

Elsevier required licence: © <2019>.
This manuscript version is made available under the CC-BY-NC-ND 4.0 license

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

The definitive publisher version is available online at

<https://www.sciencedirect.com/science/article/pii/S0020025518308545?via%3Dihub>

On Exploiting Priority Relation Graph for Reliable Multi-Path Communication in Mobile Social Networks

Limei Lin^{a,b}, Li Xu^{a,b,*}, Yanze Huang^{a,b,c}, Yang Xiang^d, Xiangjian He^e

^aSchool of Mathematics and Computer Science, Fujian Normal University, Fuzhou, Fujian, 350007, P.R. China

^bKey Laboratory of Network Security and Cryptology, Fujian Normal University, Fuzhou, Fujian, 350007, P.R. China

^cSchool of Mathematics and Physics, Fujian University of Technology, Fuzhou, Fujian, 350118, P.R. China

^dSchool of Software and Electrical Engineering, Swinburne University of Technology, Hawthorn, 3122, V.I.C. Australia

^eSchool of Computing and Communications, University of Technology Sydney, Sydney, 2007, N.S.W. Australia

Abstract: A mobile social network (MSN) consists of certain amount of mobile users with social characteristics, and it provides data delivery concerning social relationships between mobile users. In MSN, ordinary people contact each other more frequently if they have more social features in common. In this paper, we apply a new topology structure–priority relation graph (PRG) to evaluate the data delivery routing in MSN on the system-level. By using the natural order of nodes’ representation, the diameter, the regular degree and the multi-path technology, we determine the priority relation graph-based social feature routing (PRG-SFR) algorithm to find disjointed multi-paths in MSN. Here, the multi-path technology can be exploited for ensuring that, between each pair of sender and receiver, the important information can be delivered through a highly reliable path. Then we calculate the tolerant ability of ‘faults’ and estimate the availability of MSN on the theoretical level. Finally, we analyze the efficiency of PRG-SFR algorithm from the numerical standpoint in terms of fault tolerance, forwarding number, transmission time and delivery rate. Moreover, we make comparisons between PRG-SFR algorithm and certain state-of-the-art technologies.

Keywords: Reliable communication, Priority relation graph, Multi-path technology, Social features, Mobile social networks.

1 Introduction

Mobile social networks (MSNs) [14, 15] combine social networks [29, 35] with mobile communication networks [3, 6] (see Fig. 1), and they can provide the data delivery involving social relationships among mobile users.

To obtain an efficient data delivery routing, a mobile user needs to find a preferential data carrier in his/her contacts in order to improve the routing performance. Many stud-

^{0*} Corresponding author: xuli@fjnu.edu.cn

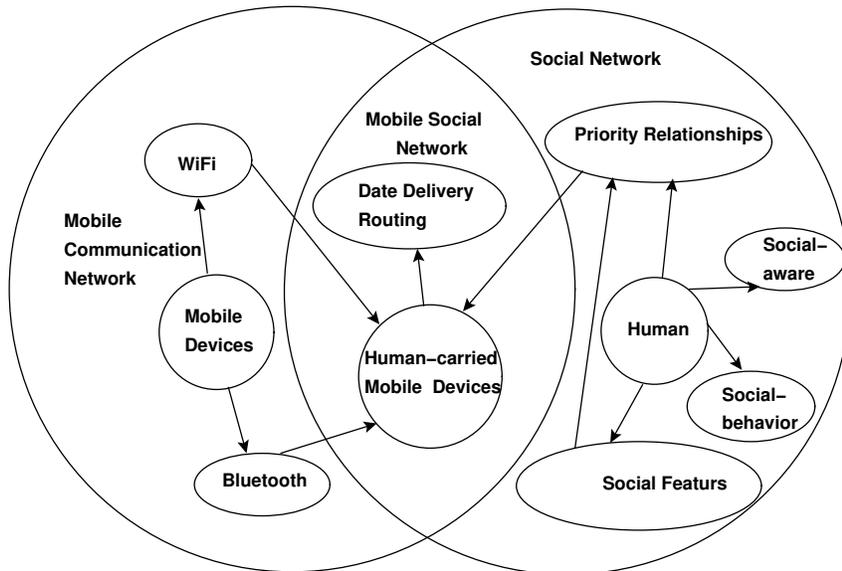


Figure 1: The mobile social networks.

ies [5, 8, 9, 16, 17, 38] have focused the choice of preferential data carriers to address the data delivery communication problem, and they all use social-aware factors and social-behaviors. We also focus on the routing protocols which take into account the degree of similarities in forwarding messages. Gao *et al.* [10] proposed a mechanism for multicasting in delay tolerant networks from the perspective of a social network. Li *et al.* [19] provided a mechanism for routing in socially selfish delay tolerant networks and Lu *et al.* [25] further established a social-based privacy preserving protocol for vehicular DTNs. Altman *et al.* [1] provided a combined optimal control mechanism of activation and transmission in DTNs. Talipov *et al.* [32] proposed the discover-predict-deliver for smartphone DTNs. Venkatramanan *et al.* [33] considered a co-evolution of content spread and popularity in mobile opportunistic networks. T. T. Mapoka *et al.* [26] proposed a novel GKM protocol for multiple multicast groups, and it is named as slot based multiple group key management scheme. Athanasopoulou *et al.* [2] developed a new back-pressure-based packet-by-packet adaptive routing in communication networks.

All of these above studies consider state information, which is hard to obtain without a long-term collection process. To avoid the shortcoming, Wu *et al.* [36, 37] chose the preferential communication carriers based on the social feature distances among mobile users, and provided a mechanism for multi-path routing based on the hypercube in a DTN. However, they only consider the special case that two mobile users have a data delivery relationship when they differ in only one feature value. They do not consider the wide contact demands among all users.

To further the above approaches, a *priority relationship* [20] between two users was defined if they differ in no more than k feature values. We [21] have proposed a reliable communication in mobile social networks based on trustworthiness-hypercube. In this paper, we observe that the PRG has many structural properties, including the *nature order*, *diameter*, *regular degree* and *multi-path technology*. Then we propose an efficient

and systematic *priority relation graph-based social feature routing* (PRG-SFR) algorithm to find disjointed multi-paths in MSN by using these four structural properties. Exploiting the structural properties of PRG can improve the performance of data forwarding in terms of fault tolerance, forwarding number, transmission time and delivery rate.

1. The *natural order* can be used to provide the disjointed communication paths between any two users.
2. Smaller *diameter* of PRG implies that there exists a shorter communication connecting two nodes.
3. The *regular degree* of PRG ensures the fault tolerant with two basic functionality approaches. The first one is whether the network logically contains a certain topological structure. The second one considers a multiprocessor system as functional if a fault-free communication path exists between any two fault-free nodes;
4. The *multi-path technology* ensures the important information to be delivered through a highly reliable path between each pair of sender and receiver.

The PRG-SFR algorithm is more general due to the fact that the data is forwarded not only along the first-priority relationship but also along the appropriate priority relationship. The PRG-SFR algorithm has the competitive properties in terms of fault tolerance, forwarding number, transmission time and delivery rate.

Organization. The remainder of this paper is organized as follows. Section 2 shows the preliminary work. Section 3 gives some available and symmetric properties of PRG, and then proposes the multi-path reliable communication algorithm. Section 4 formally analyzes the proposed routing algorithm in terms of fault tolerance, forwarding number, transmission time and delivery rate on theoretical level. Section 5 gives the numerical analysis. Section 6 extends the PRG-based social feature routing to the general case. Our work is concluded in Section 7.

2 Preliminaries

In this section, some terminologies and notations in combinatorial network theory are first presented in Subsection 2.1. Subsection 2.2 briefly introduces the basic idea of data delivery in MSN. Subsection 2.3 provides the model assumption.

2.1 Combinatorial Network

We usually use $G = (V(G), E(G))$ to represent a graph, where $V(G)$ is the *node-set* of G and $E(G)$ is the *edge-set* of G . Let $|V(G)|$ be the size of the node-set and $|E(G)|$ be the size of the edge-set. Two nodes corresponding to an edge are called the *end nodes* of the edge. The *distance* between node x and node y , denoted by $d(x, y)$, is the length of the shortest path between x and y . The *diameter* of G is defined as the maximal distance between any two nodes in G . The *degree* of a node u in G , denoted by $d(u)$, is the number

of edges of G incident with u . For any node x , the set of all nodes which are adjacent to x is denoted by $N(x)$. The set $N(v)$ is called the *neighborhood* of node v in G , and $u \in N(v)$ is said to be a neighbor of v in G . A *path* P is a sequence of distinct nodes with an edge joining consecutive nodes. If the length of a path (resp., cycle) of G is n , then the path (resp., cycle) is an n -*path* (resp., n -*cycle*). For notations and terminologies not defined here, we follow [39].

2.2 Mobile Social Networks

Since people carry mobile devices, the information about social behavior and structure can be acquired as a key to design and provide efficient data communication services (e.g., routing and data dissemination, bandwidth and computing capacity allocation, and storage) [3]. Firstly, Gao *et al.* [11] solved the social-aware multicast in disruption-tolerant networks. Secondly, Bulut and Szymanski [4] exploited friendship relations for efficient routing in MSNs. Li *et al.* [18] provided an efficient and scalable data center multicast routing technology. Liu and Wu [24] established a practical routing in a cyclic Mobispace. Musolesi and Mascolo [28] proposed a context-aware adaptive routing for delay-tolerant mobile networks. J. Ryu *et al.* [30] gave a timescale decoupled routing and rate control in intermittently connected networks. Physically, MSN is composed of human-carried mobile devices that employ short-range wireless communication technologies (i.e., WiFi, BlueTooth and GPRS) to communicate with each other. In addition, users deliver messages via store-carry-forwarding which are guided by users' relationships. Functionally, MSN allows users to share information in a completely decentralized manner that doesn't require the aid of wireless communication infrastructure. In fact, the communications between users of MSN rely on the corresponding relation graph. A high-level illustration of this architecture is presented in Fig. 2.

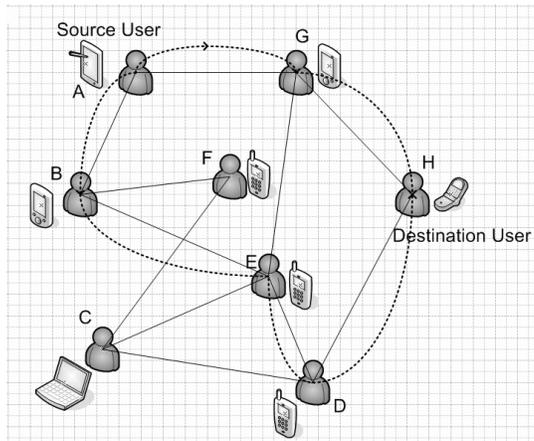


Figure 2: A high-level illustration of MSN.

