 0.1 hops less than the Loc scheme.

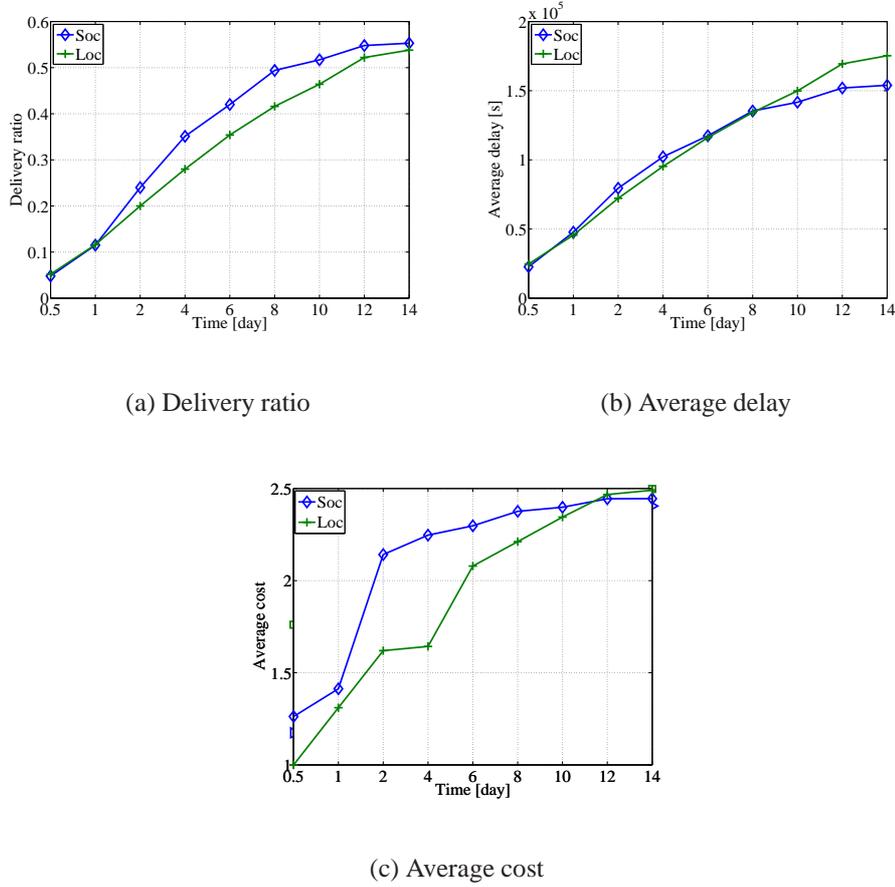


Figure 5.9: Performance comparison of Soc and Loc strategies on synthetic data trace

One instance of comparison of Soc and Loc on the synthetic traces ($ns = 1.5m/s$, $nn = 100$) is shown in Fig. 5.9. Similar as other data traces in the real world, both Soc and Loc have closely performance. Specifically, Soc outperforms Loc by about 3% in terms of delivery ratio as shown in Fig. 5.9a in the entire experiment period. The average delay of Loc and Soc as presented in Fig. 5.9b is twisted together at the beginning of the experiment and then the Soc scheme has lower delay than Loc. The average cost as shown in Fig. 5.9c presents that Soc has 0.2 hops higher cost than Loc in the beginning of the experiment and the cost of Soc is 0.1 hops lower than that of Loc in the end of the experiment.

Besides, we conduct a group of experiments on the synthetic data traces with different node speeds ranging from $0.5m/s$ to $2.5m/s$ ($nn = 100$). We record the experimental results in the end of each experiment as shown in Fig. 5.10. The delivery ratio of two schemes as a function of node speed on synthetic data traces

5.3. Performance Evaluation

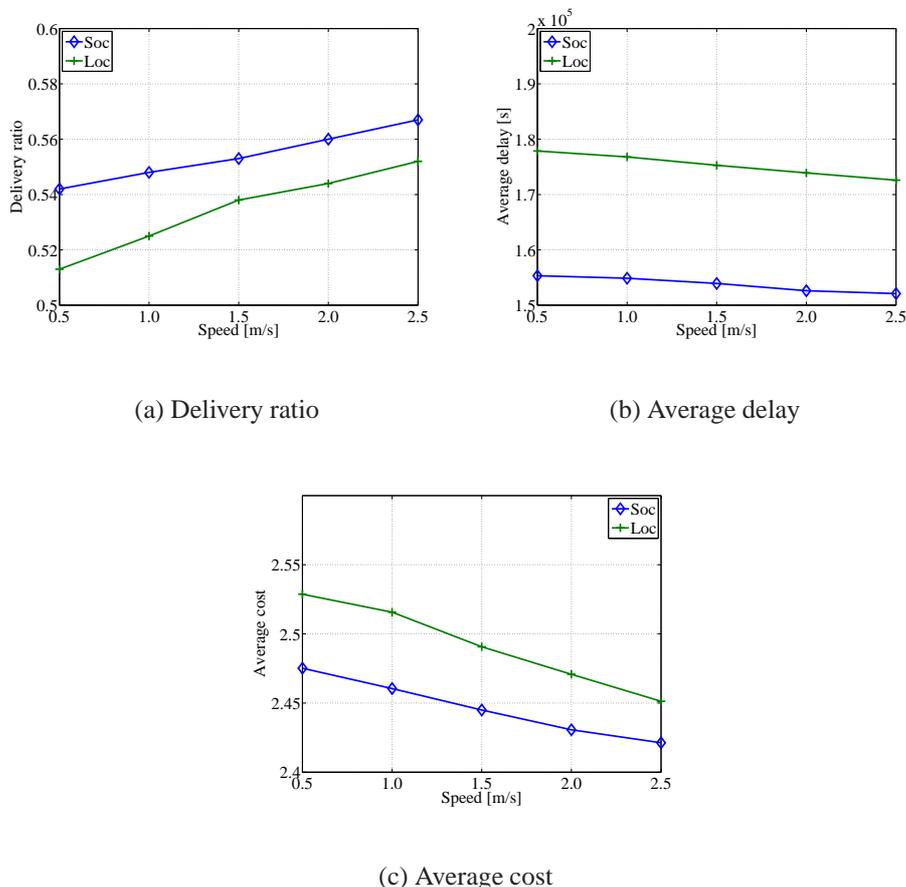


Figure 5.10: Performance comparison of Soc and Loc strategies on synthetic data traces as a function of node speed

is shown in Fig. 5.10a. The delivery ratio has slightly increment as the increase of speed. This attributes to the fact that faster speed increases the chance of encounters among nodes in the network. For each node speed, the Soc scheme outperforms Loc by about 3% in terms of delivery ratio. The average delay as shown in Fig. 5.10b suggests that the higher speed leads to the lower average delay. The Soc scheme takes 15% less delay than the Loc scheme in all speeds. Similar to average delay, the average cost of Soc and Loc is presented in Fig. 5.10c. The cost of both schemes decreases as the increase of the speed. Besides, in each speed, Soc has about 0.1 hops less cost than Loc.

Additionally, we compare Soc with Loc in terms of different network size on the synthetic data traces. The results are recorded in the end of each experiment as shown in Fig. 5.11. The number of nodes in five different synthetic traces is 20, 40, 60, 80 and 100 respectively ($n_s = 1.5m/s$). The delivery ratio increases as

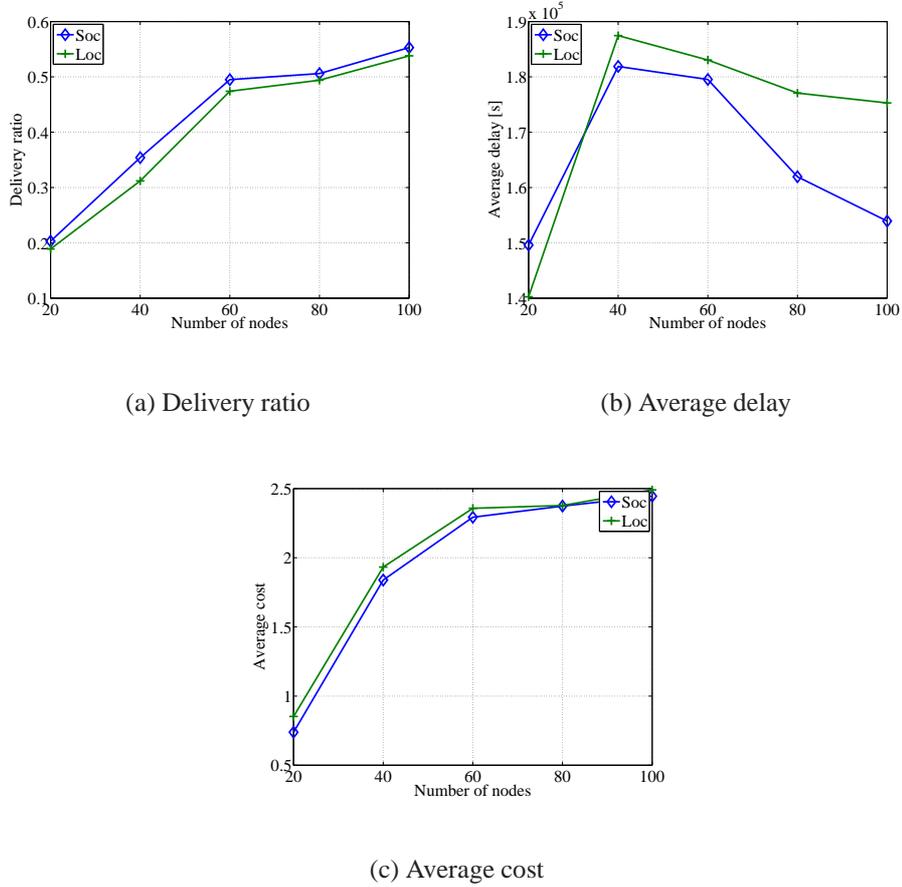


Figure 5.11: Performance comparison of Soc and Loc on synthetic data traces as a function of network size

the increase of the number of nodes in the network. Accordingly, the general trend of average delay and average cost declines as the number of nodes increases. The delivery ratio of two schemes is shown in Fig. 5.11a. In line with our previous results, the delivery ratio of Soc is 3% higher than that of Loc with respect to different network sizes. The average delay as shown in Fig. 5.11b for both schemes suggests that it is very low in the case of the number of nodes equal to 20. It increases when nodes number is 40 and then decreases. The Soc scheme takes more average delay when the network size is small while it uses less when the network size becomes large. The average cost as shown in Fig. 5.11c presents that the Soc scheme has slightly (0.1 hops) better performance than the Loc scheme.

Overall, our comparison of the Soc scheme and the Loc scheme on both synthetic traces and real data traces suggests that they have similar performance in data routing, whose difference is within 5% in most cases. Nowadays user location infor-

mation is considered to be highly private and people usually are not willing to share it over the network, while encounter-based social information like social degree is much less sensitive and easy to obtain. Therefore, the usage of encounter-based social information is more privacy preserving than location-based social information for data routing in DTNs.

5.4 Summary of Contributions

The proposed encounter-based scheme integrates social metrics including social centrality and social similarity to calculate a comprehensive routing metric. We provide comprehensive performance comparisons of Soc together with other encounter-based social schemes. The proposed encounter-based strategy outperforms with other encounter-based social strategies up to 15% in terms of delivery ratio. Our experiment results also show that routing strategies using location-based social information and encounter-based social information have no significant difference in performance: they perform closely in delivery ratio, delay and cost with a slight difference within 5% in most cases. Our analysis indicates that location-based social information is not critical in designing routing strategies, although it can provide accurate position information and mobility pattern of nodes. Due to the fact that physical location information is sensitive and hard to collect, our work suggests that encounter-based social information could be a good substitute for data routing in DTNs for the sake of privacy preserving.

Chapter 6

Community-based Routing Strategy

One characteristic of two proposed social-based routing strategies along with other previous routing schemes in DTN is that they apply a utility function for data forwarding decision making and choose the nodes with higher utility values as relays for data delivery. We call this kind of schemes as utility-based routing strategies. The most widely used utility metric is the frequency of node encounters, where messages are forwarded to a node who is more frequently meeting the destination than the current node [56, 85, 15]. Another important utility metric is the inter-contact time of node pairs, where messages are forwarded to the nodes with smaller inter-contact time to reduce transmission delay [7, 20, 58]. Several strategies consider geographic distance as a metric and try to forward the messages along the shortest geographic distance path [49, 65, 18]. Inspired by the research of social network analysis, the utilities such as social similarity and social centrality are also proposed to enhance data forwarding via social connections [33, 40, 63, 41, 94]. Most existing utility-based routing strategies employ a single or multiple utility metrics to compose their utility functions. The commonly used utilities and their representative strategies are listed in Table 6.1.

Table 6.1: Routing utilities in DTNs

Utility metric	Definition	Strategies
Encountering frequency	Number of encounters in a period of time	PROPHET [56], FC [42], Seek and Focus [85], MaxProp [15], RAPID [7], and etc.
Inter-contact time	The time interval between two contacts	RAPID [7], Two-Hop-Relay [20], ASBIT [58], and etc.
Geographic distance	The distance of user locations	MobySpace [49], CAR [65], MV [18], and etc.
Social centrality	The degree/betweenness of a node in the network	SDM [33], SimBet [24, 12], BubbleRap [40], PeopleRank [63], and etc.
Social similarity	The common social features between two nodes, such as common friends	SimBet [24, 12], Social-Greedy [41], Social feature-based [94], and etc.

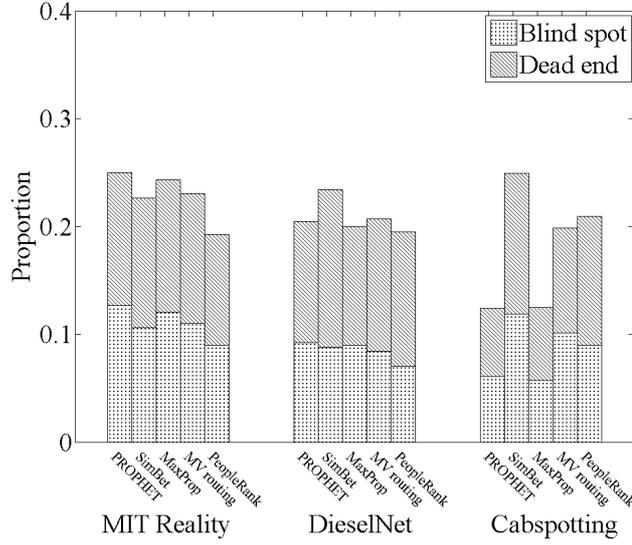


Figure 6.1: Proportion of blinds spots and dead ends in utility-based strategies

There are two problems in utility-based routing strategies: *blind spot* and *dead end*. Consider the scenario that the source node n_i wants to send a message to the destination node n_d . In the utility-based routing, n_i needs to forward the message to a node with higher utility. However, if n_i and all its neighbors have the similar utility near to zero, it will be difficult to decide which node should relay the message. Such problem is called blind spot for the reason that the next hop route is hard to be seen from the current node. The dead end problem occurs when n_i has a higher utility value than all its neighbors, in which case the message is stuck in n_i and not able to be delivered further.

