

On Exploiting Transient Social Contact Patterns for Data Forwarding in Delay-Tolerant Networks

Wei Gao, *Member, IEEE*, Guohong Cao, *Fellow, IEEE*,
Tom La Porta, *Fellow, IEEE*, and Jiawei Han, *Fellow, IEEE*

Abstract—Unpredictable node mobility, low node density, and lack of global information make it challenging to achieve effective data forwarding in Delay-Tolerant Networks (DTNs). Most of the current data forwarding schemes choose the nodes with the best cumulative capability of contacting others as relays to carry and forward data, but these nodes may not be the best relay choices within a short time period due to the heterogeneity of transient node contact characteristics. In this paper, we propose a novel approach to improve the performance of data forwarding with a short time constraint in DTNs by exploiting the transient social contact patterns. These patterns represent the transient characteristics of contact distribution, network connectivity and social community structure in DTNs, and we provide analytical formulations on these patterns based on experimental studies of realistic DTN traces. We then propose appropriate forwarding metrics based on these patterns to improve the effectiveness of data forwarding. When applied to various data forwarding strategies, our proposed forwarding metrics achieve much better performance compared to existing schemes with similar forwarding cost.

Index Terms—Forwarding, transient contact pattern, social network, delay-tolerant network, centrality, community

1 INTRODUCTION

IN Delay-Tolerant Networks (DTNs) [13], mobile devices such as PDAs, laptops and Smartphones are only intermittently connected due to mobility and low node density. As a result, it is hard to maintain an end-to-end connection. To forward data to a destination, node mobility is exploited to let nodes (relays) physically carry data, and forward data opportunistically upon contact with others. The key problem for data forwarding in DTNs is therefore how to ensure that the data is carried by relays with the best chance to forward data to the destination within the time constraint.

The capability for a mobile node to forward data to the destination is indicated by the *data forwarding metric*. Some schemes [22], [2] propose data forwarding metrics by exploiting the stochastic node contact process based on experimental and theoretical analysis. Some others [8], [31] propose metrics based on the prediction of node mobility and its probability of contacting the destination. However, the performance of these schemes is limited due to the randomness of human mobility and thus the low prediction accuracy.

In contrast, recent research focuses on exploiting the social contact patterns of human beings for data forwarding in DTNs [9], [18], [17], because they are more reliable and less susceptible to the randomness of human mobility. Based on traces collected at various DTN scenarios such as university campus and conference sites, social contact patterns mainly include the following two perspectives: 1) *Centrality*, which indicates that some nodes are the common acquaintances of other nodes and thus have better capabilities of contacting others. 2) *Community*, which indicates that people are naturally organized into groups according to their social relations.

When social contact patterns are exploited, data forwarding is divided into two stages [18]. First, node centrality is evaluated at the global scope which includes all the nodes in the network, to ensure that data is carried and forwarded by relays with higher capability of contacting other nodes. Second, when a relay contacts a node within the same community of the destination, data is forwarded to that community. Afterward, node centrality is evaluated within the local community scope, and data is forwarded directly to the destination. In both stages, most of the current social-aware data forwarding schemes evaluate the centrality of mobile nodes based on their cumulative social contact patterns over a long period of time.

We observe that the transient social contact patterns of mobile nodes during short time periods in DTNs are usually different from their cumulative contact patterns. The relays selected by existing schemes, therefore, may not be the best choices for forwarding data with a short time constraint. Based on this observation, we improve the performance of data forwarding by exploiting the transient social contact patterns from the following perspectives.

- *Transient contact distribution*, which may be highly skewed during different time periods. For example,

- W. Gao is with the Department of Electrical Engineering and Computer Science, University of Tennessee at Knoxville, 302 Min Kao Building, 1520 Middle Drive, Knoxville, TN 37919. E-mail: weigao@utk.edu.
- G. Cao and T. La Porta are with the Department of Computer Science and Engineering, The Pennsylvania State University, IST Building, University Park, PA 16802. E-mail: {gcao, tlp}@cse.psu.edu.
- J. Han is with the Siebel Center for Computer Science, Department of Computer Science, University of Illinois at Urbana-Champaign, Room 2132, 201 N. Goodwin Avenue, Urbana, IL 61801. E-mail: hanj@cs.uiuc.edu.

Manuscript received 26 May 2011; revised 25 Sept. 2011; accepted 4 Nov. 2011; published online 15 Nov. 2011.

For information on obtaining reprints of this article, please send e-mail to: tmc@computer.org, and reference IEEECS Log Number TMC-2011-05-0277. Digital Object Identifier no. 10.1109/TMC.2011.249.

a student A may contact his classmate B frequently in the daytime but not at night, while B 's roommate C contacts B frequently at night but not in the daytime. Then, if there is data destined to B , it is better to deliver the data to A in the daytime, but to C at night. However, cumulative contact distribution cannot differentiate these two cases. Thus, we use transient contact distribution to better represent the skewness of contact distribution to improve the performance of data forwarding.