Fig. 2 shows MSN composed of eight users, each of whom carries a mobile device. Eleven solid lines form the relation graph and six dotted lines represent two data forwarding routings (paths $A \rightarrow G \rightarrow H$ and $A \rightarrow B \rightarrow E \rightarrow D \rightarrow H$). The source user can

deliver the information to other users in their contacts through the mobile devices. In Fig. 2, the bold lines are given based on the solid lines. Hence, the relation graph ensures effective data delivery routing. The latency of the data delivery routing is decided by the contact frequency. Mei *et al.* [27] suggested that people contacted with each other more frequently if they had more social features in common.

2.3 Model Assumption

To achieve our goal of developing an efficient multi-path communication algorithm based on a priority relation graph in MSN, we introduce some assumption metrics related to the internal social features.

We use a high dimension feature profile to characterize individuals. However, usually only a small subset of features (key features) is important, which are extracted from feature data mining [23]. Obviously, the key features of a special mobile social network are independent of each other. Assume that in MSN, there are exactly n key internal social features. For feature i ($1 \leq i \leq n$), there are m_i distinct values. For simplicity, the set $M_i = \{0, 1, \dots, m_i - 1\}$ is used to represent the values of feature i . The notation $u_1 u_2 \dots u_n$ is applied to represent a user u with n key features and $u_i \in M_i$. The *feature distance* of two users is the number of features where these two users have different feature values. An example is given to show how to code the features (see Table 1).

Table 1: An illustration of coding of features

Features	Identity		Profession			
Feature values	Teacher	Student	Computer	Math	Physics	Biology
Feature Codes	0	1	0	1	2	3

In Table 1, the identity feature is labeled by 1 and the profession feature is label by 2. The identity feature has two values: teacher and student, and $M_1 = \{0, 1\}$. The profession feature has four values: computer, math, physics and biology, and $M_2 = \{0, 1, 2, 3\}$. A teacher with computer profession is coded as 00 and a student with math profession is coded as 11. The feature distance of them is 2.

Motivated by the two real traces: Infocom 2006 [31] and MIT reality mining [7], we group individuals according to the set of key features, where individuals with the same values of all key features are put into one group. The routing packet is transmitted from group to group until it reaches the destination group. The notion of group has some optimal properties, such as closeness centrality, betweenness centrality [12] and degree centrality [15]. When the difference of social features between two individuals increases in the Infocom 2006 trace [31] and the MIT reality mining data [7], the total number of mutual contacts decreases. From Infocom 2006 [31] and MIT reality mining [7], it is discovered that users can contact each other when their feature distance is less than \sqrt{n} , where n denotes the number of key social features.

The efficient routing algorithm is embodied in the following four performance metrics (see Table 2):

Table 2: Routing performance metrics

Fault tolerance (FT)	The minimal number of groups which must be deleted to destroy all paths between a pair of groups
Forwarding number (FN)	The average number of forwarding of each packet before the destination is reached
Transmission time (T_l)	The average duration between the generation and the arrival of a packet with feature distance l
Delivery rate (DR_t)	The probability of successful delivery within time t

Therefore, the efficient routing entails a maximal fault tolerance, an acceptable forwarding number, a low transmission time and a high delivery rate.

3 Priority Relation Graph-based Multi-path Reliable Communication

We presented the concepts [20] of *priority relationship* between mobile groups and *priority relation graph* (PRG) to describe MSN more comprehensively. In this section, we show some available properties about the priority relationship and PRG to propose efficient multi-path data delivery routing.

Definition 1 [20] *Let n be the number of key features in a mobile social network and let k be an integer with $k < \sqrt{n}$. Two users in MSN have a priority relationship if their features distance is no more than k . If the feature distance between two users is one, then their relationship is called a first-priority relationship (FPR).*

Property 1 *The data delivery based on the priority relationship can reduce the forwarding time to improve the routing performance.*

It is easy to see that, as long as the feature distance is no more than k , two users are preferential contacted and they preferentially deliver data/messages to each other. Clearly, with different feature distance, the levels of priority relationship between two users are also different. The smaller the feature distance is, the more preferential that the relationship is between two users. The difference causes the difference of the decision time for data delivery. This means that the longer the feature distance k is, the greater the decision time is. Therefore, the latency of data delivery routing is decided by feature distance. Since the number k is far less than n , the forwarding time between two users with a priority relationship is less. This implies that the data delivery based on the priority relationship can reduce the forwarding time to improve the routing performance.

In Subsection 2.3, we define users with the same feature values of all key features as a group. Clearly, there is a many-to-one mapping between the M-space and the FPR-space [20] (see Fig. 3).

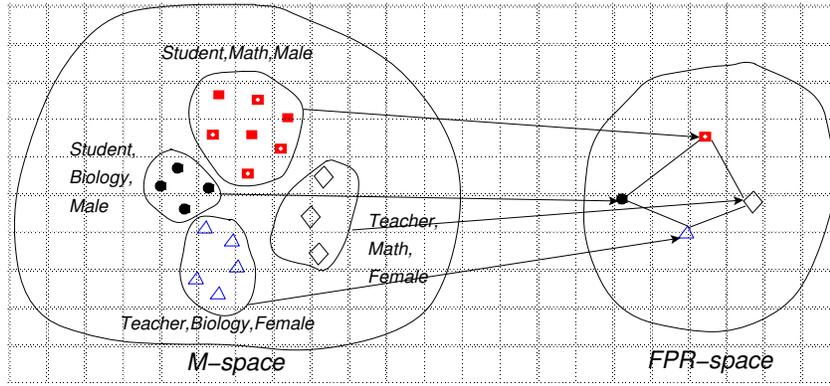


Figure 3: The M-space converts into the FPR-space.

Fig. 3 gives three key features, which are identity, profession and gender. The shape “■” represents a user who is a male student with math profession; the shape “◇” represents a user who is a female teacher with math profession; the shape “△” represents a user who is a female teacher with biology profession; the shape “●” represents a user who is a male student with biology profession. A teacher (student) preferentially makes contact with another teacher (student) when they are female (male). Moreover, a teacher (student) preferentially contacts with another student when they have the same profession. Four solid lines represent the priority relationships among these four groups (nodes).

An $(n_{\{m_1, \dots, m_n\}}, k)$ -hypercube [20] is applied to provide the *priority relation graph* (PRG) [20] of MSN to obtain an efficient multi-path data delivery routing. The definition of $(n_{\{m_1, \dots, m_n\}}, k)$ -hypercube ($k < \sqrt{n}$) is given as follows.

Definition 2 [20] Let n , k ($k < \sqrt{n}$) and m_i be positive integers and $m_i \geq 2$ ($1 \leq i \leq n$). The $(n_{\{m_1, \dots, m_n\}}, k)$ -hypercube, denoted by $Q_{n_{\{m_1, \dots, m_n\}}, k}$, has the node-set $V(Q_{n_{\{m_1, \dots, m_n\}}, k}) = \{v_1 \cdots v_i \cdots v_n \mid v_i \in \{0, 1, \dots, m_i - 1\} (1 \leq i \leq n)\}$ and the edge-set $E(Q_{n_{\{m_1, \dots, m_n\}}, k}) = \{uv \mid u \text{ and } v \text{ differ in no more than } k \text{ positions}\}$.

We denote the $(n_{\{m_1, \dots, m_n\}}, k)$ -hypercube $Q_{n_{\{m_1, \dots, m_n\}}, k}$ as the (n, k) -hypercube $Q_{n, k}$ when $m_i = 2$ for all $i = 1, 2, \dots, n$. Fig. 4 gives a part of structure of $(5, 2)$ -hypercube.

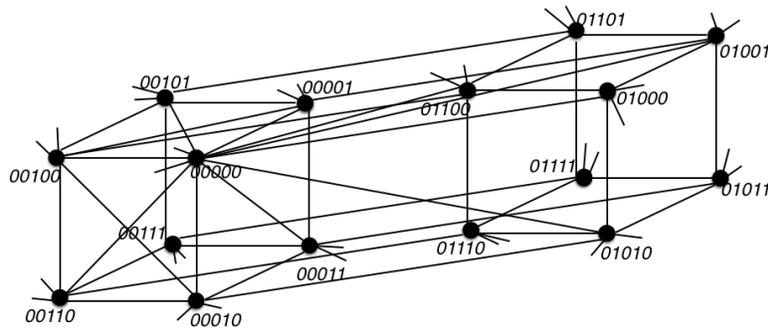


Figure 4: Part of structure of $(5, 2)$ -hypercube $Q_{5, 2}$.

Throughout this paper, an edge is called an s -edge, denoted by e_s , if it connects two nodes having exactly s ($\leq k$) different positions. If uv is an s -edge, then u (resp., v) is an s -neighbor of v (resp., u). The set of all s -edges is called an s -edge-set, denoted by E_s . It implies that $E(Q_{n_{\{m_1, \dots, m_n\}}, k}) = \bigcup_{s=1}^k E_s$. For any node u , the set of all s -neighbors of u is denoted by $N_s(u)$. In particular, when $m_i = 2$ ($i = 1, 2, \dots, n$) and $k = 1$, $Q_{n_{\{m_1, \dots, m_n\}}, k}$ is Q_n [34]. This means that Q_n is a special case of $Q_{n_{\{m_1, \dots, m_n\}}, k}$, hence $Q_{n_{\{m_1, \dots, m_n\}}, k}$ is expected to have more general properties.

Definition 3 [20] Let n be the number of key features in a mobile social network and let k be an integer with $k < \sqrt{n}$. Every node in the priority relation graph (PRG) is a group, which consists of users having the same feature values in all n key features with each other. Two groups in PRG have a communication link if and only if they have a priority relationship, i.e., their feature distance is no more than k . Note that, every position of a node is a feature of the user.

We [20] previously proposed that the PRG of MSN is the $(n_{\{m_1, \dots, m_n\}}, k)$ -hypercube $Q_{n_{\{m_1, \dots, m_n\}}, k}$ (see Fig. 5). It implies that the feature-based PRG of MSN is a regular graph. Fig. 5 shows the PRG of a school network. Fig. 5 (a) shows the physical locations of mobile users in the local school network where mobile users have three key social features (Identity, Faculty and Profession). The mobile users with the same values in all key features with each other form a group in Fig. 5 (b). Fig. 5 (c) gives the PRG of the school network based on the three key social features. A teacher (student) has a higher chance of contacting other teachers (students) when they have the same profession and in the same school rather than the teachers (students) with other professions or in other schools.