The blind spot and dead end problems are rarely noticed in the previous works, but they commonly exist in utility-based strategies. Fig. 6.1 shows the percentage of blind spot and dead end when applying five utility-based routing strategies in three DTN datasets (MIT Reality [25], DieselNet [16] and Cabspotting [77]). According to the figure, in most utility-based strategies, there are more than 20% of data transmissions encountering the blind spot and dead end problems in the MIT reality trace. Similar percentages are observed in the DieselNet and Cabspotting traces, varying from 14% to 27%. Such problems will clearly affect the delivery ratio of DTN routing, and they are not yet well addressed in the past.

The community-based strategies forward data according community structure

based on the fact that people tend to group into communities by their social relationships. By dividing the network into multiple communities, the nodes within a community have strong connections, while their links across communities are weak ties. The community structure favors intra-community communication where nodes are closely connected (thus reduces the chance of blind spot and dead end), but also encounters the difficulty of inter-community communication via weak links. Existing community-based routing strategies employ naive inter-community mechanism such as flooding [51], or rely on complicated operations to discover direct links [13] or overlapping nodes [33] between communities, which are time-consuming and inefficient. Generally speaking, community-based routing strategies confront two challenges. On the one hand, the existing community partitioning algorithms are complicated and static, which is hard to adapt to the dynamic and mobile DTN environments, thus a *distributed community partitioning* mechanism is desired. On the other hand, to overcome the blindness of data forwarding among communities via weak ties, it needs to measure the *utilities across communities* for a better routing decision making.

The above analysis reveals several weakness of existing DTN routing strategies. In this chapter, we propose a novel community-based routing strategy called Social and Mobile Aware Routing sTrategy (SMART) for DTNs. SMART tackles the above problems to achieve the following objectives: (1) significantly alleviating blind spot and dead end problems; (2) distributed community partitioning; and (3) efficient inter-community communicating.

As we model DTN social graph in Chapter 3, we define the “encounter” between two mobile devices as the event that they move into each other’s communication range. We use a *weighted social graph* to formulate encounters among mobile devices. Each device is denoted by a node in the social graph. If two devices encounter, there is an *edge* between them, which builds a social link between two nodes. We consider two nodes with social link as friends. The number of friends that a node has is called *degree* of the node. The weight of an edge corresponds to the strength of social links, which can be represented by the number of encounters over a fixed time interval.

6.1 Mobile and Social Characteristics of DTN

In this section, we utilize three real DTN data sets as described in Chapter 3 to explore people mobile and social characteristics. The mobility of DTN users reveals

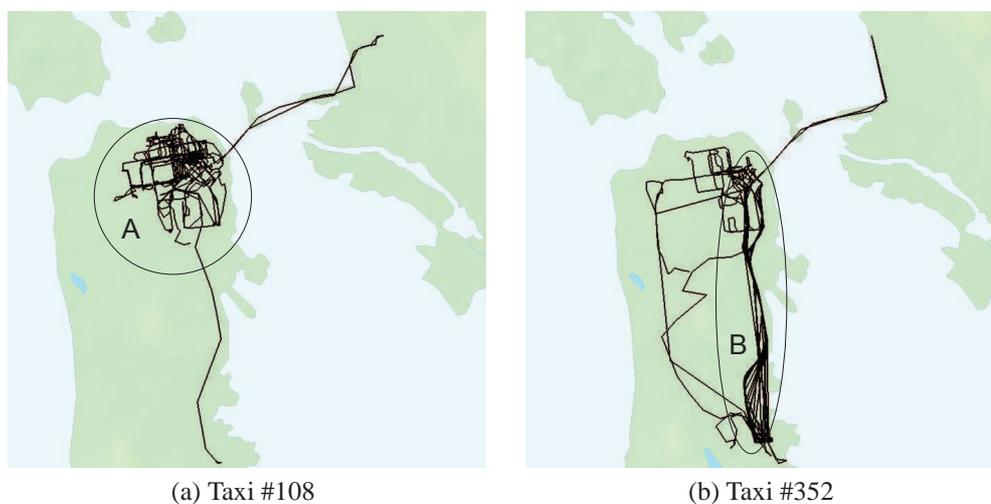


Figure 6.2: Two taxi trajectories in Cabspotting trace

locality and their interactions reveal *community* structure.

People movements are not random and appear regularities restricting in some geographic areas. For example, people move regularly from home to office and vice versa; buses run along stations according to schedules; taxi drivers tend to pick up guests in some popular areas, etc. We observe the mobility trajectories of two taxis (#108 and #352) chosen from the Cabspotting trace, which are illustrated in Fig. 6.2. It can be seen that most of the time taxi #108 moves in area *A*, while taxi #352 moves in area *B*. We call such kind of movement of which the trajectory is mostly restricted in a small area as *locality*.

To further investigate the locality of movements, we analyze the distribution of human mobility scope. We define *geographic mass point* as the centroid of a node's trajectory, which is calculated by averaging the GPS coordinates of its trajectory. We compute the geographic mass points of all nodes in the network and study the geographic distance their trajectory coordinates away from their respective mass points. Taking Cabspotting data trace as an example, Fig. 6.3 shows the probability distribution function (PDF) and cumulative distribution function (CDF) of the trajectory coordinates departing from their mass points in the trace. According to the figure, although the farthest coordinates is 30km away from the mass point, most of the movements are nearby their mass points, e.g., about 80% of trajectory coordinates within an area 5km from mass points. The majority of the coordinates concentrate in areas 2km, 3km, and 4km away from mass points. This verifies the locality property of nodes movements: most movements of DTN nodes are restricted within a range of small distance, and there are only a small proportion of

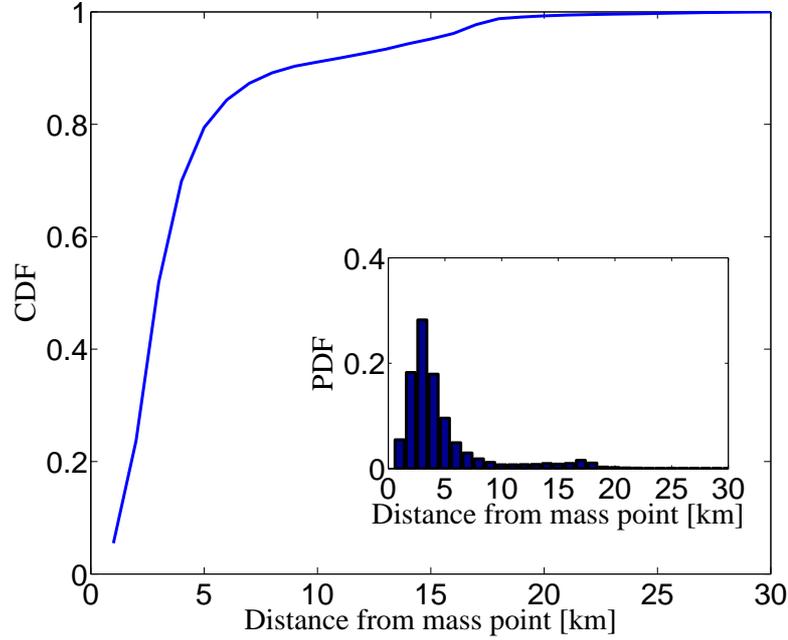


Figure 6.3: The CDF and PDF of node movements (Cabspotting).

long distant movements.

6.1.1 Distributed community partitioning

Community is defined as a social unit that shares common value. It is a tight and cohesive social entity. Intuitively, communities are formed based on locations or interests [29, 70, 8, 68]. People in the same geographic location or sharing the same interest are likely to be in the same community. In the context of DTNs, we use geographic locations to study community structure and investigate the relation between encounters and geographic distances for the discovery of communities. It is likely that there is correlation between encountering and geographic location: the closer two nodes, the more often they meet each other. We calculate the number of encounters between node pairs as a function of the distance of their geographic mass points in Cabspotting, as shown in Fig. 6.4. It illustrates that the number of encounters decreases rapidly when the distance increases, which implies that when the distance becomes longer, the number of encounters becomes smaller. There is no node pair with distance greater than $4km$ having more than 100 encounters. This verifies the strong correlation between location and encountering.

Inspired by the above observation, we propose a dynamic community partition-

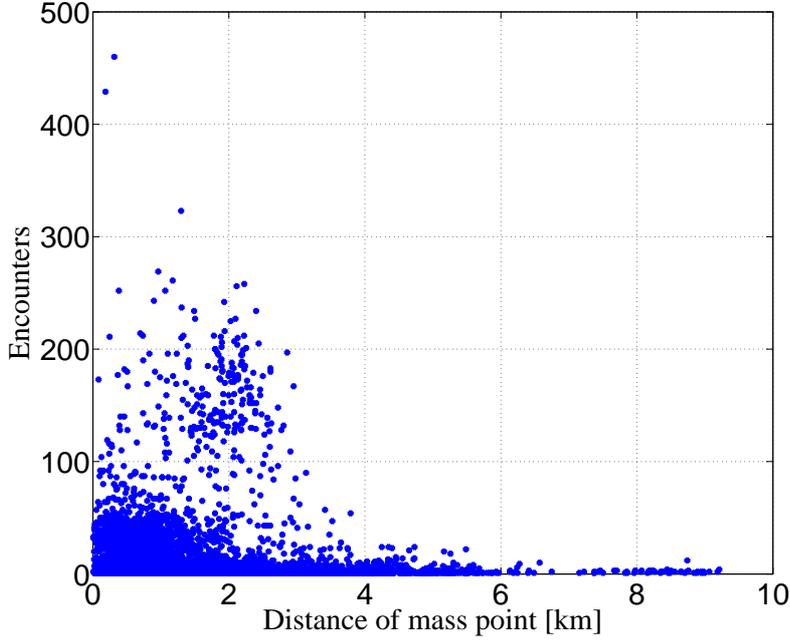


Figure 6.4: The number of encounters vs. distance (Cabspotting).

ing algorithm. Unlike existing community partitioning algorithms such as k-clique percolation [71] and Girvan Newman algorithm [67] etc. where global network topology information is needed, our algorithm only uses local encountering statistics of nodes. It is more suitable for distributed implementation in DTNs. The basic idea is adaptively grouping nodes into communities starting from a random partition (i.e. m communities) of the network. The detailed community partitioning process is described as follows.

The community construction process is divided into two stages: the bootstrap stage and the evolution stage. In the bootstrap stage, m nodes are randomly selected and each node is assigned with a unique community ID. Node without community affiliation will choose one community through encounters until every node in the network is assigned with a community ID. After this stage, the network has m communities. In the evolution phrase, each node counts the affiliation parameters (APs), which indicate the number of encounters with nodes in different communities. Then it adjusts the community affiliation according to updated APs. We use a vector to present the affiliation parameters of node n_i :

$$E_i = \{ap_{1_i}, ap_{2_i}, \dots, ap_{m_i}\},$$

where ap_{j_i} is the AP that n_i connects to community C_j denoting the number of encounters between n_i and C_j . When n_i encounters a node in community C_j , it updates its AP value accordingly, and adaptively changes its community affiliation to the community with maximal AP value in the vector. We call our algorithm *m-partition*.

Algorithm 3: m-partition algorithm

input : node N_i and AP vector
output: the community ID of node N_i

begin

- Assume there are m communities to be detected ;
- for** *Encounter with N_j* **do**
 - if** $N_i.communityID = null$ **then**
 - $N_i.communityID = N_j.communityID$;
 - else**
 - $y \leftarrow N_j.communityID$;
 - $x \leftarrow N_i.communityID$;
 - if** $y = x$ **then**
 - $ap_{x_i} = ap_{x_i} + 1$;
 - else**
 - $ap_{y_i} = ap_{y_i} + 1$;
 - if** $ap_{y_i} > ap_{x_i}$ **then**
 - $N_i.communityID = y$;

The algorithm runs dynamically as each encounter occurs in the network in a distributed fashion. Therefore, the community structure may change from time to time and is maintained dynamically. We show that the communication cost of m-partition for maintaining community members is low in DTNs. Given two communities A and B , there are m nodes in A and n nodes in B . Suppose a node n_i in community A needs to switch its community from A to B , it first obtains the new community list from the encountered node in community B . If we consider the traffic overhead for transmitting one node ID as 1, the communication overhead for obtaining community members will be n . It then floods its ID to its new community B to make other nodes in community B be aware of n_i . The communication cost will also be n . Furthermore, when a node in community A meets a node in community B , it checks the community member list to see whether any node changes their community identity. In this case, n_i changes from A to B . Suppose there are k encounters between community A and B . The cost for transmitting node ID of

n_i is k . The node in community A floods this information to make the remaining nodes ($m - k$ nodes) in A exclude the membership of n_i from their local community, which needs $m - k$ transmissions of ID of n_i . Overall, the communication overhead for maintaining community members caused by one community switch action is $O(m + 2n)$.

6.1.2 Locality of user contacts

According to the above observation, it has high correlation between user movements and their encounters. Thus the locality of user movements is also reflected to their encounters. To investigate such locality, we define two types of encounters after community partitioning: when two nodes move into each other's communication range, if both nodes are from the same community, we call such encounter a *local contact*; if the encountering nodes are from different communities, it is called a *remote contact*. Since a node tends to move in a local area, which results in frequent local contacts. In contrast, only a small proportion of user movements are long-distant, which yields remote contacts forming cross-community communications.