- *Transient connectivity (TC)*, which indicates that some nodes in DTNs may remain connected with each other during specific time periods to form *transient connected subnets (TCS)*. For example, a student remains connected with his classmates during the class and they form a TCS during that time period. Similarly, vehicles form a TCS when they are waiting for the traffic light at the intersection [32]. A node has "*indirect*" contacts with all nodes in a TCS, as long as it directly contacts any node in that TCS. The contact capability of mobile nodes is then increased by exploiting these indirect contacts.
- *Transient community structure*, which indicates that a node may belong to different communities during specific time periods. For example, a student A belongs to the community consisting of his classmates during daytime, but belongs to the community consisting of his roommates at night. In this case, to forward data to his classmate, A should evaluate its centrality within the local scope of his classmates' community during daytime. The characteristics of such transient community structure help us better identify the community boundary, and then evaluate node centrality at the appropriate scope.

The major contribution of this paper is twofold. First, we formulate the transient social contact patterns of mobile nodes based on experimental studies of realistic DTN traces, and the formulations of the three perspectives can be uniformly represented as a Gaussian function. Second, based on the transient social contact patterns, we develop data forwarding metrics to analytically predict the contact capability of mobile nodes with better accuracy. More specifically, we analytically evaluate node centrality within the given scope and time constraint using the transient contact distribution and transient connectivity. Then, we determine the appropriate scope for evaluating node centrality based on the transient community structure. Through extensive trace-driven simulations, we show that our approach significantly outperforms existing schemes in terms of data delivery ratio.

The rest of this paper is organized as follows: Section 2 reviews related work. Section 3 gives a brief overview on how to exploit transient social contact patterns for data forwarding. Section 4 formulates the transient social contact patterns based on experimental observations from realistic DTN traces. Sections 5 and 6 present our data forwarding approach in detail. Section 7 evaluates the performance of our approach by trace-driven simulations and Section 8 concludes the paper.

2 RELATED WORK

Research on data forwarding in DTNs originates from Epidemic routing [30] which floods the network. Later studies develop forwarding strategies to approach the performance of Epidemic routing with lower cost, which is measured by the number of data copies created in the network. While the most conservative approach [29] always keeps a single data copy and Spray-and-Focus [28] holds a fixed number of data copies, most schemes do not limit the number of data copies and forward data by comparing the nodes' data forwarding metrics. In Compare-and-Forward [10], a relay forward data to another node whose forwarding metric is higher than itself. Delegation forwarding [12] reduces the cost by only forwarding data to the node with the highest metric.

The data forwarding metric, which measures the nodes' capability of forwarding data to the destination, is independent from the data forwarding strategies mentioned above. Various metrics can be applied to the same forwarding strategy for different performance requirements, and the metrics can be either destination-specific or destination-independent. For destination-specific metrics, FRESH [10] uses the time elapsed since the last contact with the destination as the metric, and PROPHET [22] probabilistically estimates the delivery predictability of a node to the destination. More schemes use destination-independent metrics due to the lack of global information at individual nodes, and data forwarding metric evaluates the node's capability of contacting others in the network. Some schemes predict node contact capability by estimating their probabilities of meeting each other in different ways, such as the Kalman filter [8] and semi-Markov chains [31]. In some other schemes, node contact pattern is exploited for better prediction accuracy, based on the experimental [5], [20] and theoretical [4] analysis. The nodes' capability of contacting others can be predicted based on their previous cumulative contact records. For example, MaxProp [3] estimates the node contact likelihood based on the number of contacts in the past.

In all schemes mentioned above, node contact capability is predicted only based on the direct contacts among mobile nodes. Contact duration has only been analyzed for the properties of temporal network topology [6], or has been considered in case of limited channel bandwidth to determine the appropriate data transmission order [3], [2], [21]. Various weighting methods [3], [2] and the willingness of mobile users [21] are considered to maximize the bandwidth utilization during the limited contact duration. However, contact duration has never been studied for indirect contacts and transient connectivity among nodes, which can significantly increase the node contact capability.