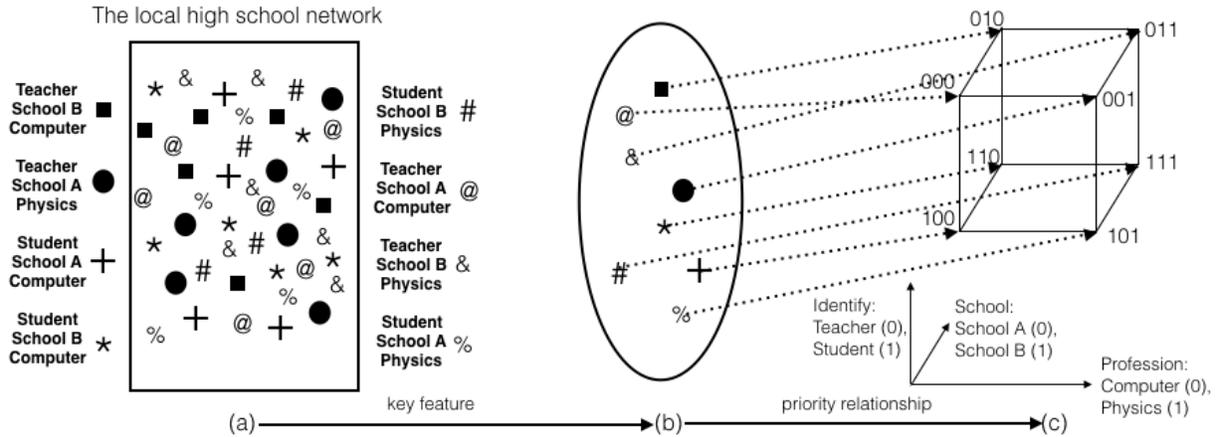


Figure 5: The PRG of the school network.

In this paper, an approach called *destination-based partitioning* is used to partition users into nodes according to the values of features of the destination. On each dimension (corresponding to a feature) in PRG, the users are assigned to the destination node or a different node based on whether or not they have the same values in all key features as

the one at the destination. This means that an $(n_{\{m_1, \dots, m_n\}}, k)$ -hypercube is “compressed” into an (n, k) -hypercube $Q_{n,k}$ even though each feature may have many different values.

To propose an efficient and systematic data delivery multi-path communication algorithm, we firstly give some available properties of (n, k) -hypercube as follows.

Property 2 *The diameter of (n, k) -hypercube is $\lceil n/k \rceil$.*

Proof. Let u and v be two nodes in the (n, k) -hypercube. Assume that u and v differ in exactly l positions. If $l \leq k$, then $d(u, v) = 1$ according to the definition of PRG. If $l > k$, then let $l = ks + t$, $u = u_1 u_2 \cdots u_n$ and $v = v_1 v_2 \cdots v_n$. Without loss of generality, assume that $u_i \neq v_i$ for $1 \leq i \leq l$. By the fact that any two nodes are connected if and only if their feature distance is no more than k , we can make up the following shortest path between u and v .

$$\begin{aligned}
u &= \underbrace{u_1 \cdots u_k}_{1st\ k} \underbrace{u_{k+1} \cdots u_{2k}}_{2nd\ k} \cdots \underbrace{u_{(s-1)k+1} \cdots u_{sk}}_{sth\ k} \underbrace{u_{sk+1} \cdots u_n}_t \\
&\rightarrow \underbrace{v_1 \cdots v_k}_{1st\ k} \underbrace{u_{k+1} \cdots u_{2k}}_{2nd\ k} \cdots \underbrace{u_{(s-1)k+1} \cdots u_{sk}}_{sth\ k} \underbrace{u_{sk+1} \cdots u_n}_t \\
&\quad \rightarrow \dots \\
&\rightarrow \underbrace{v_1 \cdots v_k}_{1st\ k} \underbrace{v_{k+1} \cdots v_{2k}}_{2nd\ k} \cdots \underbrace{v_{(s-1)k+1} \cdots v_{sk}}_{sth\ k} \underbrace{u_{sk+1} \cdots u_n}_t \\
&\rightarrow \underbrace{v_1 \cdots v_k}_{1st\ k} \underbrace{v_{k+1} \cdots v_{2k}}_{2nd\ k} \cdots \underbrace{v_{(s-1)k+1} \cdots v_{sk}}_{sth\ k} \underbrace{v_{sk+1} \cdots v_n}_t.
\end{aligned}$$

Therefore, $d(u, v) = s + 1 = \lceil l/k \rceil$.

According to the definition of diameter,

$$D(Q_{n,k}) = \max \{d(u, v) \mid u, v \in V(Q_{n,k})\} = \max \{\lceil l/k \rceil \mid k < l \leq n\} = \lceil n/k \rceil.$$

■

Property 3 *The regular degree of (n, k) -hypercube is $\binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{k}$.*

Proof. Let u be any node in the (n, k) -hypercube. According to the definition of (n, k) -hypercube, $N(u) = \cup_{l=1}^k N_l(u)$.

For any $v \in N_l(u)$, u and v differ in exactly l positions. Because any node in (n, k) -hypercube has n positions, $|N_l(u)| = \binom{n}{l}$ and $N_i(u) \cap N_j(u) = \emptyset$ for any $i \neq j \in \{1, 2, \dots, k\}$. Therefore, $|N(u)| = \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{k}$. By the arbitrariness of u in $Q_{n,k}$, the regular degree of (n, k) -hypercube is $\binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{k}$. ■

According to Properties 2 and 3, (n, k) -hypercube has smaller diameter and larger regular degree than n -hypercube. Hence, (n, k) -hypercube could be applied to give a more efficient routing than n -hypercube.

According to Property 3, the routing algorithm based on PRG can provide a disjointed multi-path data delivery routing. Because the latency of a data delivery routing is decided by the feature distance, the priority relationship reduces the forwarding time of data

delivery. Because the proposed PRG selects preferential contacts as its nodes, any link in the proposed PRG is an efficient communication link. Therefore, the proposed priority relation graph of MSN not only is a fundamental framework of data delivery routing, but also can provide an efficient and systematic disjointed multi-path data delivery routing. An available property is given as follows.

Property 4 *A priority relation graph guarantees an efficient and systematic disjointed multi-path data delivery routing.*

The multi-path technology can be exploited for ensuring that the important information can be delivered through a highly reliable path between each pair of sender and receiver. Then we calculate the tolerant ability of ‘faults’ and estimate the availability of MSN on the theoretical level. The fault-tolerance method does not consider the location of faults, but cares only the number of faults. When a user wants to deliver data to another user, the source user can choose an appropriate carrier to communicate all of his/her preferential contacts through mobile devices under the short-range communication technologies. Hence, an efficient and systematic data delivery communication algorithm based on the PRG is necessary. Data delivery routing is introduced in this section. We provide node-disjointed multi-path communication algorithm between any two mobile users.

If source user x wants to deliver a data to destination user y in MSN, then the source user needs to go through the following several steps.

- **Step 1.** Source user x uses a data mining method to find key features in the MSN.
- **Step 2.** Source user x finds the destination user y .
- **Step 3.** Source user x builds the PRG of MSN based on the set of key features, destination-based partitioning and the proposed priority relationship.
- **Step 4.** Source user x determines multi-path communication to the destination user y according to the proposed algorithm.

A multi-path data delivery routing algorithm is presented here for the objective of reaching the destination quickly, with maximal fault tolerance, an acceptable forwarding number and maximal delivery rate. Assume that the source user x and the destination user y have exactly l different feature values, and let C_l be the set of features where these users have different feature values. Some notations are given in Table 3.

Table 3: Some notations

$P_{l:x \rightarrow y}$	The set of all paths between source group x and destination group y
$P_{l:x \rightarrow y}^{j_1 j_2 \dots j_s}$	The path begins with an s -edge whose end nodes differ in the j_1 -th position, \dots , and j_s -th position
$P_{l:x \rightarrow y}^{s:b}$	The logogram of $P_{l:x \rightarrow y}^{j_1 j_2 \dots j_s}$ where b represents the decimal value converted from the binary code $j_1 j_2 \dots j_s$
$P_{l:x \rightarrow y}^s$	The set of all paths which begin with an s -edge

It is obvious that $P_{l:x \rightarrow y}^{s:b}$ is the b -th path in the path-set $P_{l:x \rightarrow y}^s$, and $P_{l:x \rightarrow y} = \bigcup_{s=1}^k P_{l:x \rightarrow y}^s$.

To describe the data delivery routing algorithm more clearly, a new concept named the *natural order* of $j_1j_2 \cdots j_s$ on $\langle n \rangle$ is defined as follows.

If $s = 1$, then a natural order of j_1 on $\langle n \rangle$ is the values of j_1 ordered from 1 to n .

If $s \geq 2$, then a natural order of $j_1j_2 \cdots j_s$ on $\langle n \rangle$ is the ordered values of $j_1j_2 \cdots j_s$ satisfying that $j_a < j_b$ for any $a < b$ ($a, b \in \{1, 2, \dots, s\}$) and the value of j_1 is arranged in order from 1 to $n - s + 1$. It is easy to see the code $12 \cdots (s - 1)s$ is first value of $j_1j_2 \cdots j_s$ in the natural order of $j_1j_2 \cdots j_s$ on $\langle n \rangle$, the code $2 \cdots (s - 1)s(s + 1)$ is the second value of $j_1j_2 \cdots j_s$ in the natural order of $j_1j_2 \cdots j_s$ on $\langle n \rangle$, ..., and the code $(n - s + 1)(n - s + 2) \cdots (n - 1)n$ is the last value of $j_1j_2 \cdots j_s$ in the natural order of $j_1j_2 \cdots j_s$ on $\langle n \rangle$.

To describe the data delivery routing algorithm more simply, a new notation will be proposed as follows.

If uv is an s -edge in (n, k) -hypercube having feature distance s at the j_1 -th, j_2 -th, ..., j_s -th positions, where $j_1 < j_2 < \cdots < j_s$, then we use the notation $(u, u^{(j_1j_2 \cdots j_s)})$ to denote the edge uv . We also use $(u, u^{(C)})$ to denote the edge uv when $C = \{j_1, j_2, \dots, j_s\}$.

Let n be the number of key features in MSN. Recall that two users with a priority relationship differ in no more than k features ($k < \sqrt{n}$). It implies that two users are more willing to contact each other when they differ in fewer features. Based on this consideration, we will establish a multi-path *priority relation graph-based social feature routing* (PRG-SFR) algorithm. The detailed PRG-SFR algorithm and its sub-algorithms are shown in Algorithm 1 and Algorithms 2-4, respectively.

Algorithm 1 *PRG-SFR (The priority relation graph-based social feature routing)*

Require: Source user x , destination user y , the set of different features C between x and y , the heading code sets D, A^1 , the auxiliary heading code set J' .

Ensure: $P_{l:x \rightarrow y}$.