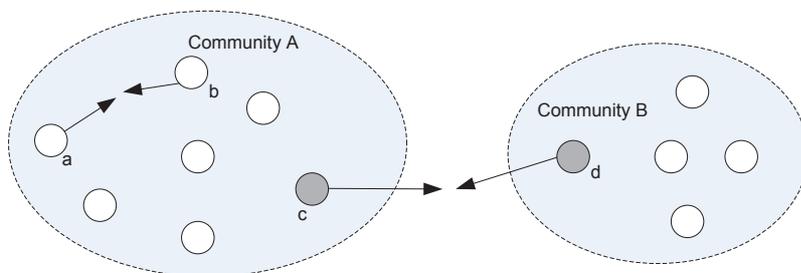


Figure 6.5: Local contact and remote contact

Fig. 6.5 shows an example of local and remote contacts. In the figure, arrows denote the movement of nodes. The pair $\langle a, b \rangle$ makes a local contact, and the pair $\langle c, d \rangle$ makes a remote contact. We calculate the proportion of local contacts and remote contacts of the three DTNs after applying our community partitioning algorithm (with $m = 10$). The results are shown in Table 6.2. It presents that local contacts are majority for all three traces and remote contacts are minor. MIT Reality has 19.7% remote contacts; DieselNet has 7.3%; and Cabspotting has 9.9%, which confirms the locality of user contacts. Although remote contacts only take a

small fraction, they play an important role for information exchange since they are the “bridges” between communities. The more remote contacts, the more active a community interact with others. It enlightens the idea of using local contacts for intra-community communication and remote contacts for inter-community communication. The detailed approach is presented in the subsequent section.

Table 6.2: Proportion of local and remote contacts

<i>Traces</i>	<i>MIT Reality</i>	<i>DieselNet</i>	<i>Cabspotting</i>
Local contact (%)	80.3%	92.7%	90.1%
Remote contact (%)	19.7%	7.3%	9.9%

6.2 Strategy Design

In this section, we introduce the social and mobile aware routing strategy (SMART) for DTNs. The basic idea of SMART is to divide the DTN into several communities using community partitioning algorithms, deliver the message in the same community via local contacts, and forward the message to other communities via remote contacts. The details are described in the following.

6.2.1 Assumptions

Before presenting the detailed design of SMART, we introduce some basic concepts and assumptions.

- Assume a dynamic community partitioning process (i.e. *m-partition*) is applied to cluster the social graph into a number of communities $= \{C_1, C_2, \dots, C_M\}$.
- Each node n_i is assigned with a community ID C_i and a set $\pi(C_i) = \{n_j | \forall n_j \text{ in } C_i\}$ indicating the members in the same community. The community ID and community members are obtained and maintained from the result of community partitioning process.
- Each node records the encounter history for a period of time ΔT , where ΔT is a time window representing one day or one week interval. Each node maintains the history of local contacts that occur within the local community and remote contacts that occur across different communities.

Table 6.3: Local contact table

n_1	n_2	\cdots	n_i	\cdots	n_N
ζ_{i1}	ζ_{i2}	\cdots	-1	\cdots	ζ_{iN}

- Local contacts are recorded into a *local contact table*, whose format is shown in Table 6.3. It records the contact frequency of a node with other nodes in the same community. In the table, n_j ($j = 1, 2, \dots, N$) is the ID of a node in the local community and ζ_{ij} ($j = 1, 2, \dots, N$) is the contact frequency between n_i and n_j in ΔT . If $n_j = n_i$, let $\zeta_{ij} = -1$ indicating that the contact with itself is not countable.

The contact frequency is the number of encounters over the time period ΔT . It is calculated by:

$$\zeta_{ij} = \frac{\sum_{t=0}^T X(t)_{ij}}{\Delta T},$$

where $X(t)_{ij} = 1$ if two nodes contact at time t , otherwise, $X(t)_{ij} = 0$.

Table 6.4: Remote contact table

C_1	C_2	\cdots	C_i	\cdots	C_M
η_{i1}	η_{i2}	\cdots	-1	\cdots	η_{iM}

- Remote contacts are recorded into a *remote contact table*, whose format is shown in Table 6.4. It records the contact frequency of a node with other communities. In the table, C_j ($j = 1, 2, \dots, M$) is the ID of a community (there are M communities in the network) and η_{ij} ($j = 1, 2, \dots, M$) is the sum of encounters that n_i with nodes in C_j over ΔT . Again, $\eta_{ij} = -1$ when $C_j = C_i$ suggesting that the local contacts is not reflected from the remote contact table.

When a source node n_s sends a message to the destination n_d , n_s checks the local community members to see whether n_d is in the same community. If n_s and n_d are in the same community, we apply an *intra-community communication* process. If n_s and n_d are in different communities, we apply an *inter-community communication* process. The two processes are described as follows.

6.2.2 Intra-community communication

If a source node n_s and a destination node n_d are in the same community by checking local community members, it is possible to apply traditional utility-based strate-

gies for data forwarding. However, to avoid the problems of blind spot and dead end as mentioned before, the intra-community routing scheme needs to be carefully designed.

To against blind spot and dead end, we need a social featured routing metric that accumulates encounter effects and decays according to the node's social status and time elapsed. We consider each encounter has an effect to the utility value, which is positively correlated to the social relation (i.e. social similarity) between two nodes. That is, two nodes with closer social relation lead to higher utility increase in each encounter. Besides, the encounter effect decays depending on its social status and time elapsed. An earlier effect will have less effect remaining due to temporal factor. Meanwhile, a node with high social status will motivate further encounters. To represent this motivation, a node with higher social status in the network should have a slower decaying speed on the encounter effect. Combining temporal and social factor, each encounter effect decays as the social status of a node and the time elapsed. To select a relay, the scheme will evaluate the accumulative effects produced by the encounters and the decaying speed of the effects. We provide the formulation of our scheme as follows.

As the first step, we give the definition of social relation and social status. The social relation denotes the social closeness between two nodes and social status shows the relative importance of nodes in the social network. Although the social relation and social status can be represented in many sophisticated manners [2, 24], we choose two representative expressions to illustrate our scheme. Namely, we use *social similarity* to represent the social relation, and *social centrality* to represent social status.

Social similarity: it is defined as the number of common friends between a pair of nodes, indicating the trustiness and cohesiveness of social ties between them [23, 22]. As explained in Chapter 5, social similarity can be calculated by the following equation.

$$\mathcal{S}_{i,j}(\tau) = |F_i(\tau) \cap F_j(\tau)| + 1, \quad (6.1)$$

where $F_i(\tau)$ ($F_j(\tau)$) is the set of friends of node n_i (n_j) at time τ . The intersection operation is to obtain the common friends between two nodes and plus 1 is to eliminate the effect of 0. When two nodes encounter, they exchange their friend lists to calculate the social similarity.

Social centrality: it is a quantification of the relative importance of nodes in the social network. There are various definitions of centrality. We use the Freeman's degree [30] to define social centrality as described in Chapter 5. For a node n_i , its

centrality is defined as follows.

$$\mathcal{C}_i(\tau) = \frac{\sum_{k=1}^N d_{ik}(\tau)}{N}, \quad (6.2)$$

where $d_{ik}(\tau) = 1$ if a direct link exists between n_i and n_k at time τ and N is the number of nodes in the community.

Both social similarity and social centrality only require local information, and they can be calculated locally in DTNs by exchanging information with neighbors.

The encounter effect between two nodes is therefore denoted by the social similarity. To model the decaying effect, we introduce a decay function with respect to social centrality and time as follows.

$$\mathcal{D}_i(t - \tau) = \frac{\mathcal{C}_i(\tau)}{t - \tau}, \quad (6.3)$$

where τ is the time when the encounter occurs. The decay function relying on both social centrality and elapsed time reduces the accumulative effects of the utility value.

If an encounter occurs in each time unit, the accumulative effects of encounters between two nodes with decay between node n_i to destination n_d can be formulated as the convolution of Eq. 6.1 and Eq. 6.3.

$$\begin{aligned} Y_{i,d}(T) &= \mathcal{S}_{i,d}(T) \otimes \mathcal{D}_i(T), \\ &= \int_{\tau=0}^T \mathcal{S}_{i,d}(\tau) * \mathcal{D}_i(T - \tau). \end{aligned} \quad (6.4)$$

However, the encounter only occurs in several time units. Therefore, the accumulative effects of encounters are represented by a discrete convolution as:

$$U_{i,d}(T) = \sum_{\tau=0}^T X(\tau)_{id} * \mathcal{S}_{i,d}(\tau) * \mathcal{D}_i(T - \tau), \quad (6.5)$$

where $X(\tau)_{id} = 1$ when an encounter occurs at time τ or when $\tau = 0$ (to initialize the utility value), otherwise, $X(\tau)_{id} = 0$. The utility function describes that when each encounter occurs, it yields an encounter effect represented by social similarity. Each effect occurs at different time decays as a decay function composed by social centrality and time, indicating the encounter effects of a node with higher social status decays slower than a node with poor connection to the network and a recent

encounter effect decays slower than a older encounter effect.

Based on the above utility function, we propose our intra-community routing principle.

Intra-community routing principle: when two nodes n_i and n_j in the same community encounter, the routing utility in Eq. 6.5 is calculated and the node with higher value will be chosen to forward the message. The forwarding path will be recorded to make it free from loop.

6.2.3 Inter-community communication

If a destination node n_d does not belong to the community of source node n_s , we need to choose some relay nodes to forward the message among communities. The idea is using “fringe nodes” to bridge the communication of inter-communities.

A fringe node is a node which is capable to remote contact with other communities. It is measured by the number of links that it connects to other communities. We select nodes with higher number of links to outside of local community as fringe nodes. Each fringe node is represented by its ID and the remote contact table as mentioned in Table 6.4 to indicate its links to other communities. In our proposed scheme, each community maintains a set of fringe nodes \mathcal{F} . The set \mathcal{F} is randomly selected initially, and is updated periodically. During a period ΔT , each node compares its remote contact table with the fringe nodes. If a node n_i finds that it has closer connection with outside communities than a fringe node n_j , it will announce itself as the new fringe node. The comparison is described as follows.

Assume $\eta_{i1}, \eta_{i2}, \dots, \eta_{iM}$ is the remote contact frequency of n_i , and $\eta_{j1}, \eta_{j2}, \dots, \eta_{jM}$ is the remote contact frequency of n_j . Define a function $\phi(x, y) = 1$ if $x \geq y$ and $\phi(x, y) = -1$ for the rest. The selection of fringe node is determined by the value

$$\mathfrak{F}_c = \sum_{k=1}^M \phi(\eta_{ik}, \eta_{jk}).$$

If \mathfrak{F}_c is larger than 1, it means n_i has better connection than n_j , and n_i becomes the new fringe node and announces to the other nodes in the local community.

According to the report in [91], a small fraction rewired links are enough to create a small world network. Our analysis to the three traces shows that the fraction of remote contacts varies from 7.3% to 19.7% (as shown in Table 6.2), so we set the number of fringe nodes as 10% of the community size. If the community size is smaller than 10, we set the number of fringe nodes as 1.

Due to dense network connection in the community, the set of fringe nodes \mathcal{F} and their remote contact tables can easily spread to nodes in the same community. Since there are more than one fringe nodes which can reach to other communities, it needs to carefully choose the forwarding node for performance consideration.

Assume the source node n_s in community C wants to send a message to the destination n_d in community C' ($C \neq C'$). We first decide whether C and C' are directly connected by looking up the set \mathcal{F} in C to check whether there exists a fringe node connecting to C' . If there is a connection from fringe node set to the destination community, we say C and C' are directly connected. Otherwise, we say they are indirectly connected.

If C and C' are directly connected, we need to forward the message to a fringe node who can reach C' . There might be more than one fringe nodes directly connecting to C' , and the candidate set is $\mathbb{C} = \{n_j | \forall n_j \in \mathcal{F} \text{ and } (n_j \text{ connects to } C')\}$. We need to decide which node in the candidate set as a relay. Our principle is to send the message to a relay having more connections with C' . By looking up the remote contact tables of fringe nodes, the algorithm chooses the one with the maximal number of encounters to C' as the relay. If C and C' are indirectly connected, we select the fringe node with maximal number of encounters with outside of C by summing up entries in remote contact tables. The source node forwards the message to the selected fringe node by intra-community routing strategy. After the data is forwarded to the dedicated fringe node, the data transmission between communities becomes a challenge.

To enable efficient inter-community communication, we propose a utility function extended from intra-community utility to forward data from the fringe node to the destination community. Namely, we extend the utility function from node-to-node to node-to-community for inter-community communication. We build the utility function between a fringe node f to the destination community C' . To construct such utility, we consider social relation between f and C' as similarity between the node f and a set of nodes in C' . However, knowing the friends of all nodes in C' would suffer too much overhead in DTNs. In this case, we provide an estimation that we only count the friends of nodes who have ever encountered with f . The similarity is defined as:

$$\mathcal{S}_{f,C'}(\tau) = |F_f(\tau) \cap F_{C'}(\tau)| + 1,$$

where $F_{C'}(\tau)$ indicates the friends of a set of nodes in C' that ever encountered with

f until time τ .

To formulate the social status of node f , we extend the concept of centrality from the local community to the entire network. We call it community centrality, denoted by $\mathcal{C}_\Gamma(\tau)$, which is defined as the proportion of the number of communities it connecting with ($M_c(\tau)$) to the total number of communities ($M(\tau)$) at time τ . It is defined as:

$$\mathcal{C}_\Gamma(\tau) = \frac{M_c(\tau)}{M(\tau)}.$$

The decay function in the node-to-community utility becomes $\mathcal{D}_\Gamma(t - \tau) = \frac{\mathcal{C}_\Gamma(\tau)}{t - \tau}$. The overall utility function from the node f to community C' thus is defined as:

$$U_{f,C'}(T) = \sum_{\tau=0}^T X(\tau)_{fC'} * \mathcal{S}_{f,C'}(\tau) * \mathcal{D}_\Gamma(T - \tau), \quad (6.6)$$

where $X(\tau)_{fC'} = 1$ when an encounter occurs between node f and community C' at time τ or when $\tau = 0$ (to initialize $U_{f,C'}(T)$). According the utility function, the fringe node finds the next relay by choosing a node with higher utility value with destination community. The procedure continues until the data reaches destination community.

6.3 Discussion

In this section, we discuss how SMART tackles blind spot and dead end problems, and the efficiency of inter-community routing by SMART.

6.3.1 Tackling blind spot and dead end problems

The blind spot and dead end problems result from the indecisive utility value of utility-based data routing strategies. The blind spot occurs when utility values of a node and its neighbors are close and nearly to 0. It cannot decide which node should be the next relay. The emergence of blind spot is because these nodes have rare interaction with the destination, which generates similar utility values close to 0. The dead end arises with all neighbors of a node having lower utility value than it. The node cannot conduct the forwarding behavior in the network. The existence of dead end comes from the fact that the utility value of the relay reaches the peak locally.