When social contact patterns are exploited, data forwarding metric is centrality based, and is generally destination-independent. Various metrics have been proposed to calculate node centrality. In SimBet [9] and BUBBLE Rap [18], betweenness [14] is used as the centrality metric which measures the social importance of a node facilitating the communication among other nodes. Similar concept is also exploited in [24]. Gao et al. [17] propose to use Cumulative Contact Probability (CCP) as the centrality metric based on the cumulative node contact rates and the assumption of exponential distribution of pairwise node intercontact time. Later, Gao and Cao [16] extend CCP to

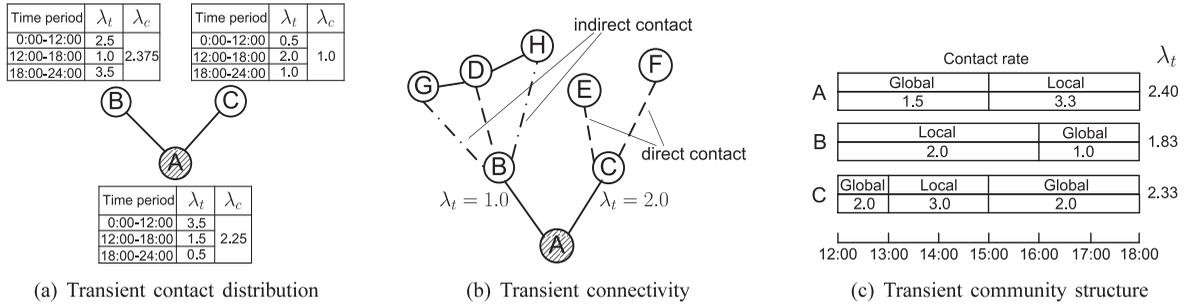


Fig. 1. Exploitation of transient social contact patterns.

the multihop scope. However, node centrality in all these schemes are evaluated based on the previous cumulative network knowledge, and none of them investigates the transient characteristics of node contact patterns. BUBBLE Rap [18] calculates node betweenness using time window, but the window length is fixed and not specific for the time constraint of data forwarding.

Social community structure in DTNs, on the other hand, is usually used to determine the scope for evaluating node centrality, and can be detected in a fully distributed manner in various ways [19]. k -clique-based method [27] enables the detection of overlapping communities, and modularity-based method [25] works on weighted network contact graph. Based on such community detection techniques, BUBBLE Rap [18] exploited social community structures for data forwarding. Node centrality is evaluated at various scopes according to the community boundary of the destination, and data is, hence, forwarded in a hierarchical manner. However, the community structure in [18] is detected based on the cumulative node contact characteristics, and is considered as fixed during the data forwarding process. Time-varying community structure is studied in [7], but is not specifically exploited for data forwarding. Comparatively, in this paper we exploit multiple perspectives of transient social contact patterns at the fine-grained scale to improve the effectiveness of data forwarding.

3 OVERVIEW

3.1 The Big Picture

When a node i determines whether to forward its carried data to another node j , most of current forwarding strategies in DTNs can be summarized as a uniform comparison-based framework. i compares a locally maintained quantity Q_i with j 's data forwarding metric m_j , and only forward data to j if $Q_i < m_j$. Different strategies vary in Q_i they use. Compare-and-Forward [10] maintains Q_i as the data forwarding metric of node i , and in Delegation [12] Q_i is the highest forwarding metric among all relays that i has contacted. The data forwarding metric is generally independent from the forwarding strategy being used.

In this paper, we focus on developing appropriate data forwarding metric by exploiting the transient social contact patterns in DTNs. Our proposed metric is expected to provide more accurate estimation of node contact capability within the specific time period, and can be applied to various data forwarding strategies. Whenever a node i contacts another node j , our forwarding scheme can be summarized by Algorithm 1. Data d_k carried by node i is

transmitted to node j if j is d_k 's destination or within the community of its destination. Otherwise, the forwarding metrics of i and j are calculated by exploiting their transient contact patterns. Data forwarding decision is then made based on these metrics and the specific forwarding strategy being used.

Algorithm 1. DataForwarding (i, j)

```

1 for each data  $d_k$  that node  $i$  carries do
2   if  $j == Dst(d_k)$  then
3     Node  $i$  forward  $d_k$  to  $j$ 
4   else if  $Community(j) == Community(Dst(d_k))$ 
      &&  $Community(i) \neq Community(Dst(d_k))$  then
5     Node  $i$  forward  $d_k$  to  $j$ 
6   else
7     Calculate the data forwarding metrics  $m_i$  and  $m_j$ 
      for nodes  $i$  and  $j$  as described in Section 3.3
8     if  $Q_i < m_j$  then
9       Node  $i$  forwards  $d_k$  to  $j$ 

```

3.2 Network Model and Assumptions

Opportunistic node contacts in DTNs are described by the network contact graph $G(V, E)$, where the stochastic contact process between a pair of nodes $i, j \in V$ is modeled as an edge $e_{ij} \in E$. We assume that node contacts are symmetric; i.e., node j contacts i whenever i contacts j , and the network contact graph is therefore undirected.

Without loss of generality, we consider forwarding one data item to a specific destination. We assume that the data size is small, so that it can be carried by any node and can be completely transmitted during a contact. The considerations of limited node buffer and channel bandwidth are orthogonal to the major focus of this paper, because they are irrelevant with the estimations of node contact capability. More detailed discussions can be found in [15].