- 1: Let $l = |C|$ and let $P_{l:x \rightarrow y}, D, A^1, J'$ be the empty sets;
 - 2: **for** j selected from natural order of j_1 on $\langle n \rangle$ **do**
 - 3: Execute subroutine beginning with 1-edges (see Algorithm 2) for the data delivery routing;
 - 4: $P_{l:x \rightarrow y} = P_{l:x \rightarrow y} \cup P_{l:x \rightarrow y}^j$ and $D = D \cup \{j\} \cup A^1$;
 - 5: **end for**
 - 6: **for** any value selected from the natural order of $j_1j_2 \cdots j_s$ on $\langle n \rangle$ ($2 \leq s \leq k$) **do**
 - 7: Execute subroutine beginning with s -edges ($2 \leq s \leq k$) (see Algorithm 3) for the data delivery routing;
 - 8: $P_{l:x \rightarrow y} = P_{l:x \rightarrow y} \cup P_{l:x \rightarrow y}^{j_1 \cdots j_s}$ and $D = D \cup \{j_1, j_2, \dots, j_s\} \cup A^1$;
 - 9: **end for**
 - 10: **if** the auxiliary heading code set $J' \neq \emptyset$ **then**
 - 11: Execute subroutine beginning with nonempty auxiliary heading code set J' (see Algorithm 4) for the data delivery routing;
 - 12: Send the set $P_{l:x \rightarrow y} = P_{l:x \rightarrow y} \cup P_{l:x \rightarrow y}^{J'}$;
 - 13: **end if**
 - 14: **return** $P_{l:x \rightarrow y}$;
-

Algorithm 2 Subroutine beginning with 1-edges

Require: x, y, j, C .**Ensure:** $P_{l:x \rightarrow y}^j; A^1$.

- 1: Let $P_{l:x \rightarrow y}^j$ be the empty set;
 - 2: $C \setminus \{j\}$ is divided into segments that have an equal size of k (≥ 2) (except the last segment) in the order of elements in C ;
 - 3: If $j \notin C$ and the size of the last segment is smaller than k , then j is added to the last segment;
 - 4: If $j \notin C$ and the size of the last segment is k , then let j be the new last segment;
 - 5: Let A^1 be the first segment;
 - 6: The consequent path of $P_{l:x \rightarrow y}^j$ after the 1-edge starting segment from x is generated step by step by going through a matching process along the segments obtained above;
 - 7: **return** $P_{l:x \rightarrow y}^j; A^1$;
-

Algorithm 3 Subroutine beginning with s -edges ($2 \leq s \leq k$)

Require: $x, y, j_1, j_2, \dots, j_s, C, D$.**Ensure:** $P_{l:x \rightarrow y}^{j_1 j_2 \dots j_s}, A^1, J'$.

- 1: Let $P_{l:x \rightarrow y}^{j_1 j_2 \dots j_s}$ and J' be empty sets and let $J = \{j_1, j_2, \dots, j_s\}$;
 - 2: $C \setminus J$ is divided into segments of k elements in order;
 - 3: If $|J \cap C| = c$ and the size of the last segment is smaller than $k - c + 1$, then the elements in $J \cap C$ are added to the last segment;
 - 4: Let A^1 be the first segment;
 - 5: **if** $\{j_1, j_2, \dots, j_s\} \cup A^1$ belongs to D **then**
 - 6: $J' = J' \cup \{j_1, j_2, \dots, j_s\}$;
 - 7: **else**
 - 8: The consequent path of $P_{l:x \rightarrow y}^{j_1 j_2 \dots j_s}$ after the s -edge starting segment from x is generated step by step by going through a matching process along the segments obtained above;
 - 9: **end if**
 - 10: **return** $P_{l:x \rightarrow y}^{j_1 j_2 \dots j_s}, A^1, J'$;
-

To explain PRG-SFR algorithm better, we give the following two examples depending on different values of n in MSN.

Example 1 *Suppose there are six key features in a special MSN and two users have a priority relationship if and only if they differ in no more than two features. The values of key features of two groups x and y are 000000 and 010010, respectively.*

According to PRG-SFR algorithm, we can obtain the reliable multi-path communication between $x = 000000$ and $y = 010010$ on the $Q_{6,2}$ in Fig. 6 and Table 4. The weight of each edge in Fig. 6 represents the positions where the corresponding feature values are different at the two end nodes. For example, the leftmost path of the first subgraph goes through $000000 \rightarrow 100000 \rightarrow 110010 \rightarrow 010010$.

Example 2 *Suppose there are seven key features in a special MSN and two users have a*

Algorithm 4 Subroutine beginning with nonempty auxiliary heading code set J'

Require: x, y, C, D, J' .

Ensure: The set of remaining paths $P_{l:x \rightarrow y}^{J'}$.

- 1: Let $P_{l:x \rightarrow y}^{J'}$ be the empty set;
 - 2: **for** each J_1 from J' **do**
 - 3: Let $c = |J_1 \cap C|$;
 - 4: The first k elements are selected from $C \setminus J_1$, and they form the set A^1 ;
 - 5: **if** $A^1 \cup J_1$ is not contained in D **then**
 - 6: The selection is stopped;
 - 7: **else**
 - 8: Continue to choose the first s elements ($s \in \{k-1, k-2, \dots, 1\}$) from $C \setminus J_1$ and they form the set A^1 ;
 - 9: **if** $A^1 \cup J_1$ is not contained in D **then**
 - 10: The selection is stopped;
 - 11: **end if**
 - 12: **end if**
 - 13: **if** $t + s - 2c \leq k$ **then**
 - 14: The order segments of $A = C \setminus J_1$ are denoted by $A^1, A^2, \dots, A^m, A^{m+1}$;
 - 15: **else**
 - 16: The order segments of $A = C \setminus J_1$ are denoted by $A^1, A^2, \dots, A^{m+1}, A^{m+2}$;
 - 17: **end if**
 - 18: A path $P_{l:x \rightarrow y}^{J_1}$ is constructed based on the order of $\{J_1, A^1, \dots, A^m, A^{m+1}\}$ or $\{J_1, A^1, A^2, \dots, A^{m+1}, A^{m+2}\}$;
 - 19: $P_{l:x \rightarrow y}^{J'} = P_{l:x \rightarrow y}^{J'} \cup P_{l:x \rightarrow y}^{J_1}$;
 - 20: **end for**
 - 21: **return** $P_{l:x \rightarrow y}^{J'}$;
-

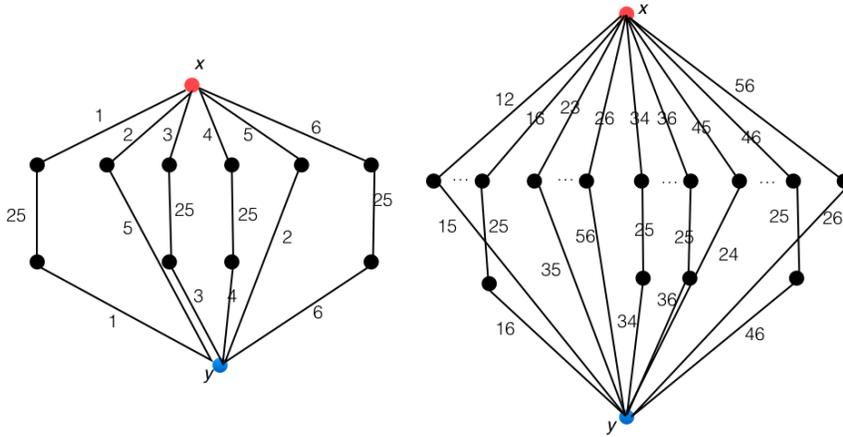


Figure 6: Multi-path communication between two users $x = 000000$ and $y = 010010$.

priority relationship if and only if they differ in no more than two features. The values of

Table 4: The disjointed paths between two users $x = 000000$ and $y = 010010$

$P_{l:x \rightarrow y}^1 = (x, 100000, 110010, y)$	$P_{l:x \rightarrow y}^{13} = (x, 101000, 111010, y)$	$P_{l:x \rightarrow y}^{26} = (x, 010001, y)$
$P_{l:x \rightarrow y}^2 = (x, 010000, y)$	$P_{l:x \rightarrow y}^{14} = (x, 100100, 110110, y)$	$P_{l:x \rightarrow y}^{34} = (x, 001100, 011110, y)$
$P_{l:x \rightarrow y}^3 = (x, 001000, 011010, y)$	$P_{l:x \rightarrow y}^{15} = (x, 100010, y)$	$P_{l:x \rightarrow y}^{35} = (x, 001010, y)$
$P_{l:x \rightarrow y}^4 = (x, 000100, 010110, y)$	$P_{l:x \rightarrow y}^{16} = (x, 100001, 110011, y)$	$P_{l:x \rightarrow y}^{36} = (x, 001001, 011011, y)$
$P_{l:x \rightarrow y}^5 = (x, 000010, y)$	$P_{l:x \rightarrow y}^{23} = (x, 011000, y)$	$P_{l:x \rightarrow y}^{45} = (x, 000110, y)$
$P_{l:x \rightarrow y}^6 = (x, 000001, 010011, y)$	$P_{l:x \rightarrow y}^{24} = (x, 010100, y)$	$P_{l:x \rightarrow y}^{46} = (x, 000101, 010111, y)$
$P_{l:x \rightarrow y}^{12} = (x, 110000, y)$	$P_{l:x \rightarrow y}^{25} = (x, y)$	$P_{l:x \rightarrow y}^{56} = (x, 000011, y)$

key features of two groups x and y are 0000000 and 0110101, respectively.

According to PRG-SFR algorithm, we can obtain the multi-path communication between $x = 0000000$ and $y = 0110101$ on the $Q_{7,2}$ in Fig. 7. Table 5 give the detailed routings. For simplicity, let $P_{l:x \rightarrow y}^{j_1 j_2 \dots j_s}$ be $P_l^{j_1 j_2 \dots j_s}$ in Table 5. The weight of each edge in Fig. 7 represents the positions where the corresponding features are different at the two end nodes. For example, the leftmost path of the first subgraph goes through $0000000 \rightarrow 1000000 \rightarrow 1110000 \rightarrow 1110101 \rightarrow 0110101$.