We conduct experiments to show the percentage of blind spot and dead end

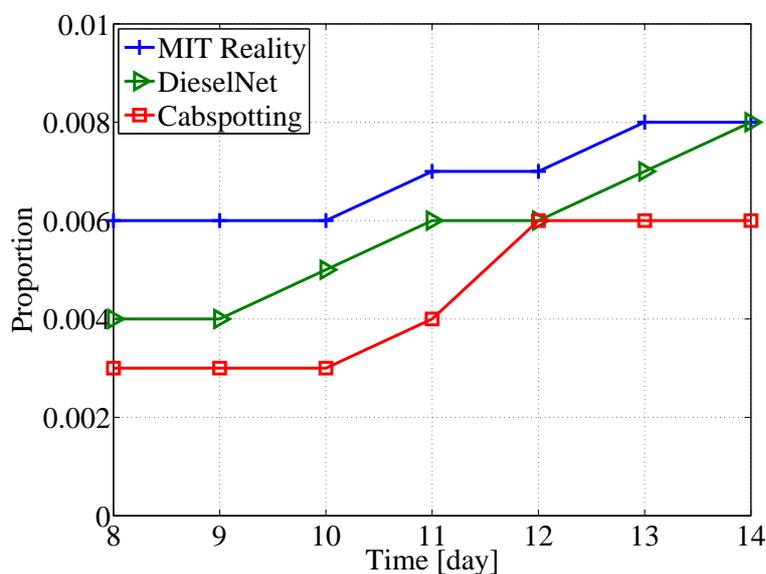


Figure 6.6: Percentage of blind spot and dead end in SMART.

occurring in SMART. By definition, we consider a node encounters blind spot if it has similar utility value closing to 0 as its neighbors and it cannot find the next relay until the data is expired. A node runs into dead end if its utility value is larger than all its neighbors and the data is stuck into the node until it is expired. We look for the data routing failures caused by blind spot and dead end in three data sets (MIT Reality, DieselNet and Cabspotting). We sum the two types of failures and draw the curve as a function of time as shown in Fig. 6.6. According to the figure, there is only a tiny percentage of blind spot and dead end appearing in SMART, most of the time lower than 1%. Compared to the experiment results shown in Fig. 6.1, the percentage drops dramatically. Thus we claim that SMART significantly alleviates the blind spot and dead end problems.

6.3.2 Efficiency of inter-community communication

The inter-community communication is a difficult task in community-based strategies since it lacks of strong links between different communities. Existing community-based strategies use naive routing mechanism such as flooding [51], or rely on discovering the direct links [13] or overlapping nodes [33] between communities, which are time-consuming and inefficient.

SMART enhances the capability of inter-community communication by selecting fringe node and a node-to-community utility function. The fringe node is selected by the criterion that has rich connection to the remaining network. To bridge the communication between the fringe node and the destination community, SMART introduces a node-to-community utility function, which considers the destination community as an entity. Analogous to intra-community utility, we compose utility function between fringe node and the destination community, and build routing channels among different communities.

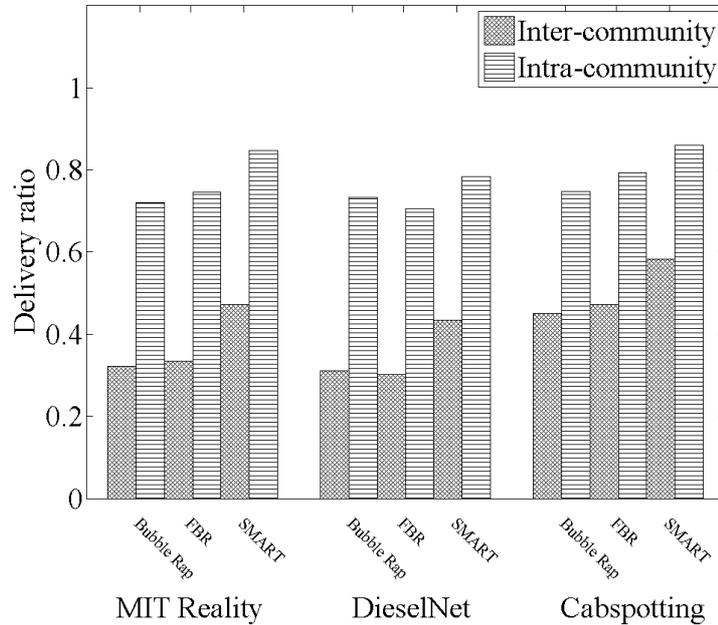


Figure 6.7: Delivery ratio of community-based strategies

We compare the performance of SMART with other community-based strategies, including Bubble Rap [40] and Friendship Based Routing (FBR) [13] to show its efficiency. The comparison of delivery ratio on three data traces (MIT Reality, DieselNet and Cabspotting) is shown in Fig. 6.7. It is illustrated that for inter-community communication, the delivery ratio of Bubble Rap and FBR is 32% and 33% respectively, while the delivery ratio is improved in SMART greatly, which achieves 47%. For delivery ratio of intra-communication, SMART achieves 85%, which also outperforms the other strategies (with 72% in Bubble Rap and 75% in FBR) on MIT Reality. The performance comparison on DieselNet and Cabspotting also show the efficient inter-community communication of SMART. Together with our distributed community partitioning algorithm, the proposed SMART strategy

shows its effectiveness in dealing with the community-based routing problems.

6.4 Performance Evaluation

In this section, we conduct experiments to evaluate the performance of SMART and compare it with other existing DTN routing strategies.

6.4.1 Experiment setup

We still use HuggleSim simulator [39] for the experiment evaluation. 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. In this group of experiments, each message only keeps one copy in the network. We run every experiment 20 times for statistical convergence. 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 SMART algorithm is implemented and is compared to other existing DTN routing algorithms.

6.4.2 Impact of community numbers

We first investigate the impact of the number of communities on the performance of SMART. We apply the proposed m -partition algorithm for community partitioning on the three DTN traces and then use SMART to route messages.

Fig. 6.8, Fig. 6.9 and Fig. 6.10 show the performance metrics as a function of community number m (varying from 1 to the size of the data sets) and time. The delivery ratio of MIT Reality trace is shown in Fig. 6.8a. According to this figure, when no community partitioning algorithm is applied ($m = 1$), the delivery ratio is quite low and it increases slowly with time. As the community number is set to an appropriate value (e.g. $m = 10$), the delivery ratio increases dramatically, which is almost 2 times as that when $m = 1$. For $10 \leq m \leq 90$, the delivery ratio becomes stable and has only small fluctuation. When the community number approaches to the size of the data set ($m = 97$), the performance drops dramatically since the impact of community structure disappears. The average delay is illustrated in Fig. 6.8b. It is seen that the average delay is almost the same for all community numbers

6.4. Performance Evaluation

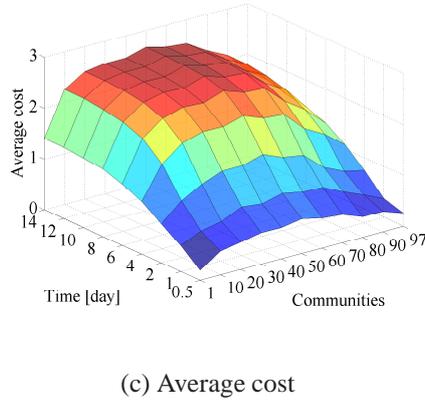
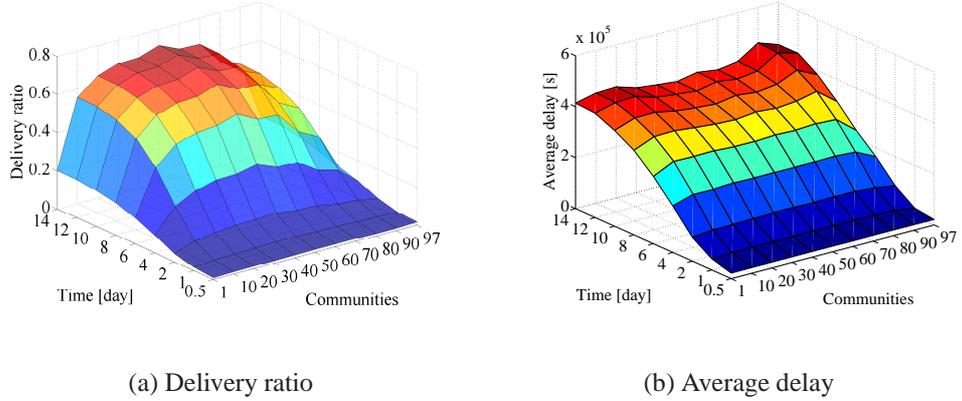


Figure 6.8: The performance metrics as a function of community number and time (MIT Reality).

and it only varies with time. The average cost is shown in 6.8c. Similar to delivery ratio, the average cost is influenced by m and it increases to a stable value when $10 \leq m \leq 90$. Similar results are also found in DieselNet as shown in Fig. 6.9 and Cabspotting as shown in Fig. 6.10. It suggests that SMART performs better when the community structure is outlined, while the performance of SMART is low when no community structure is indicated in the network. It also reveals that the proper value of m is within a wide range. We tend to choose a smaller value of m to reduce the cost of maintaining the community structure. Thus in the rest of our experiments, we fix our community number to $m = 10$.

6.4.3 Impact of community partitioning algorithms

We will show that the performance of SMART routing scheme based on different community partitioning algorithms. We evaluate the performance of SMART using

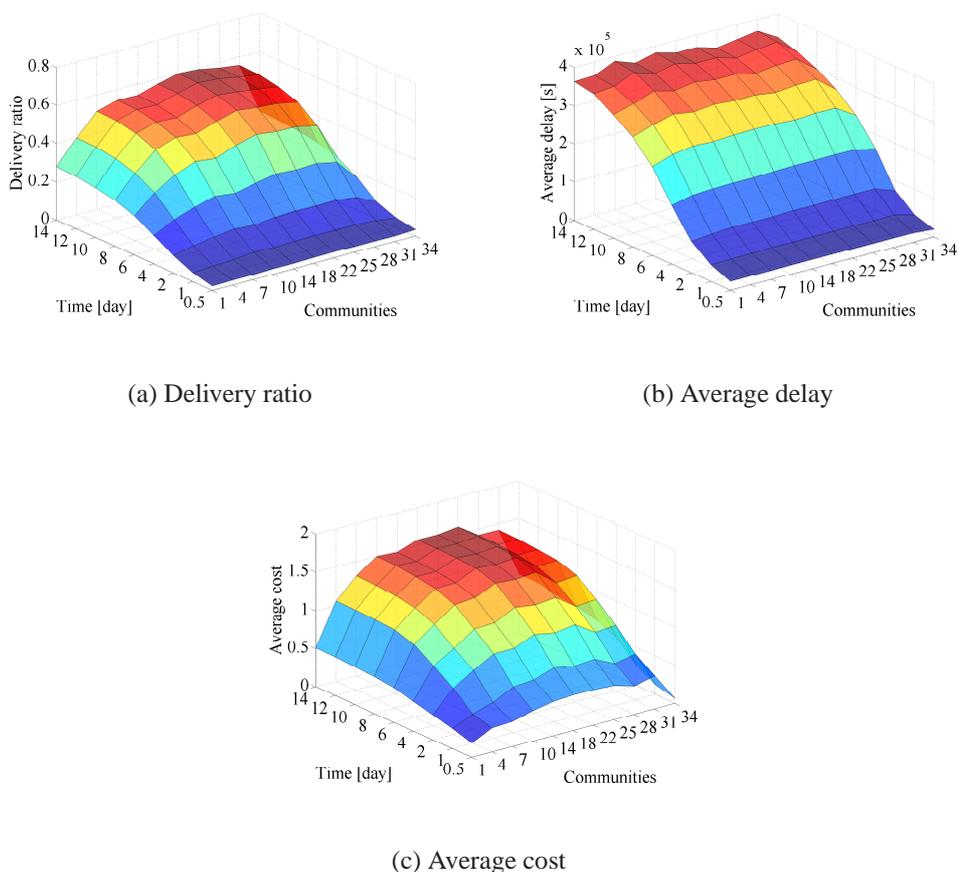


Figure 6.9: The performance metrics as a function of community number and time (DieselNet).

different community partitioning algorithms, including m-partition, k-clique percolation algorithm [71] (which considers the adjacent k-clique as communities), and Girvan Newman algorithm [67] (which continues removing edges with the highest betweenness until a certain threshold is reached).

Fig. 6.11 presents experimental results of MIT Reality trace. As shown in Fig. 6.11a, the m-partition method outperforms Girvan Newman by 10% and k-clique percolation by 2% in delivery ratio. In terms of average delay, as shown in Fig. 6.11b, m-partition performs slightly better than the other two algorithms. The three algorithms takes similar average cost as shown in Fig. 6.11c. Similar results are also observed on DieselNet data set as shown in Fig. 6.12 and Cabspotting data set as shown in Fig. 6.13.

Despite the different algorithms used for community partitioning, the routing performance is quite similar for all three DTN data sets. It indicates that our proposed m-partition community detection method can adapt to the implementation

6.4. Performance Evaluation

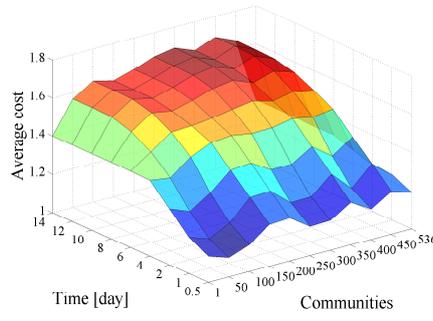
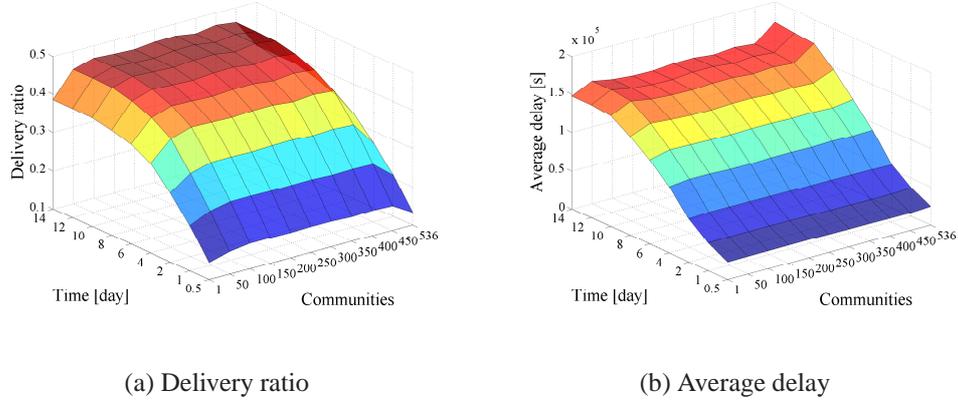


Figure 6.10: The performance metrics as a function of community number and time (Cabspotting).

of SMART which performs similar as other sophisticated community partitioning algorithms relying on global information. Since Girvan Newman and k-clique percolation need global network topology information which is very difficult to obtain in DTNs, the proposed m-partition algorithm is more suitable for distributed implementation in the real world.