3.3 Basic Idea

We exploit the three perspectives of transient contact patterns described in Section 1 for developing data forwarding metric. The first two perspectives are exploited to provide more accurate estimation on the node's capability of contacting others within the given scope and time period, and the third perspective is used to determine the appropriate scope for such estimation. We consider that the time constraint for data forwarding is shorter than one day. In this case, the transient contact patterns of mobile nodes differ a lot from their cumulative contact patterns, and should be exploited for improving the effectiveness of data forwarding.

TABLE 1
Trace Summary

Trace	MIT Reality	UCSD	Infocom
Network type	Bluetooth	WiFi	Bluetooth
Number of devices	97	275	78
Number of internal contacts	114,046	123,225	182,951
Duration (days)	246	77	4
Contact detection period (secs)	120	20	120
Pairwise contact freq. (per day)	4.6	0.024	7.52
Average contact duration (hours)	0.57	10.45	0.142

The basic ideas of exploiting transient social contact patterns are illustrated in Fig. 1, where relay A at 12:00 carries data with a 6-hour forwarding time constraint, and needs to determine whether to forward the data to nodes B or C .

First, transient contact distribution can be exploited for more accurate estimation about node contact frequency. In Fig. 1a, we define λ_t^i as the transient rate of a node contacting others during the i th time period in a day, and list the transient contact rates of nodes during different time periods. The cumulative contact rate during the whole day is then calculated as $\lambda_c = \sum_{i=1}^n (T_i \cdot \lambda_t^i) / \sum_{i=1}^n T_i$, where T_i is the length of i th time period. Based on the Compare-and-Forward strategy [10], A forwards data to B but not C if λ_c is used as the forwarding metric. However, considering that the transient contact distribution of B is skewed and λ_t of B between 12:00 and 18:00 is essentially low, C is a better relay than B due to its higher λ_t during that time period.

Second, transient connectivity and subsequent indirect contacts among mobile nodes can also be exploited to improve the performance of data forwarding. As shown in Fig. 1b, although C has a higher λ_t than B , such λ_t is calculated by counting only the direct node contacts. In contrast, node D has transient connectivity with nodes G and H . By contacting D , node B has indirect contacts with G and H . Node B therefore has higher contact capability than node C and is a better relay choice. In Section 5, we will describe the details of our proposed data forwarding metric by exploiting the transient contact distribution and transient connectivity.

Third, for effective data forwarding, the data forwarding metric needs to be calculated at the appropriate scope which is determined by the transient community structure. As shown in Fig. 1c, node C will enter the destination community at 13:00, but will later leave the community at 15:00 which is far before data expires. As a result, the rate of node C contacting others needs to be evaluated at both global and local community scopes during the corresponding time periods, and the contact rate λ_t during the time period between 12:00 and 18:00 is hence different from that in Fig. 1a. Similar case also applies to nodes A and B , and we can see that neither node B nor C has better contact capability than node A . In contrast, if transient community structure is not considered, the λ_t of node C will be $(2.0 + 3.0 \times 5)/6 = 2.83$ and data will be incorrectly forwarded to C . In general, we conclude that the exploitation of transient community structure evaluates node contact capability at a more fine-grained level during the time period for data forwarding, and we will describe the details of such exploitation in Section 6.

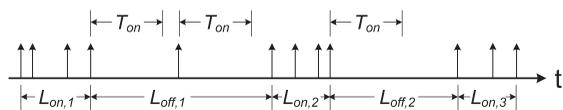


Fig. 2. Alternative appearances of on-period and off-period.

4 TRACE-BASED PATTERN FORMULATION

In this section, we formulate transient social contact patterns based on experimental observations from realistic DTN traces. These patterns are formulated with the daily period, which has been shown as the dominant period of node contact patterns in previous trace studies [11], [23], [18].

4.1 Traces

We study the transient social contact patterns on three sets of DTN traces. These traces record contacts among mobile devices moving in different environments including conference site (Infocom [18]) and university campus (MIT Reality [11], UCSD [23]). Devices in these traces are equipped with Bluetooth or WiFi interfaces. Bluetooth-enabled devices in the Infocom and MIT Reality traces periodically detect their peers nearby, and a contact is recorded when two devices move close to each other. WiFi-enabled devices in the UCSD trace search for nearby WiFi Access Points (APs) and associate themselves to the APs with the best signal strength. A contact is recorded when two devices are associated to the same AP. As summarized in Table 1, the traces differ in their scales, detection period, as well as the contact density and duration.¹

4.2 Transient Contact Distribution

For each pair of nodes, we formulate their transient contact distribution as alternative appearances of “on-period” and “off-period.” An on-period $[t_s, t_e]$ consists of a set of node contacts, which happen consecutively during time period $[t_s, t_e]$ with intercontact time shorter than T_{on} . T_{on} is a predefined threshold. This pattern formulation is illustrated in Fig. 2, where each vertical arrow indicates a contact. $L_{on,i}$ ($L_{off,i}$) denotes the length of the i th on-period (off-period). Note that an off-period may last longer than one day; e.g., a student may not contact his classmates during the entire weekend.