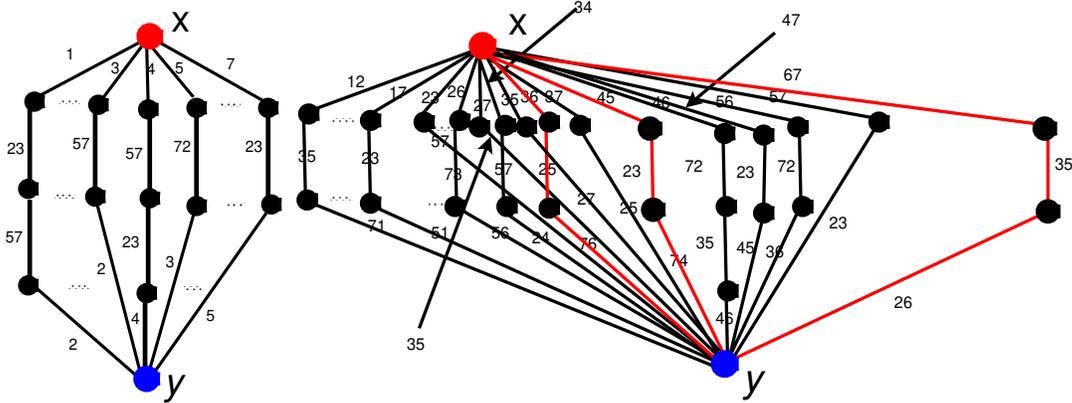


Figure 7: Multi-path communication between two users $x = 0000000$ and $y = 0110101$.

4 Theoretical Analysis

In this section, we evaluate the efficiency of the PRG-SFR algorithm, which is theoretically analyzed in terms of fault tolerance, forwarding number, transmission time and delivery rate. There are two commonly used metrics for the performance analysis, which are introduced as follows.

The *similarity* with a priority relationship between two users depends on their feature distance. For any two users with a priority relationship, the greater the feature distance is, the less similarity the users have. Hence, the similarity is inversely proportional to

Table 5: The disjointed paths between two users $x = 0000000$ and $y = 0110101$

$P_l^1 = (x, 1000000, 1110000, 1110101, y)$	$P_l^{15} = (x, 1000100, 1100101, y)$	$P_l^{36} = (x, 0010010, 0110110, y)$
$P_l^2 = (x, 0100000, 0110100, y)$	$P_l^{16} = (x, 1000010, 1100010, 1110110, y)$	$P_l^{37} = (x, 0010001, y)$
$P_l^3 = (x, 0010000, 0010101, y)$	$P_l^{17} = (x, 1000001, 1110001, y)$	$P_l^{45} = (x, 0001100, 0111100, y)$
$P_l^4 = (x, 0001000, 0001101, 0111101, y)$	$P_l^{23} = (x, 0110000, y)$	$P_l^{46} = (x, 0001010, 0101011, 0111111, y)$
$P_l^5 = (x, 0000100, 0100101, y)$	$P_l^{24} = (x, 0101000, 0101101, y)$	$P_l^{47} = (x, 0001001, 0111001, y)$
$P_l^6 = (x, 0000010, 0100011, 0110111, y)$	$P_l^{25} = (x, 0100100, y)$	$P_l^{56} = (x, 0000110, 0100111, y)$
$P_l^7 = (x, 0000001, 0110001, y)$	$P_l^{26} = (x, 0100010, 0110011, y)$	$P_l^{57} = (x, 0000101, y)$
$P_l^{12} = (x, 1100000, 1110100, y)$	$P_l^{27} = (x, 0100001, y)$	$P_l^{67} = (x, 0000011, 0010111, y)$
$P_l^{13} = (x, 1010000, 1010101, y)$	$P_l^{34} = (x, 0011000, 0011101, y)$	
$P_l^{14} = (x, 1001000, 1001101, 1111101, y)$	$P_l^{35} = (x, 0010100, y)$	

their feature distance. Assume that the similarity of end nodes of s -edges ($1 \leq s \leq k$) in (n, k) -hypercube is modeled by $p(s) = \frac{n-s}{n}$. It implies that $p(s)$ is the acquaintance between two users who have a priority relationship and differ in exactly s features. It is easy to check that $p(s) = \frac{n-s}{n} = s \cdot p(1) - (s-1)$. That is to say, the similarity of two end nodes of one s -edge coincides with that of any s -edge.

Recall that ordinary people contact each other more frequently when they have more social features in common [34, 36, 37]. Hence, we think that the transmission time and delivery rate of a communication link are mainly decided by the feature distance between two users and they are time-independent [13]. The transmission time (resp., delivery rate) of s -edge e_s is denoted by $T^1(s)$ (resp., $DR^1(s)$), which increases (resp., decreases), when the key feature distance s is increases.

4.1 Fault Tolerance

In this part, we prove that the obtained multi-paths by PRG-SFR algorithm are *node-disjointed*, where the *node-disjointness* means that the multi-path communications do not cross over each other, except at source node and destination node, and it increases the efficiency of routing. It implies that PRG-SFR algorithm has a stronger fault tolerance.

The **Fault tolerance (FT)** is defined as the maximal number of disjointed paths between any two users, which means that the failure of one node only results in the failure of one path and does not affect other paths.

Theorem 1 *Let n be the number of key features in MSN and let k be an integer with $k < \sqrt{n}$. Two users in MSN have a priority relationship if their feature distance is no more than k . Let u and v be the source user and the destination user, respectively. Then, there are $\binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{k}$ disjointed paths between u and v .*

Proof. Let $N(u)$ be the neighborhood of u . According to the definition of an s -neighbor ($1 \leq s \leq k$), $N(u) = \bigcup_{1 \leq s \leq k} N_s(u)$. According to PRG-SFR algorithm and its sub-algorithms proposed in the previous section, any one path $P_{l:u \rightarrow v}^{j_1 j_2 \dots j_s}$ in $P_{l:u \rightarrow v}^s$ begins with an s -edge, i.e., the successor node of u in the path $P_{l:u \rightarrow v}^{j_1 j_2 \dots j_s}$ is in $N_s(u)$. Because $|N_s(u)| = \binom{n}{s}$,

the number of the path-set $P_{l:u \rightarrow v}^s$ is $\binom{n}{s}$. Thus, there are $\binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{k}$ paths between u and v .

Next, these paths are mutually disjoint because Steps 2 and 3 in Subroutine beginning with 1-edges algorithm (Algorithm 2) and Subroutine beginning with s -edges ($2 \leq s \leq k$) algorithm (Algorithm 3) guarantee that the nodes on each path are different from those in any other path until the path reaches the destination node. ■

According to the definition of fault tolerance and Theorem 1, we can directly obtain the following result.

Corollary 1 *The fault tolerance FT of PRG-SFR algorithm is $\binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{k}$.*

4.2 Forwarding Number

The path length is used to evaluate the number of forwarding. Let $L_{l,s}$ be the maximal length of all paths in the path-set $P_{l:u \rightarrow v}^s$ and let L_l be the maximal length of all paths in the path-set $P_{l:u \rightarrow v}$.

Definition 4 *The average number of forwarding of each packet before the destination being reached is called the forwarding number (denoted by FN) of the packet. Structurally, the forwarding number is the average length of all disjointed paths.*

Since there is a maximal of the lengths of all disjointed paths between any two given users in MSN, their forwarding number has a upper bound.

Theorem 2 *Let n be the number of key features in MSN and let k be an integer with $k < \sqrt{n}$. Two users in MSN have a priority relationship if their feature distance is no more than k different feature values. Let u and v be the source user and the destination user, respectively, and let the feature distance of them be l , then the maximal length of all $\binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{k}$ disjointed paths between u and v is*

$$L_l = \begin{cases} 3, & \text{if } k \geq l, \\ \lceil \frac{l}{k} \rceil + 2, & \text{if } k < l. \end{cases}$$

Proof. Let u and v be any two nodes in (n, k) -hypercube, which differ in exactly l positions. Let $C = \{i_1, i_2, \dots, i_l \mid i_1 < i_2 < \dots < i_l \text{ and the } i_s\text{-th position } (1 \leq s \leq l) \text{ of } u \text{ is different from } v\}$.

First, the maximal length of all paths in the path-set $P_{l:u \rightarrow v}^1$ is considered. According to Algorithm 2, the worst case of all paths is the case of $j \notin C$. Hence, $|C \cup \{j\}| = l + 1$. We distinguish among the following three cases depending on the sizes of k and l .

If $k > l$, we have $P_{l:u \rightarrow v}^j = (x, x^{(j)}, x^{(j)(C \cup \{j\})})$. Hence, $L_{l,1} = 2$.

If $k = l$, we have $P_{l:u \rightarrow v}^{j_1} = (x, x^{(j)}, x^{(j)(C)}, x^{(j)(C)(j)})$. Hence, $L_{l,1} = 3$.

If $k < l = mk + p$, then $A = C \cup \{j\} = A^1 \cup \dots \cup A^m \cup A^{m+1}$ where $|A^i| = k$ for $1 \leq i \leq m$ and $|A^{m+1}| = p + 1 \leq k$. We have $P_{l:u \rightarrow v}^{j_1} = (x, x^{(j)}, x^{(j)(A^1)}, \dots, x^{(j)(A^1) \dots (A^m)}, y)$. Hence, $L_{l,1} = m + 2$.

Second, the maximal length of all paths in the path-set $P_{l:u \rightarrow v}^s$ ($2 \leq s \leq k$) is considered. According to Algorithms 3 and 4, the worst case of all paths is the case of $|\{j_1, j_2, \dots, j_s\} \cap C| = 0$. Let $J = \{j_1, j_2, \dots, j_s\}$. We distinguish among the following three cases depending on the size of k and l .

If $k \geq l$, we have $P_{l:u \rightarrow v}^{j_1 \dots j_s} = (x, x^{(j_1 \dots j_s)}, x^{(j_1 \dots j_s)(C)}, x^{(j_1 \dots j_s)(C)(J)})$. Hence, $L_{l,s} = 3$.

If $k < l = mk + p$ and $p + s \leq k$, then $A = C \cup \{j_1, j_2, \dots, j_s\} = A^1 \cup \dots \cup A^m \cup A^{m+1}$ where $|A^i| = k$ for $1 \leq i \leq m$ and $|A^{m+1}| = p + s \leq k$. We have $P_{l:u \rightarrow v}^{j_1 j_2} = (x, x^{(j_1 j_2)}, x^{(j_1 j_2)(A^1)}, \dots, x^{(j_1 j_2)(A^1) \dots (A^m)(A^{m+1})})$. Hence, $L_{l,2} = m + 2$.

If $k < l = mk + p$ and $p + s > k$, then $A = C \cup \{j_1, j_2, \dots, j_s\} = A^1 \cup \dots \cup A^m \cup A^{m+1} \cup A^{m+2}$ where $|A^i| = k$ for $1 \leq i \leq m$, $|A^{m+1}| = p$ and $|A^{m+2}| = s$. We have $P_{l:u \rightarrow v}^{j_1 \dots j_s} = (x, x^{(j_1 \dots j_s)}, x^{(j_1 \dots j_s)(A^1)}, \dots, x^{(j_1 \dots j_s)(A^1) \dots (A^{m+1})}, y)$. Hence, $L_{l,s} = m + 3$.