6.4.4 Performance analysis

We compare SMART with five existing DTN routing strategies: PROPHET [56], SimBet [24], Bubble Rap [40], Friendship Based Routing (FBR) [13], and Epidemic routing [89]. PROPHET is a utility-based strategy according to encounter histories. It forwards data to the nodes with higher delivery rate based on contact history. SimBet is a utility-based strategy according to social features. It considers social properties including similarity and centrality to make data forwarding decision. Bubble Rap is a community-based strategy. It depends on community

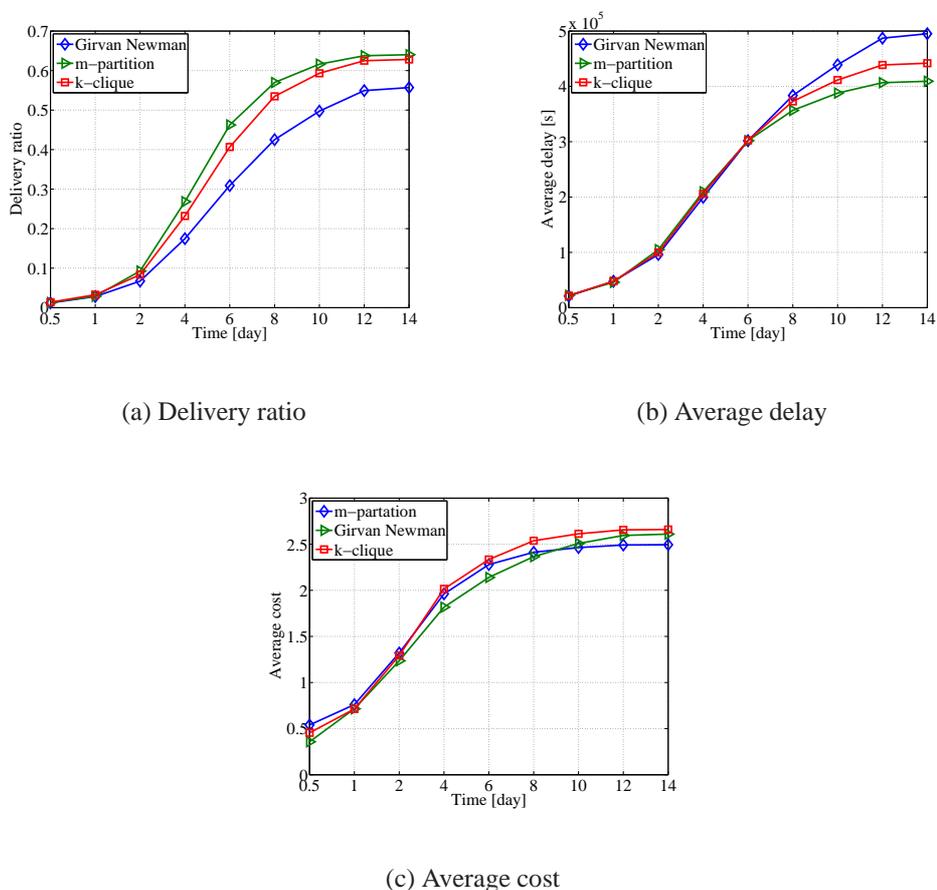
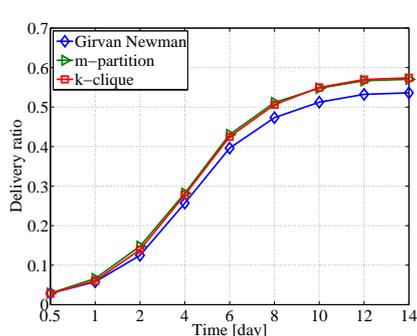


Figure 6.11: The performance of SMART under different community partitioning algorithms (MIT Reality).

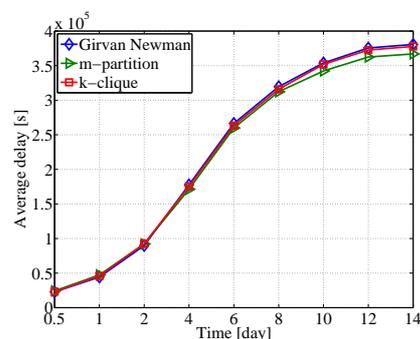
structure and routes data based on rankings calculated from social centrality. FBR algorithm as another community-based algorithm presented in [13], it constructs temporal community and use the nodes with direct connection to the destination community for data delivery. Epidemic routing is a flooding strategy. It has high delivery cost, but its delivery ratio and delay approach the theoretical bound.

The performance comparison in three data sets is presented in Fig. 6.14, Fig. 6.15 and Fig. 6.16. Fig. 6.14 shows the performance of various algorithms as a function of time on MIT Reality trace. The delivery ratio is compared in Fig. 6.14a. It shows that SMART outperforms PROPHET, SimBet, FBR and Bubble Rap. The delivery ratio of SMART is about 10% higher compared to Bubble Rap and FBR, 15% higher than that of SimBet and nearly 20% higher than that of PROPHET. The results also confirm that SMART outperforms utility-based strategies nearly 20% by solving blind spot and dead end problems. The reason PROPHET performs the worst is due to the strong community structure of MIT Reality trace. When source

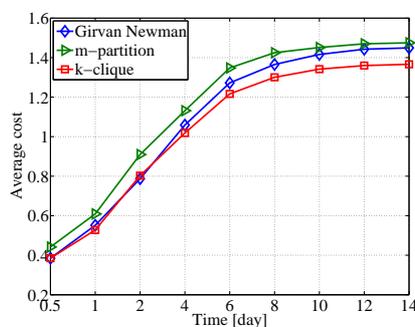
6.4. Performance Evaluation



(a) Delivery ratio



(b) Average delay



(c) Average cost

Figure 6.12: The performance of SMART under different community partitioning algorithms (DieselNet).

and destination are inter-connected by a long path, PROPHET will encounter the blind spot and dead end problems, which degrade its performance. SimBet exploits social properties to enhance the delivery ratio but it also encounters high proportion of blind spot and dead end problems. Bubble Rap and FBR takes advantages of community structure, so they perform better than PROPHET, but not as well as SMART. Since Epidemic routing represents the theoretical upper bound of delivery ratio, the performance of SMART is below the upper bound. Average delay is compared in Fig. 6.14b. Again, the delay of SMART is lower than the other four strategies (most of the time their performance are very close), but higher than the lower bound (Epidemic routing). Average cost is compared in Fig. 6.14c. The cost of PROPHET is the highest. This indicates that “transitivity” in PROPHET is not accurate enough to predict the relay selection, thus it has longer relay path than others. The SMART is a little higher than others due to the decaying effect, which makes SMART take more relays for data delivery.

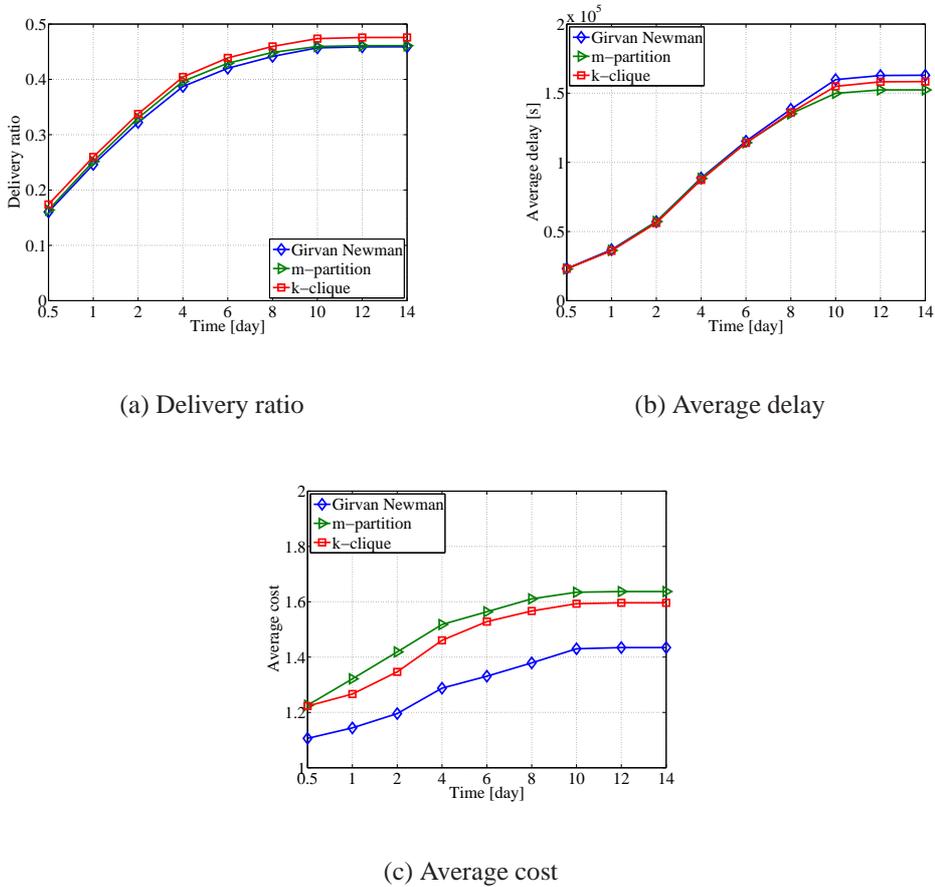


Figure 6.13: The performance of SMART under different community partitioning algorithms (Cabspotting).

Fig. 6.15 presents the performance results of various algorithms as a function of time on DieselNet data set. The delivery ratio is depicted in Fig. 6.15a. SMART outperforms Bubble Rap by 3%, FBR by 5% and PROPHET by 8%. It has nearly 20% higher of delivery ratio than SimBet. Regarding the average delay and the average cost of each strategy as shown in Fig. 6.15b and Fig. 6.15c, SMART has very close average delay with Epidemic, which is less than other strategies. The average cost of SMART is about 50% of that of PROPHET and higher than FBR and SimBet. DieselNet has very similar network structure with MIT Reality and thus has similar trend on delivery ratio with MIT Reality. However, due to the regular and repetition routine of buses in DieselNet, it makes the SimBet meet dead ends quite often and takes 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 and FBR perform close to SMART. SMART has similar cost with social-related strategies but much lower cost than PROPHET.

6.4. Performance Evaluation

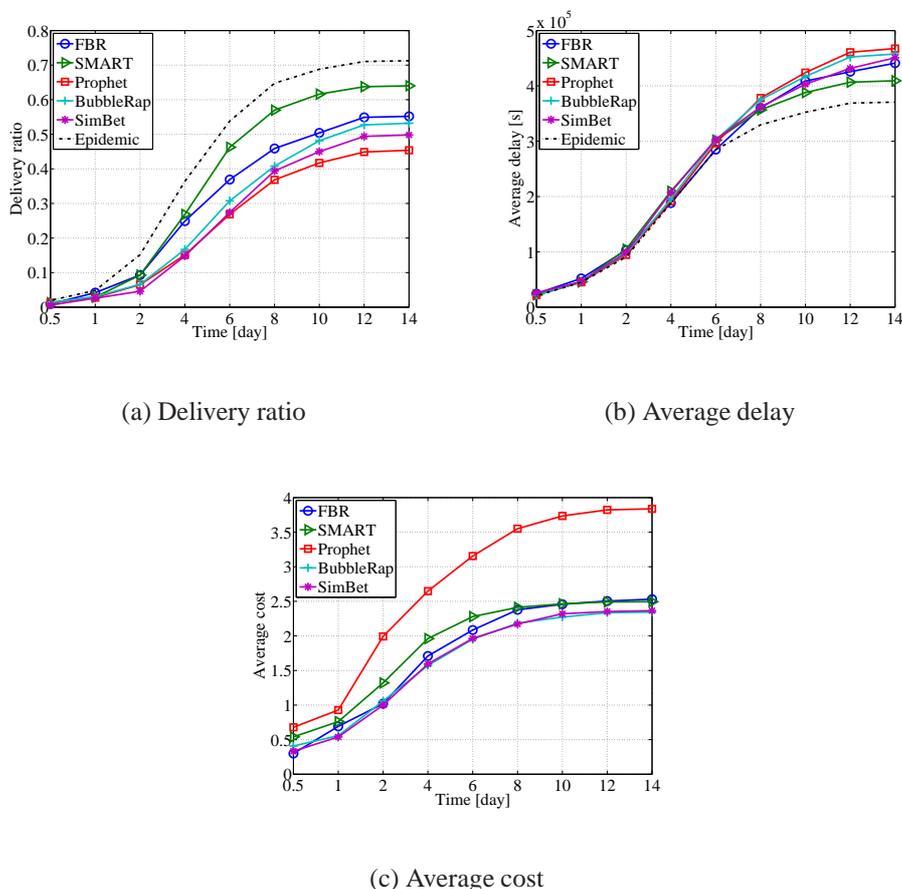


Figure 6.14: The performance comparison of various strategies on MIT Reality Mining trace

Comparison of different algorithms' performance on Cabspotting trace is shown in Fig. 6.16. Fig. 6.16a depicts the delivery ratio of varied algorithms as a function of time. The SMART has very similar performance as PROPHET. It outperforms FBR by 5%. 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. 6.16b, SMART costs as low as Epidemic algorithm delay, which is much lower than others. The average costs of various algorithms are similar as shown in Fig. 6.16c.

In a summary, the proposed SMART strategy outperforms the utility-based and community-based strategies on various DTN data sets in most of the performance metrics.