In our formulation, only the contact process during on-periods is considered as stable and predictable, and is exploited to predict future node contact capability. Individual contacts happened during off-periods are considered as random and unpredictable. This formulation is validated by the skewed contact distribution during different hours in a day, which is shown in Fig. 3. In Fig. 3a of the MIT Reality trace, over 50 percent of the contacts happen between 12:00 to 16:00, while only about 7 percent of the contacts happen between 22:00 to 7:00. Similar distribution is shown in Figs. 3b and 3c for the UCSD and Infocom traces. Such skewed distribution shows that most contacts happen during the on-periods which are generally shorter than the off-periods, and only few contacts can be found during the off-periods at random.

1. For any contact with the same starting and ending time, we set its duration to be a half of the detection period.

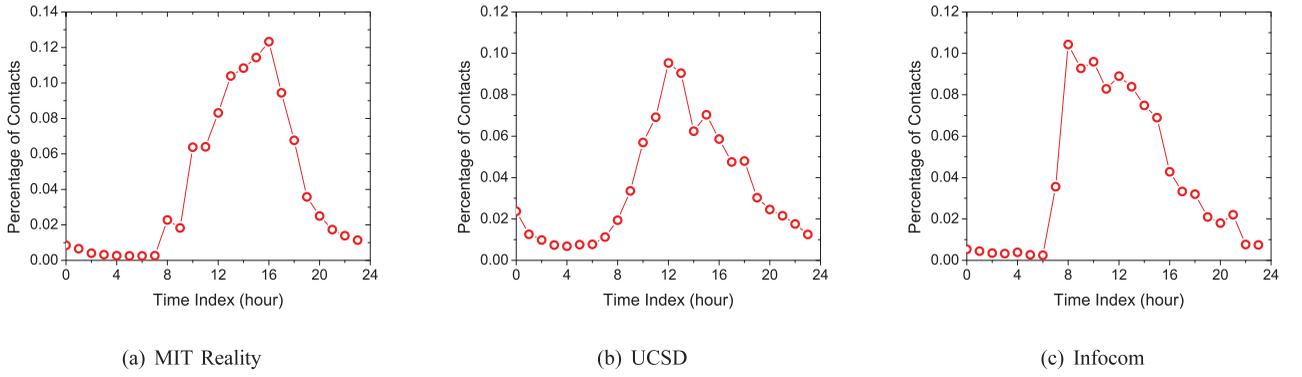


Fig. 3. Skewed distribution of node contacts.

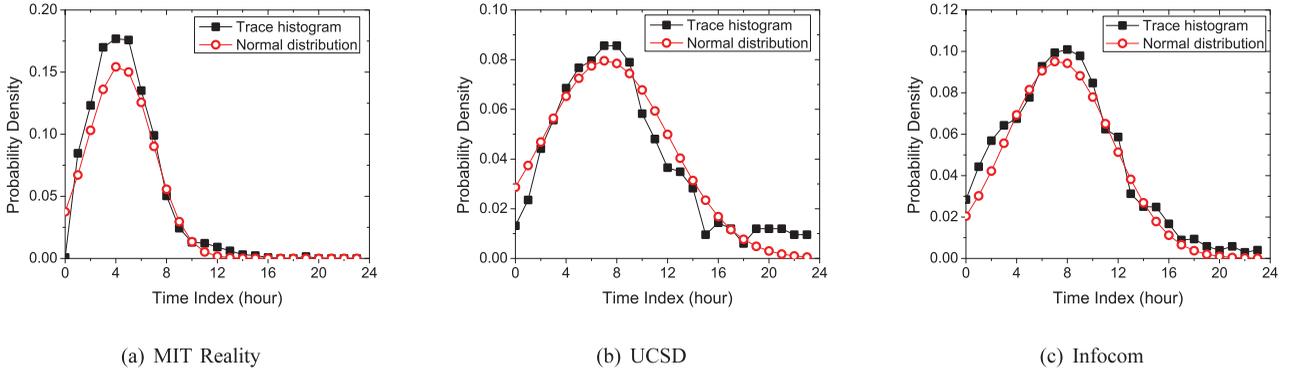


Fig. 4. Distributions of on-period lengths.