Therefore, the maximal length of all $\binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{k}$ disjointed paths between u and v is

$$L_l = \begin{cases} 3, & \text{if } k \geq l, \\ \lceil \frac{l}{k} \rceil + 2, & \text{if } k < l. \end{cases}$$

■

According to Definition 4 and Theorem 2, we can directly obtain the following result.

Corollary 2 *The upper bound of forwarding number of PRG-SFR algorithm is $FN \leq \lceil \frac{l}{k} \rceil + 2$ when $k < l$ or $FN \leq 3$ when $k \geq l$.*

4.3 Transmission Time

The PRG-SFR algorithm can find a sufficient number of loop-free paths. However, since the transmission time is used as routing cost, data tends to be delivered through the paths with less transmission time. Longer paths are mainly used as backups in case of the failure of shorter paths.

In FPR-space, the transmission time of an edge is inversely proportional to its similarity of end nodes. Therefore, the transmission time of an s -edge, denoted by $T^1(s)$, is a monotone increasing function of s . It is discovered that the transmission time of an s -edge is smaller than the sum of transmission times of s 1-edges (see Fig. 8).

Observation 1. $k * T^1(1) > T^1(k) > T^1(k - 1) > \dots > T^1(1)$ (see Fig. 8).

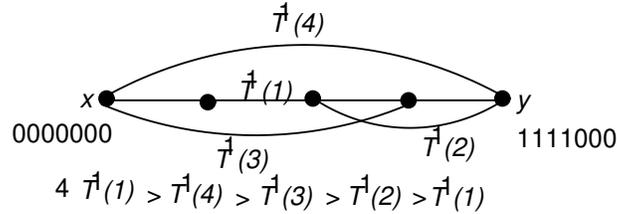


Figure 8: The comparison of transmission times among 1-edge, 2-edge, ... and k -edge.

The transmission time along the path $P_{l:x \rightarrow y}^{j_1 j_2 \dots j_s}$ is the sum of all inter-contact times, each of which is the transmission time of an edge on the path.

Definition 5 *Let u and v be the source user and the destination user, respectively, and let the feature distance of them be l , then the transmission time in the path-set $P_{l:u \rightarrow v}$, denoted by T_l , is the minimal value among the transmission times of all disjointed paths.*

The transmission time between the source user u and the destination user v in MSN is computed according to the following theorem.

Theorem 3 Let n be the number of key features in MSN and let k be an integer with $k < \sqrt{n}$. Two users in the MSN have a priority relationship if their feature distance is no more than k . Let u and v be the source user and the destination user, respectively, and let the feature distance of them be l . Then the transmission time in the path-set $P_{l:u \rightarrow v}$ is

$$T_l = \begin{cases} T^1(1), & \text{if } l = 1, \\ T^1(2) + T^1(l - 2), & \text{if } k \geq l, \\ T^1(1) + \lfloor \frac{l}{k} \rfloor * T^1(k) + T^1(p - 1), & \text{if } k < l, \end{cases}$$

where $l = mk + p$ ($1 \leq p \leq k$).

Note: $T^1(0) = 0$ and $0 * T^1(z) = 0$ where $1 \leq z \leq k$.

Note that: The proof of Theorem 3 is in Appendix.

4.4 Delivery Rate

The delivery rate of a packet is the probability of successful delivery within the time t . The packet is usually transmitted through multiple relay groups. Assume that the average time needed to pass each relay group is \bar{t} . In FPR-space, the delivery rate $DR_{\bar{t}}^1(s)$ of an s -edge is proportional to its similarity. Therefore, $DR_{\bar{t}}^1(s)$ is a monotone decreasing function of s ($\leq k$) within time \bar{t} . It is discovered that the delivery rate of an s -edge is larger than the product of the delivery rates of s 1-edges within time \bar{t} (see Fig. 9).

Observation 2. $(DR_{\bar{t}}^1(1))^k < DR_{\bar{t}}^1(k) < DR_{\bar{t}}^1(k - 1) < \dots < DR_{\bar{t}}^1(1)$ (see Fig. 9).

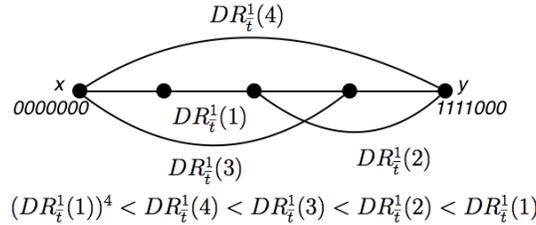


Figure 9: The comparison of delivery rates among 1-edge, 2-edge, ... and k -edge within time \bar{t} .

The delivery rate along the path $P_{l:x \rightarrow y}^{j_1 j_2 \dots j_s}$ within time t is the product of all delivery rates of all edges on the path, each of which is the delivery rate within time \bar{t} .

Definition 6 In FPR-space, let u and v be the source user and the destination user, respectively, and let the feature distance of them be l . The delivery rate from u to v within time t , denoted by DR_t^l , is the maximal value among the delivery rates of all $\binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{k}$ disjointed paths within time t .

Theorem 4 Let n be the number of key features in MSN and let k be an integer with $k < \sqrt{n}$. Two users in the MSN have a priority relationship if their feature distance is no

more than k . Let u and v be the source user and the destination user, respectively, and let the feature distance of them be l . The delivery rate from u to v within time t is

$$DR_t^l = \begin{cases} DR_{\bar{t}}^1(1), & \text{if } l = 1, \\ DR_{\bar{t}}^1(2) \cdot DR_{\bar{t}}^1(l-2), & \text{if } k \geq l, \\ DR_{\bar{t}}^1(1) \cdot (DR_{\bar{t}}^1(k))^{\lfloor \frac{l}{k} \rfloor} \cdot DR_{\bar{t}}^1(p-1), & \text{if } k < l, \end{cases}$$

where $l = mk + p$ ($1 \leq p \leq k$) and \bar{t} is the average time needed to pass each relay user. Note: $DR_{\bar{t}}^1(0) = 1$ and $(DR_{\bar{t}}^1(z))^0 = 1$ where $1 \leq z \leq k$.

Note that: The proof of Theorem 4 is in Appendix.

5 Numerical Analysis

We compare the performance of PRG-SFR algorithm with the hypercube-based social feature routing (HSFR) algorithm via numerical analysis in Matlab. The analysis results are verified in terms of fault tolerance, forwarding number, transmission time and delivery rate. On the other hand, Wu *et al.* [34] had pointed out their algorithm has competitive performance compared to the spray-and-wait algorithm, the spray-and-focus algorithm, and the social-aware routing algorithm from the above four aspects. Hence, if we could show the superiority of PRG-SFR algorithm compared to HSFR algorithm, then we could conclude that PRG-SFR algorithm has better performance than the other algorithms.

Let n be the number of key features. Assume that two users have a priority relationship of their feature distance is no more than k . Let l be the feature distance of source user u and the destination user and let m be the number obtained in Algorithm 4.

Firstly, we compare the fault tolerances of PRG-SFR algorithm and HSFR algorithm. According to Corollary 1, the fault tolerance of our proposed PRG-SFR algorithm is $\binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{k}$.

Wu *et al.* [37] proposed the n -hypercube-based social feature routing (HSFR) with disjointed paths. According to the definition of fault tolerance, the fault tolerance of HSFR algorithm is n .

Because $\binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{k} > n$ where $k \geq 2$, PRG-SFR algorithm can tolerate more faulty nodes than HSFR algorithm.

Fig. 10 plots two comparisons of fault tolerance between PRG-SFR algorithm and HSFR algorithm. It can be seen that the fault tolerance of PRG-SFR algorithm increases notably faster than HSFR algorithm as n increases. Therefore, PRG-SFR algorithm has stronger fault tolerance than HSFR algorithm. The fault tolerance has the tendency to further grow as the number of key features increases. Note that, in the real scenario, the number of key features can not be very large, thus the fault tolerance can be controlled.

Secondly, we compare the forwarding number of PRG-SFR algorithm and HSFR algorithm. According to Corollary 2, the upper bound of forwarding number of PRG-SFR algorithm is $FN \leq \lceil \frac{l}{k} \rceil + 2$ when $k < l$ or $FN \leq 3$ when $k \geq l$.

In the traditional hypercube-based routing, each forwarding can only change one feature at one time. Therefore, when a packet holder meets another user who has more than one feature distance away, the packet is not forwarded to that user although the user

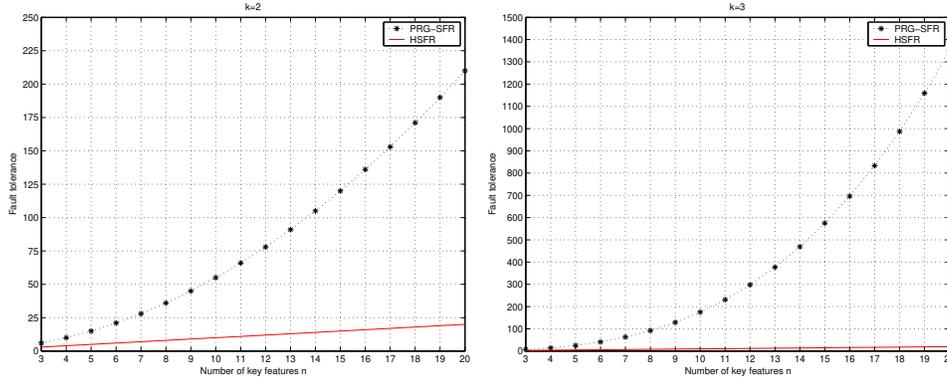


Figure 10: The fault tolerance of PRG-SFR algorithm and HSFR algorithm.

is closer to the destination. The shortcut is a prefix of the coordination sequence. Wu *et al.* [37] added shortcuts along the n disjointed multi-paths through HSFR algorithm. Moreover, the maximal length of all disjointed paths between u and v using HSFR algorithm without shortcuts is calculated as $L_l = l + 2$. The maximal length of all disjointed paths between u and v using HSFR algorithm with shortcuts is calculated as $L_l = l - k + 3$, when one k -hop shortcut is included. According to the definition of forwarding number, the upper bound of forwarding number of HSFR algorithm is $FN \leq l - k + 3$.

Because $\lceil \frac{l}{k} \rceil + 2 \leq l - k + 3$ where $2 \leq k \leq l$, the forwarding number of disjointed paths in PRG-SFR algorithm is less than HSFR algorithm with/without shortcuts.