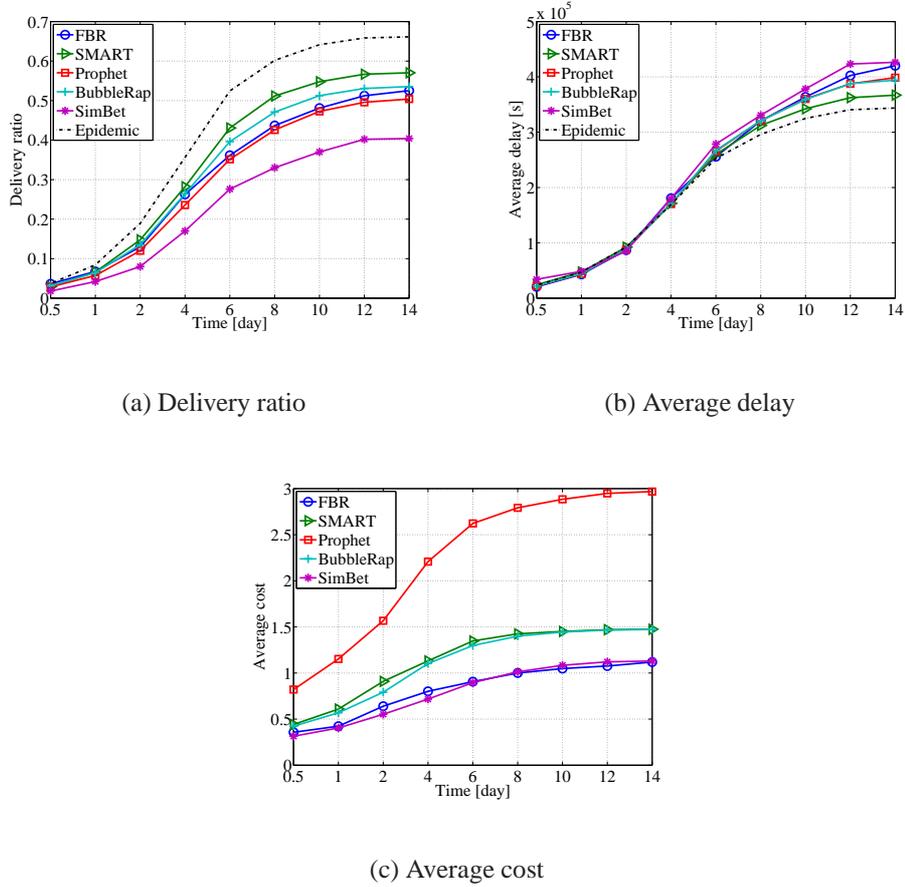
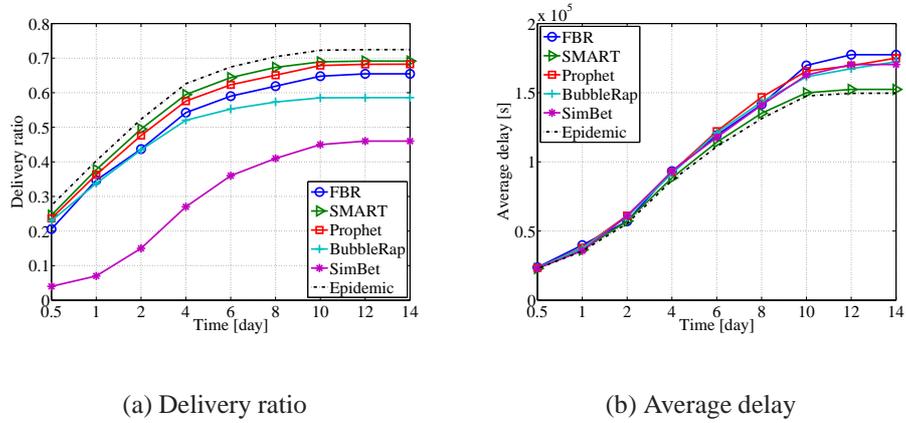


Figure 6.15: The performance comparison of various strategies on DieselNet trace

6.5 Summary of Contributions

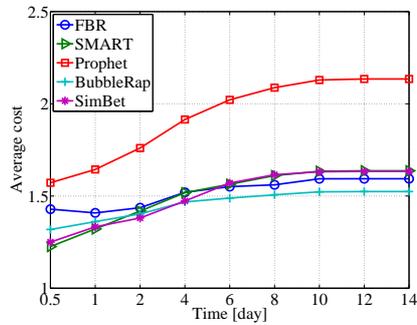
Data routing in DTNs is challenging due to the fact that nodes are constantly moving and the opportunity of communication between node pairs is intermittent. Existing routing strategies encounter the problems of blind spot and dead end, and also lack of efficient implementation in DTNs. In this Chapter, we first investigate the characteristics of DTNs by analyzing three data sets collected from cell-phones, buses and taxis. We reveal the social and mobile features of DTNs: they have community structure and their movement shows locality. Based on these features, we propose the social and mobile aware routing strategy called SMART. In this strategy, a DTN is divided into a number of communities using an adaptive community partitioning algorithms. Two data routing processes are introduced: intra-community communication and inter-community communication. For intra-community communication, a utility function convoluting social similarity

6.5. Summary of Contributions



(a) Delivery ratio

(b) Average delay



(c) Average cost

Figure 6.16: The performance comparison of various strategies on Cabspotting trace

and social centrality with a decay factor is used to choose relay nodes. For inter-community communication, the nodes moving frequently across communities are chosen as relays to carry the data to destination efficiently. It is shown that such routing strategy significantly alleviates the blind spot and dead end problems. It adapts to the community structure by enhancing performance for inter-community communication. We conduct extensive experiments to compare the performance of SMART with other DTN routing strategies. It presents that the proposed routing strategy works well in various DTN traces.

Chapter 7

Discussion and Future Works

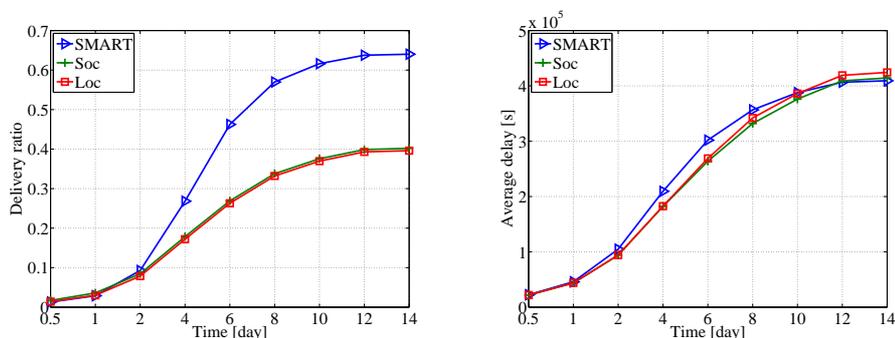
In this chapter, we have a fair comparison of three proposed social-based routing strategies and discuss the future work that can be done in the area of social-based data routing in DTN.

7.1 A Comparison of Three Strategies

We conduct a comprehensive comparison of three proposed social-based routing strategies on three real data traces in terms of delivery ratio, average delay and average cost. Similar as previous evaluation setup, we use HuggleSim for the simulation. 1000 messages are generated with randomly selected sources and destinations. In this group of experiment, each message only keeps one copy of message in the network. We run each simulation 20 times for the result convergence.

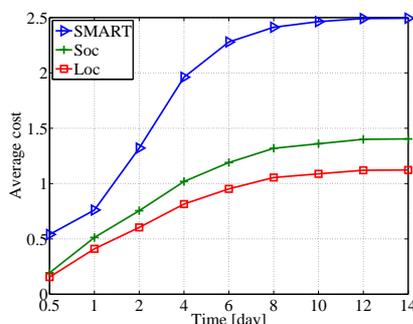
Fig. 7.1 shows the performance of three social-based routing strategies on MIT Reality data trace. The delivery ratio of three different social-based routing strategies is presented in Fig. 7.1a. The location-based routing strategy Loc has similar performance with the encounter-based routing strategies Soc, although Soc has slightly higher delivery ratio compared with Loc. This is consistent to the results when comparing location-based strategy and encounter-based strategy in Chapter 5. The delivery ratio of SMART is 50% higher than that of Soc and Loc. The average delay of three social-based routing schemes is shown in Fig. 7.1b. It suggests SMART has higher average delay than both Soc and Loc, whereas Soc and Loc perform similar in terms of average delay. The average cost of SMART, as shown in Fig. 7.1c, is 40% higher than both Soc and Loc. The cost of Soc is also higher than that of Loc. However, the difference between them is still not significant.

Fig. 7.2 shows the performance of three proposed social-based routing strategies on DieselNet data trace. The delivery ratio is shown in Fig. 7.2a. The community-based strategy SMART outperforms both the location-based social strategy Loc and encounter-based routing strategy Soc over 100%. The performance of Soc and Loc are nearly the same with the delivery ratio of Soc slightly higher. The



(a) Delivery ratio

(b) Average delay



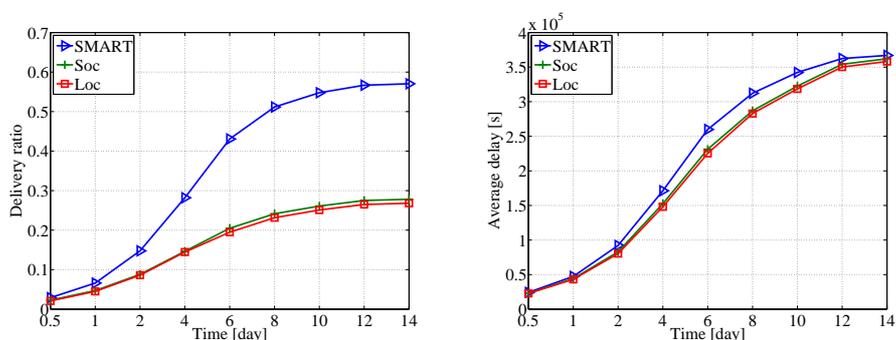
(c) Average cost

Figure 7.1: Performance comparison of three social-based strategies on MIT Reality data trace

average delay, presented in Fig. 7.2b, shows differences among different strategies as well. Specifically, Loc and Soc has similar average delay, which is consistent with the comparison between Loc and Soc in Chapter 5. The average delay of SMART is only 10% higher than that of Loc and Soc. Fig. 7.2c shows the average cost of three social-based routing schemes on DieselNet. It is clear that SMART has much higher cost than both Soc and Loc (around 1 hop longer). In contrast, the cost of Soc and Loc are still similar to each other.

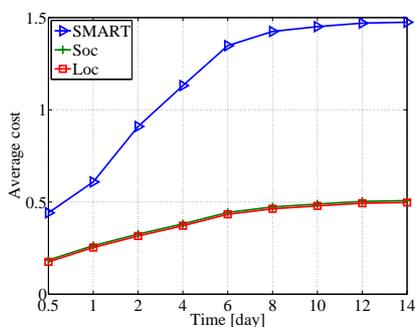
Fig. 7.3 presents the performance of three social-based routing schemes on Cabspotting data trace. The delivery ratio of SMART, as shown in Fig. 7.3a, reaches 0.7 in the end of the experiment, which is 133% higher than that of Loc and Soc. The delivery ratio of Loc and Soc are low and similar to each other. In this experiment, Loc has slightly higher delivery ratio than Soc. Fig. 7.3b shows the average delay of three different routing strategies. In the figure, SMART has higher av-

7.1. A Comparison of Three Strategies



(a) Delivery ratio

(b) Average delay

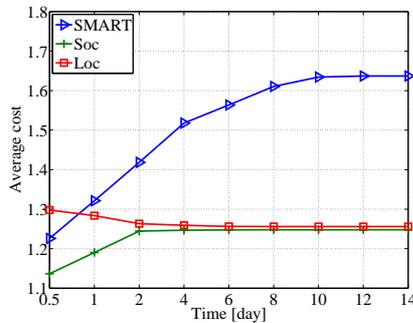
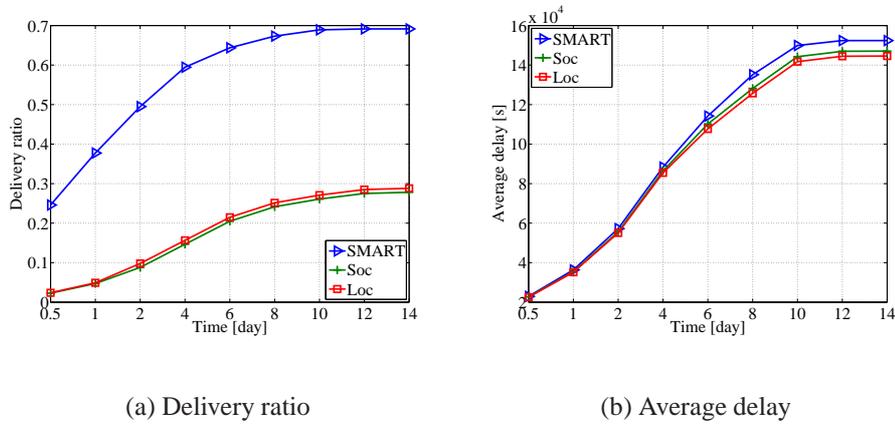


(c) Average cost

Figure 7.2: Performance comparison of three social-based strategies on DieselNet data trace

verage delay, but compared with delivery ratio, the outstanding proportion is much lower. Soc uses slightly higher average delay compared with Loc. The average cost is shown in Fig. 7.3c. Again, Soc and Loc have similar cost after the second day of the simulation. The cost of SMART varies from Soc and Loc significantly. It is about 30% higher than others.

Overall, the community-based routing strategy SMART has much better performance than both location-based routing strategy Loc and encounter-based routing strategy Soc. Whereas, encounter-based routing strategy has similar performance with location-based strategy. This is consistent with the comparison results that we have presented in Chapter 5.



(c) Average cost

Figure 7.3: Performance comparison of three social-based strategies on Cabspotting data trace

7.2 Future Works

In this thesis, we mainly focus on using social information for the enhancement of data routing performance. For the future research, we will target on the following future research directions: resource utilization and allocation, privacy preserving and content centric routing.

One key future research direction for data dissemination in delay tolerant networks is the resource utilization and allocation. In a DTN, especially with human involved, mobile nodes 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. 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. Similarly, if the buffer of a node is full, a node

may first drop the message for those nodes that it is not familiar with. The other one is how to save 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.