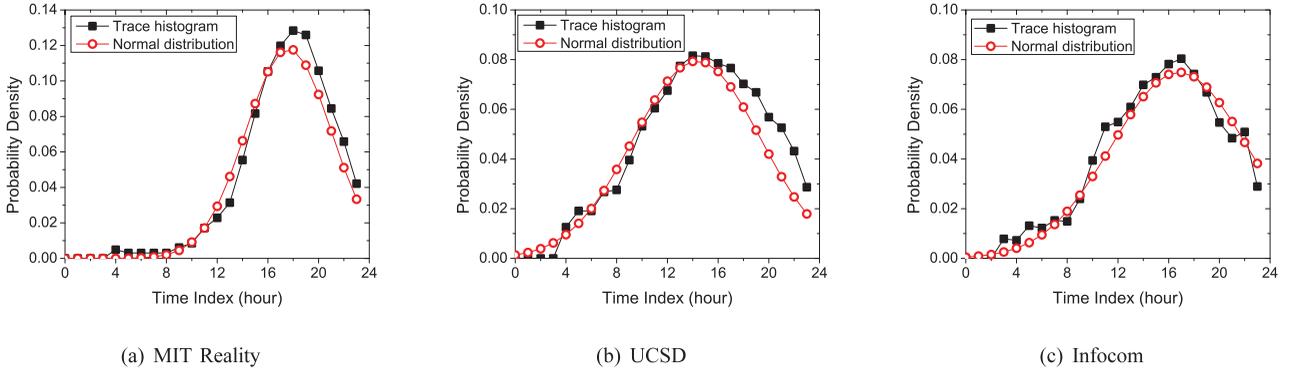


Fig. 5. Distributions of off-period lengths.

We fix $T_{\text{on}} = 8$ hours and study the distributions of on-period and off-period lengths. The length of off-period is restricted in modulus of 24 hours, and the results over all pairs of nodes are shown in Figs. 4 and 5. We have two observations. First, the distributions of on-period and off-period lengths are accurately approximated by normal distribution, using the mean and variance of the recorded lengths of on/off-periods as parameters. These parameters

are summarized in Table 2. Second, in Table 2, we have $\mu_{\text{on}} + \mu_{\text{off}} \approx 24$ hours for all traces, and over 80 percent of contacts happen during on-periods which occupy less than 30 percent of the total time. These results validate that transient contact distribution can be formulated as alternative appearances of on-period and off-period.

4.3 Transient Connectivity

The transient connectivity of a node is represented by the size of its *Transient Connected Subnet* during different time periods. The TCS of a node i during time period $[t_1, t_2]$ consists of all the nodes that have end-to-end communication links with node i during that time period.

The temporal change of TCS size is illustrated in Fig. 6 where the lines between nodes indicate contacts. At time t_1 (Fig. 6a), node A has indirect contacts² with nodes C and D

2. In the rest of this paper, we consider that a node “indirectly contact” another whenever there is end-to-end communication link between them.

TABLE 2
Numeric Parameters of on-Period and off-Period

Trace		MIT Reality	UCSD	Infocom
On-period	Mean (μ_{on})	5.3239	8.1611	7.3416
	Variance (σ_{on}^2)	6.5889	25.109	17.487
	Percent. of contacts (%)	89.312	85.624	80.372
Off-period	Mean (μ_{off})	18.631	15.334	16.8228
	Variance (σ_{off}^2)	11.387	25.191	28.337
	Percent. of contacts (%)	11.687	14.376	19.628

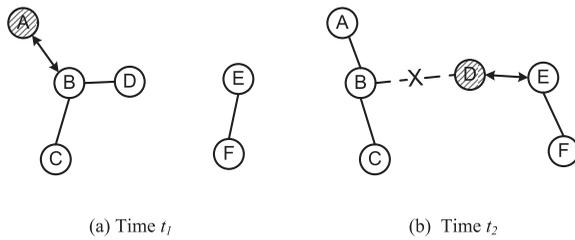


Fig. 6. Illustration of TCS size change.

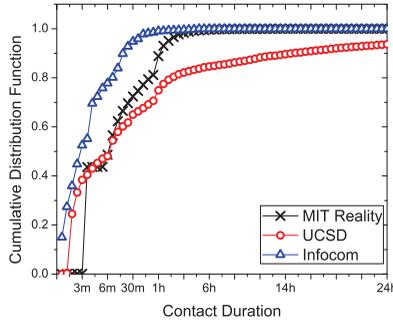


Fig. 7. Cumulative distribution function of contact durations.

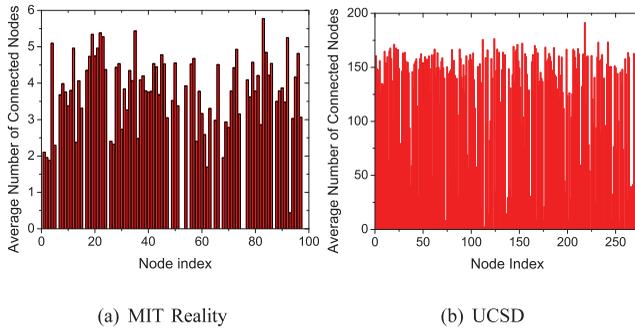


Fig. 8. The average TCS sizes.

by directly contacting node B , and A 's TCS therefore has four nodes. Later at time $t_2 > t_1$ (Fig. 6b), the TCS size of node A reduces to 3 because node D moves away and joins another TCS containing nodes E and F .