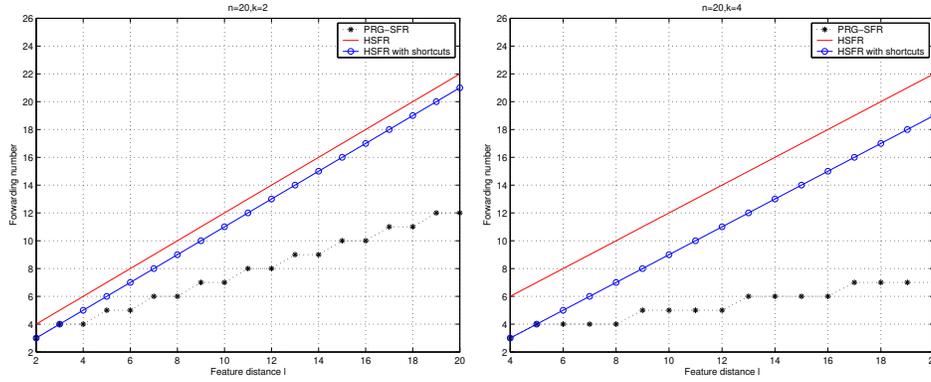


Figure 11: The forwarding number of PRG-SFR algorithm and HSFR algorithm with/without shortcuts.

Fig. 11 plots two comparisons of forwarding number using PRG-SFR algorithm and HSFR algorithm with/without shortcuts for various sets of n , k and l . It shows that PRG-SFR algorithm has smaller forwarding number than that of HSFR algorithm with/without shortcuts when $l \geq k$. The forwarding number have the tendency to further grow as the feature distance l increases. Note that, in the real scenario, the feature distance l ($l \leq n$) can not be very large, thus the forwarding number can be controlled.

Thirdly, we compare the transmission time of PRG-SFR algorithm and HSFR algorithm. According to Theorem 3, the transmission time between the source user u and the destination user v is

$$T_l = \begin{cases} T^1(1), & \text{if } l = 1, \\ T^1(2) + T^1(l - 2), & \text{if } k \geq l, \\ T^1(1) + \lfloor \frac{l}{k} \rfloor * T^1(k) + T^1(p - 1), & \text{if } k < l, \end{cases}$$

where $l = mk + p$ ($1 \leq p \leq k$). Hence, the maximal transmission time between the source user u and the destination user v is $T_l = T^1(1) + \lfloor \frac{l}{k} \rfloor * T^1(k) + T^1(p - 1)$, where $l = mk + p$ ($1 \leq p \leq k$).

According to the definition of transmission time, the transmission time of HSFR algorithm without shortcuts is $T_l = l * T^1(1)$, and the transmission time of HSFR algorithm with shortcuts is $T_l = (l - k) * T^1(1) + T^1(k)$, when one k -hop shortcut is included. It is easy to know that the transmission time along an edge e_s ($1 \leq s \leq k$) is inversely proportional to the similarity $p(s)$. In what follows, $T^1(s)$ is independent and exponentially distributed with mean $1/p(s)$ as $T^1(s) = exp(1/p(s))$.

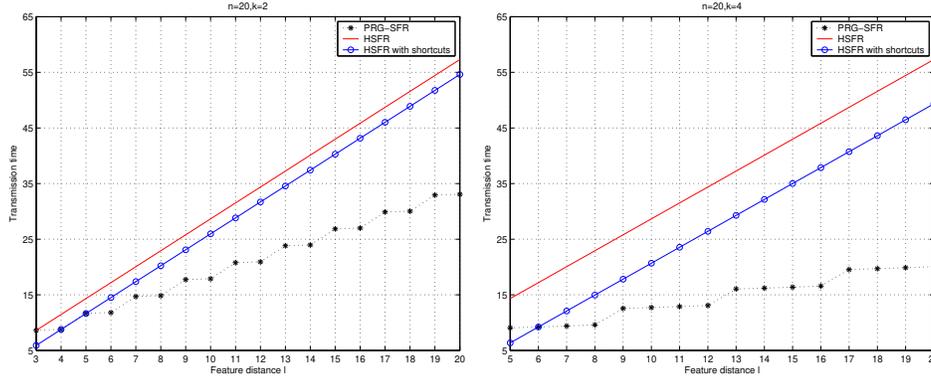


Figure 12: The transmission time of PRG-SFR algorithm and HSFR algorithm with/without shortcuts.

Fig. 12 shows two comparisons of transmission time between PRG-SFR algorithm and HSFR algorithm with/without shortcuts for various sets of n , k and l . It can be seen that the transmission time of PRG-SFR algorithm is less than that of HSFR algorithm without shortcuts in both subgraphs when $l \geq k + 1$. Also, the PRG-SFR algorithm has shorter transmission time than that of HSFR algorithm with shortcuts in both subgraphs when $l > k + 1$. Therefore, we can conclude that PRG-SFR algorithm has a lower latency than HSFR algorithm with/without shortcuts in general. The transmission time of PRG-SFR algorithm has the tendency to further grow as the feature distance l increases. Note that, in the real scenario, the feature distance l ($l \leq n$) can not be very large, thus the transmission time can be controlled.

Finally, we compare the delivery rate of PRG-SFR algorithm and HSFR algorithm.

According to Theorem 4, the delivery rate from u to v within time t is

$$DR_t^l = \begin{cases} DR_{\bar{t}}^1(1), & \text{if } l = 1, \\ DR_{\bar{t}}^1(2) \cdot DR_{\bar{t}}^1(l-2), & \text{if } k \geq l, \\ DR_{\bar{t}}^1(1) \cdot (DR_{\bar{t}}^1(k))^{\lfloor \frac{l}{k} \rfloor} \cdot DR_{\bar{t}}^1(p-1), & \text{if } k < l, \end{cases}$$

where $l = mk + p$ ($1 \leq p \leq k$) and \bar{t} is the average time needed to pass each relay group.

If a user wants to deliver a packet within time t based on HSFR algorithm with/without shortcuts, assume that the relay for a packet to pass one node on the path from the source to destination is \bar{t} . According to the definition of delivery rate, the delivery rate of the packet based on HSFR algorithm with/without shortcuts is computed as $DR_t^l = (DR_{\bar{t}}^1(1))^l$, when no shortcuts are applied, and $DR_t^l = (DR_{\bar{t}}^1(1))^{l-k} * DR_{\bar{t}}^1(k)$, when one k -hop shortcut is included. It is easy to know that the delivery rate along an edge e_s ($1 \leq s \leq k$) is directly proportional to the similarity $p(s)$. Recall that the function $DR_t^1(s)$ is a monotone decreasing function of s within time t . In what follows, assume that $DR_t^1(s)$ is independent and exponentially distributed with mean $1 - 1/p(s)$ as $DR_t^1(s) = \exp(1 - 1/p(s))$.

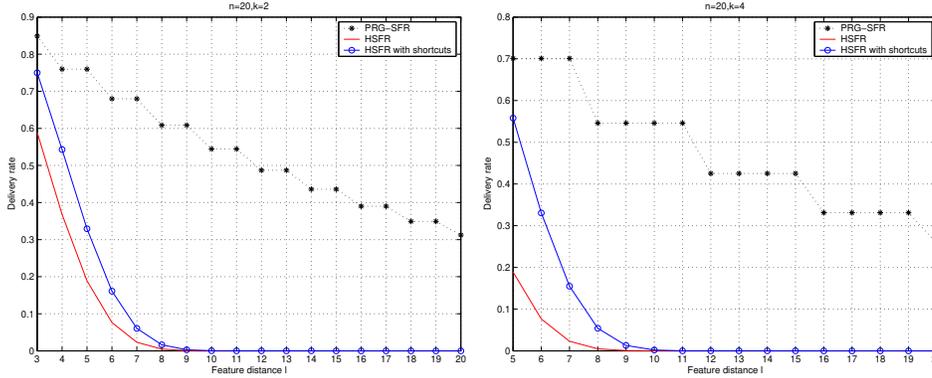


Figure 13: The delivery rate of PRG-SFR algorithm and HSFR algorithm with/without shortcuts.

Fig. 13 shows two comparisons of delivery rate of PRG-SFR algorithm and HSFR algorithm with/without shortcuts for various sets of n , k and l . It can be seen that the delivery rate of PRG-SFR algorithm is higher than that of HSFR algorithm in both subgraphs when $l \geq k + 1$. Therefore, the delivery rate of PRG-SFR algorithm is larger than that of HSFR algorithm. The delivery rate of PRG-SFR algorithm has the tendency to further decrease as the feature distance l increases. Note that, in the real scenario, the feature distance l ($l \leq n$) can not be very large, thus the delivery rate can be controlled.

Wu *et al.* [34,36,37] compared the performance of hypercube-based routing with spray-and-wait algorithm, spray-and-focus algorithm, and seek-active algorithm in terms of fault tolerance, forwarding number, transmission time and delivery rate. They found that HSFR algorithm performs better than these three schemes. Therefore, we can conclude that our proposed PRG-SFR algorithm is more competitive than HSFR algorithm with/without shortcuts, spray-and-wait algorithm, spray-and-focus algorithm, and seek-

active algorithm in terms of fault tolerance, forwarding number, transmission time and delivery rate.

Since smaller forwarding number implies that fewer intermediate users are required, and lower transmission time implies that the data can be delivered more quickly, according to the numerical analysis, the forwarding cost based on PRG-SFR algorithm is lower in fault tolerance, forwarding number, transmission time and delivery rate.

In this section, only the numerical analysis is presented, and the real data has not been used in this paper. It is very difficult to obtain the real data. However, we showed higher performance of PRG-SFR algorithm than that of HSFR algorithm proposed by Wu *et al.* [34, 36, 37] from four aspects mathematically. Also, it is important to note that, in the real scenario with real nodes, the results match the simulation results as shown in Wu *et al.* [34, 36, 37]. Moreover, in our future work, we will cooperate with large companies to obtain real data and make further statistical analysis.

6 Extension

In the previous sections, the multi-path routing based on (n, k) -hypercube is discussed. Obviously, this is scalable. This routing scheme is extended to $(n_{\{m_1, \dots, m_n\}}, k)$ -hypercube-based routing, with multiple distinct values in each dimension without compression (see Fig. 14).

The multi-path routing algorithm based on $Q_{n_{\{m_1, \dots, m_n\}}, k}$ is similar to PRG-SFR algorithm. Although we do not provide the details here, there does exist a more tedious algorithm when the size of the model is larger. In order to overcome this difficulty, assume that all nodes which differ in k features are served as a *clique*. The basic scheme is extended from a clique to more cliques. It is easy to see each clique forms a complete subgraph (see Fig. 14).

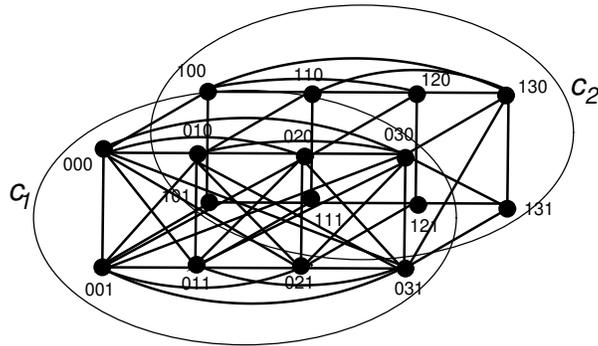


Figure 14: All nodes which differ in two features are served as a clique in a $(3_{\{2,4,2\}}, 2)$ -hypercube.