Another important future research direction for data dissemination in DTNs is the privacy preserving. Although we consider location-based social information may violate user privacy, there are few works targeting on privacy preserving in terms of data routing in DTNs. Especially, different types of social information also refer to privacy issue. The study of privacy concern in social networks [6, 78, 54, 79] have 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, the future focus is that how to preserve user privacy during the routing process in DTNs from social perspective.

Finally, as the proliferation of content centric network in delay tolerant networks [88, 69, 46, 90], data dissemination needs to adapt to such network structure. In content-centric delay tolerant networks, nodes do not need to care where the data 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 delay tolerant network become to address the following two questions: where is the data and how to cache data? Thus, the routing for content-centric DTNs needs to be in end-to-data diagram. Since the DTN is opportunistically connected, optimal caching mechanism for content-centric DTNs also needs to be carried out.

Chapter 8

Conclusions

In this dissertation, we discuss several social-based routing strategies for DTNs. We pointed out that data dissemination in DTNs is a challenge since the dynamic network topology, limited network information and interrupted connectivity which causes the end-to-end communication to become impractical. There are many of routing strategies proposed to enhance the routing performance in DTNs, which can be categorized into three groups from social perspective: location-based routing, encounter-based routing and community-based routing. One limitation of the existing routing scheme is that they apply sole node or network feature to construct the routing metric, which cannot adapt to various situations in the network. Besides, most utility-based routing strategies meet blind spot and dead end problems during the routing process. Lastly, rare efficient inter-community routing schemes are proposed for community-based routing schemes.

This thesis models social graph based on geographic information and encounter pattern. Based on location-based social information and encounter-based social information, a location-based social routing strategy comprehensively incorporating location information is proposed in Chapter 4 to adapt to different network situations and thus enhance routing performance, and an encounter-based social routing strategy combining multiple encounter-based social properties is devised in Chapter 5 to against sensitive location information and preserve privacy in the network.

Specifically, we discussed the location-based social routing in Chapter 4. Based on the observation that the routing strategy relying on sole aspect of location information cannot adapt to different network situation in DTN, we comprehensively consider two important geographical aspects, geographical distance and mobility pattern, to compose the routing utility metric. The evaluation results show that the proposed comprehensive location-based routing strategy outperforms other location-based routing strategies that relying on sole aspect of location information in terms of delivery ratio, average delay and average cost.

Due to the fact that location-based social information is sensitive and privacy concern to nodes in DTN, while encounter-based social information is less sensitive, we proposed the encounter-based routing strategy in Chapter 5. The design of

the scheme combines multiple social properties by convolution. It presents the feature that the most recent encounters have more effects to the routing utility value. The utility value decays with respect to social centrality and elapsed time which provide different decay speeds for different nodes. The evaluation results show that the proposed encounter-based routing strategy outperforms than other encounter-based routing schemes. Besides, we compared it with the location-based routing strategy, the results suggest that they have very similar performance and therefore encounter-based social information is a good substitute of location-based social information to route data in DTN for the purpose of privacy preserving.

Finally, the thesis carried out a social and mobile aware routing strategy called SMART in Chapter 6 that identified the blind spot and dead end problems and inefficiency of inter-community communication efficiency. SMART introduces a convolutionary routing metric for intra-community routing to address blind spot and dead end problems. It reduces the blind spot and dead end problems below 1%. For efficient inter-community communication, it selects fringe node and utilize utility function similar to intra-community routing to improve the inter-community communication efficiency. Overall, SMART achieves significant improvement compared with other utility and community-based routing strategies. Furthermore, comprehensive comparisons of our proposed three routing strategies have been conducted to outline the benefit of social information applying for data routing in DTNs.

Bibliography

- [1] T. Abdelkader, K. Naik, A. Nayak, N. Goel, and V. Srivastava. Sgbr: A routing protocol for delay tolerant networks using social grouping. *IEEE Trans. Parallel Distrib. Syst.*, 24(12):2472–2481, Dec. 2013.
- [2] B. Adams, D. Phung, and S. Venkatesh. Sensing and using social context. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, 5(2):11, 2008.
- [3] E. Altman, G. Neglia, F. De Pellegrini, and D. Miorandi. Decentralized stochastic control of delay tolerant networks. In *INFOCOM '09*, pages 1134–1142, 2009.
- [4] J. An, Y. Ko, and D. Lee. A social relation aware routing protocol for mobile ad hoc networks. In *IEEE International Conference on Pervasive Computing and Communications, PerCom '09*, pages 1–6, 2009.
- [5] L. Arantes, A. Goldman, and M. V. dos Santos. Using evolving graphs to evaluate dtn routing protocols. *ExtremeCom 2009*, 2009.
- [6] R. Baden, A. Bender, N. Spring, B. Bhattacharjee, and D. Starin. Persona: An online social network with user-defined privacy. In *Proceedings of the ACM SIGCOMM 2009 Conference on Data Communication, SIGCOMM '09*, pages 135–146, New York, NY, USA, 2009. ACM.
- [7] A. Balasubramanian, B. Levine, and A. Venkataramani. Dtn routing as a resource allocation problem. In *Proceedings of the 2007 conference on Applications, technologies, architectures, and protocols for computer communications (SIGCOMM '07)*, pages 373–384, New York, NY, USA, 2007. ACM.
- [8] B. Ball, B. Karrer, and M. Newman. Efficient and principled method for detecting communities in networks. *Physical Review E*, 84(3):036103, 2011.
- [9] M. Behrisch, L. Bieker, J. Erdmann, and D. Krajzewicz. Sumo - simulation of urban mobility: An overview. In *SIMUL 2011, The Third International*

Bibliography

- Conference on Advances in System Simulation*, pages 63–68, Barcelona, Spain, 2011.
- [10] A. Beresford and F. Stajano. Location privacy in pervasive computing. *Pervasive Computing, IEEE*, 2(1):46–55, 2003.
- [11] C. Bettini, X. Wang, and S. Jajodia. Protecting privacy against location-based personal identification. *Secure Data Management*, pages 185–199, 2005.
- [12] J. A. Bitsch Link, N. Viol, A. Goliath, and K. Wehrle. SimBetAge: utilizing temporal changes in social networks for pocket switched networks. In *Proceedings of the 1st ACM workshop on User-provided networking: challenges and opportunities (U-NET '09)*, pages 13–18, New York, NY, USA, 2009. ACM Press.
- [13] E. Bulut and B. Szymanski. Exploiting friendship relations for efficient routing in mobile social networks. *IEEE Transactions on Parallel and Distributed Systems*, PP(99):1, 2012.
- [14] E. Bulut, Z. Wang, and B. Szymanski. Cost-effective multiperiod spraying for routing in delay-tolerant networks. *IEEE/ACM Transactions on Networking*, 18(5):1530–1543, 2010.
- [15] J. Burgess, B. Gallagher, D. Jensen, and B. Levine. Maxprop: Routing for vehicle-based disruption-tolerant networks. In *Proceedings of 25th IEEE International Conference on Computer Communications (INFOCOM '06)*, pages 1–11, 2006.
- [16] J. Burgess, B. N. Levine, R. Mahajan, J. Zahorjan, A. Balasubramanian, A. Venkataramani, Y. Zhou, B. Croft, N. Banerjee, M. Corner, and D. Towsley. CRAWDAD data set umass/diesel (v. 2008-09-14). <http://crawdad.cs.dartmouth.edu/umass/diesel>, 2008.
- [17] S. Burleigh and K. Scott. Bundle protocol specification. *IETF Request for Comments RFC*, 5050, 2007.
- [18] B. Burns, O. Brock, and B. Levine. Mv routing and capacity building in disruption tolerant networks. In *Proceedings of 24th IEEE International Conference on Computer Communications (INFOCOM '05)*., pages 398 – 408, 2005.

-
- [19] V. Cerf, S. Burleigh, A. Hooke, L. Torgerson, R. Durst, K. Scott, K. Fall, and H. Weiss. Delay-tolerant networking architecture. *RFC4838*, April, 2007.
- [20] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Impact of human mobility on opportunistic forwarding algorithms. *IEEE Transactions on Mobile Computing*, 6(6):606–620, 2007.
- [21] P.-C. Cheng, K. C. Lee, M. Gerla, and J. Härri. Geodtn+ nav: geographic dtn routing with navigator prediction for urban vehicular environments. *Mobile Networks and Applications*, 15(1):61–82, 2010.
- [22] J. Coleman. *Foundations of Social Theory*. Belknap Press, 1990.
- [23] J. S. Coleman. Social capital in the creation of human capital. *American Journal of Sociology*, 94:S95–S120, 1988.
- [24] E. M. Daly and M. Haahr. Social network analysis for routing in disconnected delay-tolerant manets. In *Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing (MobiHoc '07)*, pages 32–40, New York, NY, USA, 2007. ACM.
- [25] N. Eagle and A. S. Pentland. CRAWDAD data set mit/reality (v. 2005-07-01). <http://crawdad.cs.dartmouth.edu/mit/reality>, 2005.
- [26] F. Fabbri and R. Verdone. A sociability-based routing scheme for delay-tolerant networks. *EURASIP Journal on Wireless Communications and Networking*, 2011:1, 2011.
- [27] K. Fall. A delay-tolerant network architecture for challenged internets. In *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications (SIGCOMM '03)*, pages 27–34, New York, NY, USA, 2003. ACM.
- [28] J. Fleming. *Magnets and electric currents: An elementary treatise for the use of electrical artisans and science teachers*. E. & F.N. Spon, limited, 1902.
- [29] S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3&C5):75 – 174, 2010.
- [30] L. Freeman. Centrality in social networks: Conceptual clarification. *Social Networks*, 1(3):215–239, 1979.

Bibliography

- [31] L. C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41, 1977.
- [32] W. Gao and G. Cao. On exploiting transient contact patterns for data forwarding in delay tolerant networks. In *18th IEEE International Conference on Network Protocols (ICNP '10)*, pages 193–202, 2010.
- [33] W. Gao, Q. Li, B. Zhao, and G. Cao. Multicasting in delay tolerant networks: a social network perspective. In *Proceedings of the tenth ACM international symposium on Mobile ad hoc networking and computing (MobiHoc '09)*, pages 299–308, New York, NY, USA, 2009. ACM.
- [34] N. Glance, D. Snowdon, and J.-L. Meunier. Pollen: Using people as a communication medium. *Comput. Netw.*, 35(4):429–442, Mar. 2001.
- [35] M. S. Granovetter. The strength of weak ties. *American journal of sociology*, pages 1360–1380, 1973.
- [36] M. Grossglauser and M. Vetterli. Locating mobile nodes with ease: Learning efficient routes from encounter histories alone. *IEEE/ACM Trans. Netw.*, 14(3):457–469, June 2006.
- [37] S. Guo, M. H. Falaki, E. A. Oliver, S. Ur Rahman, A. Seth, M. A. Zaharia, U. Ismail, and S. Keshav. Design and implementation of the kiosknet system. In *International Conference on Information and Communication Technologies and Development (ICTD '07)*, pages 1–10, 2007.
- [38] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot. Pocket switched networks and human mobility in conference environments. In *Proceedings of the 2005 ACM SIGCOMM Workshop on Delay-tolerant Networking*, WDTN '05, pages 244–251, New York, NY, USA, 2005. ACM.
- [39] P. Hui and J. Crowcroft. How small labels create big improvements. In *Fifth Annual IEEE International Conference on Pervasive Computing and Communications Workshops, 2007 (PerCom Workshops '07)*, pages 65–70, 2007.
- [40] P. Hui, J. Crowcroft, and E. Yoneki. Bubble rap: social-based forwarding in delay tolerant networks. In *Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing (MobiHoc '08)*, pages 241–250, New York, NY, USA, 2008. ACM.

-
- [41] K. Jahanbakhsh, G. C. Shojha, and V. King. Social-greedy: a socially-based greedy routing algorithm for delay tolerant networks. In *Proceedings of the Second International Workshop on Mobile Opportunistic Networking (MobiOpp '10)*, pages 159–162, New York, NY, USA, 2010. ACM.
- [42] S. Jain, K. Fall, and R. Patra. Routing in a delay tolerant network. In *Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications (SIGCOMM '04)*, pages 145–158, New York, NY, USA, 2004. ACM.
- [43] P. Juang, H. Oki, Y. Wang, M. Martonosi, L. S. Peh, and D. Rubenstein. Energy-efficient computing for wildlife tracking: Design tradeoffs and early experiences with zebranet. *SIGARCH Comput. Archit. News*, 30(5):96–107, Oct. 2002.
- [44] H. Kang and D. Kim. Vector routing for delay tolerant networks. In *IEEE 68th Vehicular Technology Conference (VTC '08)*, pages 1–5, 2008.
- [45] B. Karp and H. T. Kung. Gpsr: greedy perimeter stateless routing for wireless networks. *MobiCom '00*, pages 243–254, New York, NY, USA, 2000. ACM.
- [46] V. Kawadia, N. Riga, J. Opper, and D. Sampath. Slinky: An adaptive protocol for content access in disruption-tolerant ad hoc networks. In *ACM MobiHoc 2011 International Workshop on Tactical Mobile Ad Hoc Networking*, 2011.
- [47] J. Krumm. A survey of computational location privacy. *Personal Ubiquitous Comput.*, 13(6):391–399, Aug. 2009.
- [48] J. LeBrun, C.-N. Chuah, D. Ghosal, and M. Zhang. Knowledge-based opportunistic forwarding in vehicular wireless ad hoc networks. In *IEEE 61st Vehicular technology conference (VTC '05)*, volume 4, pages 2289–2293. IEEE, 2005.
- [49] J. Leguay, T. Friedman, and V. Conan. Evaluating mobility pattern space routing for dtns. In *Proceedings of 25th IEEE International Conference on Computer Communications (INFOCOM '06)*, pages 1–10, 2006.
- [50] I. Leontiadis and C. Mascolo. Geopps: Geographical opportunistic routing for vehicular networks. In *IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, (WoWMoM '07)*, pages 1–6. IEEE, 2007.