The transient connectivity in DTNs depends on the distribution of contact duration, which is shown in Fig. 7. In both MIT Reality and UCSD traces, there are a large portion of contacts with nonnegligible durations. In both traces, there are a large portion of contacts with non-negligible durations. In the MIT Reality trace, there are over 20 percent of the contacts with durations longer than 1 hour, and this percentage in the UCSD trace is around 30 percent. There are 7 percent of the contacts in the UCSD trace with durations longer than 24 hours, and the average contact duration becomes longer than 10 hours. Comparatively, contacts in the Infocom trace have shorter durations, and there are only 5 percent of the contacts with durations longer than 30 minutes.

Nodes that have contacts with long durations have good chances to maintain transient connectivity with others. To study transient connectivity, we replay the traces and calculate the TCS size of each node once every hour to take the average. The results in Fig. 8 show that the TCS sizes of over 50 percent of nodes are larger than 3 in the MIT Reality trace. The transient connectivity is more prevalent in the

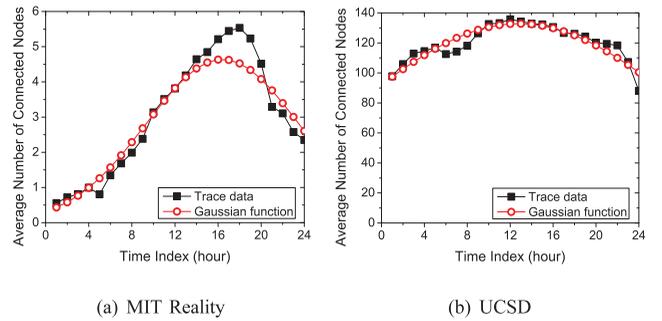


Fig. 9. The temporal distributions of the TCS sizes.

TABLE 3
Numeric Parameters on the TCS Size Distribution

Trace	MIT Reality	UCSD
A	4,6413	132.85
μ	16.406	12.795
σ^2	49.954	224.97

UCSD trace. Comparatively, transient connectivity in the Infocom trace is negligible due to the short contact durations.

Moreover, Fig. 9 shows that the average TCS size over all nodes during different time periods in a day can be accurately approximated by a Gaussian function in the form of

$$G(t) = A \cdot e^{-\frac{(t-\mu)^2}{\sigma^2}}, \quad (1)$$

and the numerical parameters in (1) are listed in Table 3. By comparing Fig. 9 with Fig. 3, it is easy to see that the TCS size during a specific time period is proportional to the amount of contacts happened during that time period.

4.4 Transient Community Structure

A pair of nodes may be within the same community for some amount of time, and leave the community afterwards. Hence, we formulate the community relationship between a pair of nodes as regular appearances of "joint-period," during which time the two nodes are within the same community.

To validate this formulation, we detect the social community structure in both traces every hour using both k -clique and modularity community detection methods [19]. The k -clique method detects a community as the union of all k -cliques (complete subgraph of size k) on the network contact graph that can be reached from each other via a series of adjacent k -cliques. The modularity method employs the Weighted Network Analysis (WNA) technique [25] based on the modularity measure proposed in [26]. In both methods, a trace-dependent community detection threshold Λ_C is applied to the transient pairwise node contact rate per hour. We consider that the social community structure only exists if there are enough node contacts in the network exceeding the given threshold Λ_C . Otherwise, the node contacts are considered as random without exhibiting any social community characteristic.

Letting S_i^t denote the set of nodes within the same community of node i at time t , we evaluate the percentage of community change of node i during time period $[t_1, t_2]$