Fig. 14 shows two cliques C_1 and C_2 , where

$$C_1 = \{0 * * \mid * \text{ is a wild card for } 0, 1, 2, \text{ or } 3 \text{ and } * \text{ is also a wild card for } 0 \text{ and } 1\}$$

and

$C_2 = \{1 * * \mid * \text{ is a wild card for } 0, 1, 2, \text{ or } 3 \text{ and } * \text{ is also a wild card for } 0 \text{ and } 1\}$.

Each pair of nodes in a clique is directly connected. We treat each clique as a node, so the size of the model is reduced and simplified. Hence, we overcome the above mentioned difficulty.

7 Conclusion

This paper proposes some properties of the priority relationship and priority relation graph (PRG) of MSN. According to these properties, we propose the PRG-SFR algorithm to obtain disjointed multi-paths between any two users. The PRG-SFR algorithm is a feature matching process where the feature distance between the source and the destination are resolved step-by-step. We analyze the PRG-SFR algorithm from four aspects: fault tolerance, forwarding number, transmission time and delivery rate. The theoretical and numerical analysis results show that our proposed PRG-SFR algorithm improves the data delivery in terms of the above four aspects, compared with the hypercube-based routing algorithm (HSFR) with/without shortcuts, spray-and-wait algorithm, spray-and-focus algorithm, and seek-active algorithm.

There are two directions that can be done in the future. In MSN, we can find disjointed multi-paths by PRG-SFR algorithm from this paper. However, there exist malicious users which will destroy the paths that contain them. We need to locate the paths on which the malicious users are. We plan to use the system-level diagnosis model [22] combing the structural properties of PRG to detect malicious users. Also, we would like to define users' trustworthiness recursively associated to the distribution and influence of malicious users on disjointed multi-paths obtained by PRG-SFR algorithm. Furthermore, we aim to find out the best path for transmission, which has highest trustworthiness, among those paths without malicious users.

On the other hand, if the transmission paths between source user and destination user are known in MSN, they may be attacked. We need to protect the information of transmission paths from attackers. According to the PRG-SFR algorithm, we can find disjointed multi-paths between any two users. In other words, each of these disjointed paths may be used for transmission. Therefore, the attackers do not know the exact paths for transmission. Hence, we will propose the anonymity algorithm in MSN based on this idea.

Acknowledgment

The authors would like to thank the anonymous referees for their critical review and valuable comments. This work was supported by the National Natural Science Foundation of China (No. 61572010, No. 61702100, No. 61702103 and No. 61771140), China Post-doctoral Science Foundation (No. 2017M612107 and No. 2018T110636), Natural Science Foundation of Fujian Province (No. 2017J01738), Fujian Normal University Innovative Research Team (No. IRTL1207), Fujian University of Technology (No. GY-Z17008), Education Department of Fujian Province (No. JAT170397).

References

- [1] E. Altman, A. P. Azad, T. Başar, F. D. Pellegrini, Combined optimal control of activation and transmission in delay-tolerant networks, *IEEE/ACM Transactions on Networking* 21 (2) (2013) 482–494.
- [2] E. Athanasopoulou, L.X. Bui, T. Ji, R. Srikant, A. Stolyar, Back-pressure-based packet-by-packet adaptive routing in communication networks, *IEEE/ACM Transactions on Networking* 21 (1) (2013) 244–257.
- [3] C. Boldrini, M. Conti, A. Passarella, BContentplace: social-aware data dissemination in opportunistic networks, in *Proc. ACM Int. Symp. Model. Anal. Simul. Wireless Mobile Syst.* (2008) 203–210.
- [4] E. Bulut, B. K. Szymanski, Exploiting friendship relations for efficient routing in mobile social networks, *IEEE Transactions on Parallel and Distributed Systems* 23 (12) (2012) 2254–2265.
- [5] R. Cohen, D. Raz, Cost-effective resource allocation of overlay routing relay nodes, *IEEE/ACM Transactions on Networking* 22 (2) (2014) 636–646.
- [6] P. Dhakan, R. Menezes, The role of social structures in mobile ad-hoc networks, in *Proc. ACM Southeast Regional Conf.* (2005) 59–64.
- [7] A. P. N. Eagle, D. Lazer, Inferring social network structure using mobile phone data, in *Proc. of the National Academy of Sciences* 106 (36) (2009) 274–278.
- [8] W. Gao, G. Cao, T. L. Porta, J. Han, On exploiting transient social contact patterns for data forwarding in delay-tolerant networks, *IEEE Transactions on Mobile Computing* 12 (1) (2013) 151–165.
- [9] L. Gao, M. Li, A. Bonti, W. Zhou, S. Yu, Multidimensional routing protocol in human-associated delay-tolerant networks, *IEEE Transactions on Mobile Computing* 12 (11) (2013) 2132–2144.
- [10] W. Gao, Q. Li, B. Zhao, G. Cao, Multicasting in delay tolerant networks: A social network perspective, in *Proc. ACM Int. Symp. Mobile Ad Hoc Netw. Comput. (MobiHoc)* (2009) 299–308.
- [11] W. Gao, Q. Li, B. Zhao, G. Cao, Social-aware multicast in disruption-tolerant networks, *IEEE/ACM Transactions on Networking* 20 (5) (2012) 1553–1566.
- [12] K.-I. Goh, E. Oh, B. Kahng, D. Kim, Betweenness centrality correlation in social networks, *PHYSICAL REVIEW E* 67 (2003) 017101-1–017101-4.
- [13] M. C. Gonzalez, C. A. Hidalgo, A.-L. Barabasi, Understanding individual human mobility patterns, *Nature* 453 (7196) (2008) 779–782.
- [14] S. Ioannidis, A. Chaintreau, L. Massoulie, Optimal and scalable distribution of content updates over a mobile social network, in *Proc. IEEE INFOCOM* 2009.

- [15] N. Kayastha, D. Niyato, P. Wang, E. Hossain, Applications, architectures, and protocol design issues for mobile social networks: a survey, *Journal: Proceedings of The IEEE-PIEEE* 99 (12) (2011) 2130–2158.
- [16] S. Kosta, A. Mei, J. Stefa, Large-scale synthetic social mobile networks with SWIM, *IEEE Transactions on Mobile Computing* 13 (1) (2014) 116–129.
- [17] C.-H. Lee, D. Y. Eun, On the forwarding performance under heterogeneous contact dynamics in mobile opportunistic networks, *IEEE Transactions on Mobile Computing* 12 (6) (2013) 1107–1119.
- [18] D. Li, Y. Li, J. Wu, S. Su, J. Yu, ESM: efficient and scalable data center multicast routing, *IEEE/ACM Transactions on Networking* 20 (3) (2012) 944–955.
- [19] Q. Li, S. Zhu, G. Cao, Routing in socially selfish delay tolerant networks, in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)* (2010) 1–9.
- [20] L. Lin, L. Xu, S. Zhou, W. Wu, The social feature-based priority relation graph of mobile social networks, *2014 IEEE 17th International Conference on Computational Science and Engineering* (2014) 1921–1926.
- [21] L. Lin, L. Xu, S. Zhou, Y. Xiang, Trustworthiness-hypercube-based reliable communication in mobile social networks, *Information Sciences* 369(C) (2016) 34–50.
- [22] L. Lin, S. Zhou, L. Xu, and D. Wang, The extra connectivity and conditional diagnosability of alternating group networks, *IEEE Transactions on Parallel and Distributed Systems* 26 (8) (2015) 2352–2362.
- [23] H. Liu, H. Motoda, *Feature selection for knowledge discovery and data mining*, Springer (1998).
- [24] C. Liu, J. Wu, Practical routing in a cyclic Mobispace, *IEEE/ACM Transactions on Networking* 19 (2) (2011) 369–382.
- [25] R. Lu, X. Lin, X. S. Shen, Spring: A social-based privacy preserving packet forwarding protocol for vehicular delay tolerant networks, in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)* (2010) 1–9.
- [26] T. T. Mapoka, S. J. Shepherd, R. A. Abd-Alhameed, A new multiple service key management scheme for secure wireless mobile multicast, *IEEE Transactions on Mobile Computing* 14 (8) (2015) 1545–1559.
- [27] A. Mei, G. Morabito, P. Santi, J. Stefa, Social-aware stateless forwarding in pocket switched networks, in *Proc. of IEEE INFOCOM* (2011) 251–255.
- [28] M. Musolesi, C. Mascolo, CAR: context-aware adaptive routing for delay-rolerant mobile networks, *IEEE Transactions on Mobile Computing* 8 (2) (2009) 246–260.
- [29] M. E. J. Newman, Juyong Park, Why social networks are different from other types of networks, *PHYSICAL REVIEW E* 68 (2003) 036122-1–036122-8.

- [30] J. Ryu, L. Ying, S. Shakkottai, Timescale decoupled routing and rate control in intermittently connected networks, *IEEE/ACM Transactions on Networking* 20 (4) (2012) 1138–1151.
- [31] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, A. Chaintreau, CRAWDAD trace cambridge/haggle/imote/infocom 2006 (v.2009-05-29) (2009).
- [32] E. Talipov, Y. Chon, H. Cha, Content sharing over smartphone-based delay-tolerant networks, *IEEE Transactions on Mobile Computing* 12 (3) (2013) 581–595.
- [33] S. Venkatramanan, A. Kumar, Co-evolution of content spread and popularity in mobile opportunistic networks, *IEEE Transactions on Mobile Computing* 13 (11) (2014) 2498–2509.
- [34] Y. Wang, W.-S. Yang, J. Wu, Analysis of a hypercube-based social feature multi-path routing in delay tolerant networks, *IEEE Transactions on Parallel and Distributed Systems* 24 (9) (2013) 1706–1716.
- [35] D. J. Watts, P. S. Dodds, M. E. J. Newman, Identity and search in social networks, *Science* 296 (2002) 1302–1305.
- [36] J. Wu, Y. Wang, Social feature-based multi-path routing in delay tolerant networks, *INFOCOM, 2012 Proceedings IEEE* (2012) 1368–1376.
- [37] J. Wu, Y. Wang, Hypercube-based multi-path social feature routing in human contact networks, *IEEE Transactions on Computers* 63 (2) (2014) 383–396.
- [38] M. Xiao, J. Wu, L. Huang, Community-aware opportunistic routing in mobile social networks, *IEEE Transactions on Computers* 63 (7) (2014) 1682–1695.
- [39] J.-M. Xu, *Combinatorial theory in networks*, Science Press Beijing/China (2013).