Bibliography

- [51] F. Li and J. Wu. Localcom: a community-based epidemic forwarding scheme in disruption-tolerant networks. In *Proceedings of the 6th Annual IEEE communications society conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON '09)*, pages 574–582, Piscataway, NJ, USA, 2009. IEEE Press.
- [52] Z. Li and H. Shen. Utility-based distributed routing in intermittently connected networks. In *37th International Conference on Parallel Processing, 2008 (ICPP '08)*, pages 604–611. IEEE, 2008.
- [53] Z. Li and H. Shen. Sedum: Exploiting social networks in utility-based distributed routing for dtns. *IEEE Trans. Comput.*, 62(1):83–97, Jan. 2013.
- [54] X. Liang, X. Li, K. Zhang, R. Lu, X. Lin, and X. Shen. Fully anonymous profile matching in mobile social networks. *IEEE J Sel Areas Commun (to appear)*, 2013.
- [55] K. C.-j. Lin, C.-w. Chen, and C.-F. Chou. Preference-aware content dissemination in opportunistic mobile social networks. In *Proceedings of the 31st Annual IEEE International Conference on Computer Communications (INFOCOM '12)*, pages 1960–1968, 2012.
- [56] A. Lindgren, A. Doria, and O. Schelén. Probabilistic Routing in Intermittently Connected Networks. volume 3126 of *Lecture Notes in Computer Science*, pages 239–254. Springer-Verlag GmbH, 2004.
- [57] C. Lochert, M. Mauve, H. Füßler, and H. Hartenstein. Geographic routing in city scenarios. *ACM SIGMOBILE Mobile Computing and Communications Review*, 9(1):69–72, 2005.
- [58] G. Luo, J. Zhang, H. Huang, K. Qin, and H. Sun. Exploiting intercontact time for routing in delay tolerant networks. *Transactions on Emerging Telecommunications Technologies*, 2012.
- [59] T. Matsuda and T. Takine. (p,q)-epidemic routing for sparsely populated mobile ad hoc networks. *IEEE Journal on Selected Areas in Communications*, 26(5):783–793, 2008.
- [60] A. Mei, G. Morabito, P. Santi, and J. Stefa. Social-aware stateless forwarding in pocket switched networks. In *INFOCOM '11*, pages 251–255, 2011.

-
- [61] A. G. Miklas, K. K. Gollu, K. K. Chan, S. Saroiu, K. P. Gummadi, and E. De Lara. Exploiting social interactions in mobile systems. In *UbiComp 2007: Ubiquitous Computing*, pages 409–428. Springer, 2007.
- [62] M. Motani, V. Srinivasan, and P. S. Nuggehalli. Peoplenet: engineering a wireless virtual social network. In *Proceedings of the 11th annual international conference on Mobile computing and networking (MobiCom '05)*, pages 243–257, New York, NY, USA, 2005. ACM.
- [63] A. Mtibaa, M. May, C. Diot, and M. Ammar. Peoplerank: social opportunistic forwarding. In *Proceedings of the 29th conference on Information communications (INFOCOM '10)*, pages 111–115, Piscataway, NJ, USA, 2010. IEEE Press.
- [64] P. Mundur, M. Seligman, and J. N. Lee. Immunity-based epidemic routing in intermittent networks. In *5th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, SECON '08*, pages 609–611, 2008.
- [65] M. Musolesi and C. Mascolo. Car: Context-aware adaptive routing for delay-tolerant mobile networks. *IEEE Transactions on Mobile Computing*, 8(2):246–260, 2009.
- [66] S. Naidu, S. Chintada, M. Sen, and S. Raghavan. Challenges in deploying a delay tolerant network. In *Proceedings of the third ACM workshop on Challenged networks*, pages 65–72. ACM, 2008.
- [67] M. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69:026113, 2004.
- [68] M. E. Newman. Fast algorithm for detecting community structure in networks. *Physical review E*, 69(6):066133, 2004.
- [69] A. D. Nguyen, P. Senac, V. Ramiro, and M. Diaz. Pervasive intelligent routing in content centric delay tolerant networks. In *IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing (DASC '11)*, pages 178–185, 2011.
- [70] N. P. Nguyen, T. N. Dinh, Y. Xuan, and M. T. Thai. Adaptive algorithms for detecting community structure in dynamic social networks. In *Proceedings*

Bibliography

of 30th Annual IEEE International Conference on Computer Communications (INFOCOM '11), pages 2282–2290, 2011.

- [71] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435:814, 2005.
- [72] I. Parris and T. Henderson. The impact of location privacy on opportunistic networks. WoWMoM '11, pages 1–6, 2011.
- [73] F. D. Pellegrini, D. Miorandi, I. Carreras, and I. Chlamtac. A graph-based model for disconnected ad hoc networks. In *INFOCOM*, pages 373–381. IEEE, 2007.
- [74] A. Pentland, R. Fletcher, and A. Hasson. Daknet: Rethinking connectivity in developing nations. *Computer*, 37(1):78–83, 2004.
- [75] C. E. Perkins and P. Bhagwat. Highly dynamic destination-sequenced distance-vector routing (dsv) for mobile computers. In *ACM SIGCOMM Computer Communication Review*, volume 24, pages 234–244. ACM, 1994.
- [76] C. E. Perkins and E. M. Royer. Ad-hoc on-demand distance vector routing. In *Second IEEE Workshop on Mobile Computing Systems and Applications (WMCSA '99)*, pages 90–100. IEEE, 1999.
- [77] M. Piorkowski, N. Sarafijanovic-Djukic, and M. Grossglauser. CRAWDAD data set epfl/mobility (v. 2009-02-24). <http://crawdad.cs.dartmouth.edu/epfl/mobility>, 2009.
- [78] K. P. Puttaswamy, A. Sala, and B. Y. Zhao. Starclique: Guaranteeing user privacy in social networks against intersection attacks. In *Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies, CoNEXT '09*, pages 157–168, New York, NY, USA, 2009. ACM.
- [79] K. P. Puttaswamy and B. Y. Zhao. Preserving privacy in location-based mobile social applications. In *Proceedings of the Eleventh Workshop on Mobile Computing Systems & Applications*, pages 1–6. ACM, 2010.
- [80] N. Ristanovic, G. Theodorakopoulos, and J.-Y. Le Boudec. Traps and Pitfalls of Using Contact Traces in Performance Studies of Opportunistic Networks. *INFOCOM '12*, 2012.

-
- [81] A. Seth, D. Kroeker, M. Zaharia, S. Guo, and S. Keshav. Low-cost communication for rural internet kiosks using mechanical backhaul. In *Proceedings of the 12th Annual International Conference on Mobile Computing and Networking (MobiCom '06)*, pages 334–345, New York, NY, USA, 2006. ACM.
- [82] T. Small and Z. J. Haas. The shared wireless infostation model: A new ad hoc networking paradigm (or where there is a whale, there is a way). In *Proceedings of the 4th ACM International Symposium on Mobile Ad Hoc Networking & Computing (MobiHoc '03)*, pages 233–244, New York, NY, USA, 2003. ACM.
- [83] T. Spyropoulos, K. Psounis, and C. S. Raghavendra. Spray and wait: an efficient routing scheme for intermittently connected mobile networks. WDTN '05, pages 252–259, New York, NY, USA, 2005. ACM.
- [84] T. Spyropoulos, K. Psounis, and C. S. Raghavendra. Spray and focus: Efficient mobility-assisted routing for heterogeneous and correlated mobility. In *Fifth Annual IEEE International Conference on Pervasive Computing and Communications Workshops, (PerCom Workshops' 07)*, pages 79–85. IEEE, 2007.
- [85] T. Spyropoulos, K. Psounis, and C. S. Raghavendra. Efficient routing in intermittently connected mobile networks: the single-copy case. *IEEE/ACM Trans. Netw.*, 16(1):63–76, 2008.
- [86] J. Su, J. Scott, P. Hui, J. Crowcroft, E. De Lara, C. Diot, A. Goel, M. H. Lim, and E. Upton. Huggle: seamless networking for mobile applications. In *Proceedings of the 9th international conference on Ubiquitous computing (UbiComp '07)*, pages 391–408, Berlin, Heidelberg, 2007. Springer-Verlag.
- [87] N. Thompson, R. Crepaldi, and R. Kravets. Locus: a location-based data overlay for disruption-tolerant networks. CHANTS '10, pages 47–54. ACM Press, 2010.
- [88] G. Tyson, J. Bigham, and E. Bodanese. Towards an information-centric delay-tolerant network. In *Proc. IEEE INFOCOM Workshop on Emerging Design Choices in Name-Oriented Networking (NOMEN)*, 2013.
- [89] A. Vahdat, D. Becker, et al. Epidemic routing for partially connected ad hoc networks. Technical report, CS-200006, Duke University, 2000.

Bibliography

- [90] L. Wang, R. Wakikawa, R. Kuntz, R. Vuyyuru, and L. Zhang. Data naming in vehicle-to-vehicle communications. In *2012 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 328–333. IEEE, 2012.
- [91] D. J. Watts and S. H. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684):440–442, 1998.
- [92] C. Wilson, B. Boe, A. Sala, K. P. Puttaswamy, and B. Y. Zhao. User interactions in social networks and their implications. EuroSys ’09, pages 205–218, New York, NY, USA, 2009. ACM.
- [93] H. Wu, R. Fujimoto, R. Guensler, and M. Hunter. Mddv: a mobility-centric data dissemination algorithm for vehicular networks. In *Proceedings of the 1st ACM international workshop on Vehicular ad hoc networks*, pages 47–56. ACM, 2004.
- [94] J. Wu and 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.
- [95] J. Wu, M. Xiao, and L. Huang. Homing spread: Community home-based multi-copy routing in mobile social networks. In *INFOCOM*, pages 2319–2327. IEEE, 2013.
- [96] F. Xia, L. Liu, J. Li, J. Ma, and A. Vasilakos. Socially aware networking: A survey. *IEEE Systems Journal*, PP(99):1–18, 2013.
- [97] M. Xiao, J. Wu, and L. Huang. Community-Aware Opportunistic Routing in Mobile Social Networks. *IEEE Transactions on Computers*, PP(99):1, 2013.
- [98] H. Zang and J. Bolot. Anonymization of location data does not work: a large-scale measurement study. MobiCom ’11, pages 145–156, New York, NY, USA, 2011. ACM.
- [99] M. Zarafshan-Araki and K.-W. Chin. Trainnet: A transport system for delivering non real-time data. *Comput. Commun.*, 33(15):1850–1863, Sept. 2010.
- [100] Y. Zhang and J. Zhao. Social network analysis on data diffusion in delay tolerant networks. In *Proceedings of the tenth ACM international symposium*

on Mobile ad hoc networking and computing (MobiHoc '09), pages 345–346. ACM, 2009.

- [101] K. Zhu, P. Hui, Y. Chen, X. Fu, and W. Li. Exploring user social behaviors in mobile social applications. In *Proceedings of the 4th Workshop on Social Network Systems, SNS '11*, pages 3:1–3:6, New York, NY, USA, 2011. ACM.
- [102] K. Zhu, W. Li, and X. Fu. Smart: A social and mobile aware routing strategy for disruption tolerant networks. *IEEE Transactions on Vehicular Technology (to appear)*, 2013.
- [103] K. Zhu, W. Li, and X. Fu. Rethinking routing information in mobile social networks: Location-based or social-based? *Elsevier Computer Communications (to appear)*, 2014.

Curriculum Vitae

Personal Data

Konglin Zhu
Albrecht-Thaer-Weg 10b
37075, Goettingen Germany
Telephone: +4955139172047
Email: zhu@cs.uni-goettingen.de

Research Interest

Mobile social networks
Online social networks
Routing in Delay Tolerant Networks

Education

2010-present Ph.D. (Computer Science), University of Goettingen
2008-2009 M.S. (Computer Science), University of California Los Angeles
2006-2008 B.S. (Computer Science), Southern Polytechnic State University
2004-2008 B.S. (Computer Science), North China University of Technology

Publications

1. Dongsheng Wei, Konglin Zhu, Xin Wang, Fairness-Aware Cooperative Caching Scheme for Mobile Social Networks, *accepted in IEEE International Conference on Communications (ICC) 2014*.
2. Konglin Zhu, Wenzhong Li, Xiaoming Fu, Rethinking Routing Information in Mobile Social Networks: Location-based or Social-based? *accepted in Computer Communications*. January 2014.
3. Konglin Zhu, Wenzhong Li, Xiaoming Fu, SMART: A Social and Mobile Aware Routing Strategy for Disruption Tolerant Networks, *accepted in IEEE Transactions on Vehicular Technology*, October 2013.
4. Konglin Zhu, Wenzhong Li, Xiaoming Fu, Jan Nagler, How Do Online Social Networks Grow? *PLOS ONE* (under second review).
5. Konglin Zhu Wenzhong Li and Xiaoming Fu, Modeling population growth in online social networks, *Springer Complex Adaptive Systems Modeling*, 2013.
6. Sufian Hameed, Alexander Wolf, Konglin Zhu, Xiaoming Fu, Evaluation of Human Altruism Using a DTN- based Mobile Social Network Application, *5th workshop on Digital Social Networks (DSN 2012) in conjunction with Gesellschaft fuer Informatik (GI 2012)*, Braunschweig, Germany, September 2012.
7. Konglin Zhu, Wenzhong Li , Xiaoming Fu, Exploring Regional and Global Population Growth in Online Social Networks (extended abstract), *Praxis der Netzwerkforschung 2012 (PdN 2012)*, Frankfurt am Main, Germany, May 2012.
8. Narisu Tao, KonglinZhu, Xiaoming Fu, Thumb: A Real-Time Resource Information Sharing Application over Mobile Phones, *IEEE International Conference on Pervasive Computing and Communication (PerCom) 2012*, demo session, Lugano, Switzerland, March 2012.
9. Konglin Zhu, Kelvin (Biao) Zhou, Xiaoming Fu, and Mario Gerla, Geo-Assisted Multicast Inter-Domain Routing (GMIDR) Protocol for MANETs,

Publications

- IEEE International Conference on Communications (ICC 2011)*, Kyoto, Japan, IEEE, June 2011.
10. Konglin Zhu, Pan Hui, Yang Chen, Xiaoming Fu, Wenzhong Li, Exploring User Social Behaviors in Mobile Social Applications, *4th ACM Workshop on Social Network Systems (SNS 2011)*, in conjunction with *ACM EuroSys 2011*, Salzburg, Austria, April 2011.
 11. Kelvin Biao Zhou, Abhishek Tiwari, Konglin Zhu, You Lu, Mario Gerla, Anurag Ganguli, Bao-hong Shen, David Krzysiak, Geo-Based Inter-Domain Routing (GIDR) Protocol for MANETs, *IEEE Military Communications Conferences (MILCOM 2009)*, Boston, MA, October 2009.

January 23rd, 2014

Goettingen