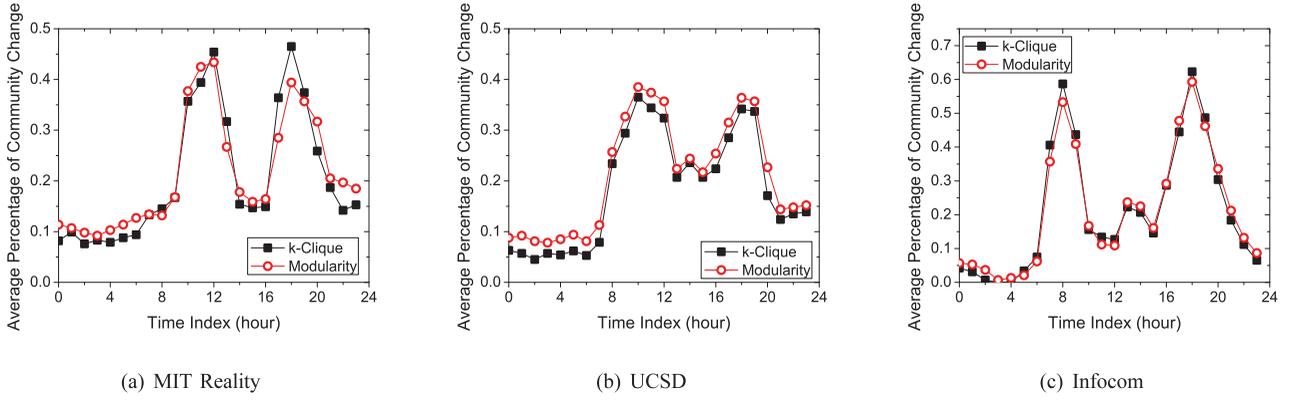


Fig. 10. Transient community structure indicated by the average percentage of community change.

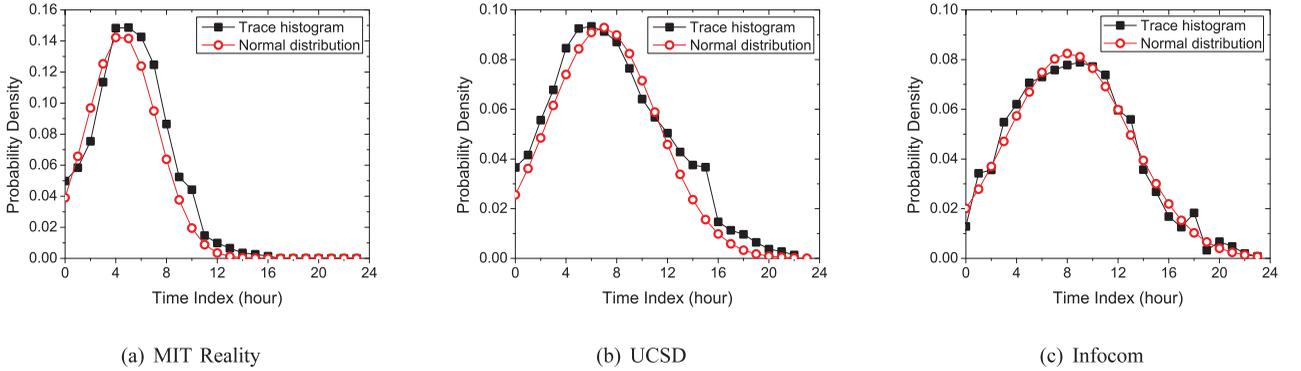


Fig. 11. Distribution of joint-period.

using the metric $1 - |\mathbf{S}_i^{t_1} \cap \mathbf{S}_i^{t_2}| / |\mathbf{S}_i^{t_1} \cup \mathbf{S}_i^{t_2}|$. The average percentage of community change over all nodes in the network during different hours in a day is shown in Fig. 10. We can see that such percentage in all traces exhibits two peaks as high as 0.6 for 3-4 hours each, and remains lower than 0.2 during the rest of time. By comparing Fig. 10 with the transient contact distribution shown in Fig. 3, we have two observations which essentially support our formulation of transient community structure. First, the percentage of community change is low during nighttime, where there are only few contacts in the network and the social community structure is occasionally noticeable. Second, this percentage during daytime is also low when the number of node contacts remains stable, and its peak values appear when the number of contacts increases or decreases rapidly.

We then study the distribution of joint-period over all pairs of nodes in the network, using the k -clique method for community detection. From the results shown in Fig. 11, we can see that the distribution of joint-period can also be accurately approximated by normal distribution, using the mean and variance of the recorded lengths of the joint-periods as parameters. These parameters are summarized in Table 4. This result is consistent with the formulation of transient contact distribution in Section 4.2, but the average length (μ_{co}) of joint-period is generally shorter than that (μ_{on}) of on-period. The major reason is that when the number of contacts increases or decreases, there is transition period before the social community structure exhibits noticeable change. Moreover, the low variance (σ_{co}^2) of normal distribution in Table 4 indicates that most nodes are

grouped into few large communities, which is consistent with the results presented in [18].

In general, we conclude that all the three perspectives of transient social contact patterns in DTNs can be represented by uniform formulations based on normal distribution or Gaussian approximation. This essentially means that transient social contact patterns exhibit stable and predictable characteristics with limited randomness. Such uniformity facilitates us to combine different perspectives of transient contact patterns together and predict node contact capability for data forwarding.

5 DATA FORWARDING METRIC

When social contact patterns are exploited for data forwarding in DTNs, the data forwarding metric of a mobile node is based on the node's centrality. In this section, we focus on accurately measuring centrality within the given scope and time constraint, by exploiting the transient contact distribution and the transient connectivity.

5.1 Overview

In our approach, we measure node centrality using the expected number of nodes that it can contact within the

TABLE 4
Numerical Parameters for the Distribution of Joint-Period

Trace	MIT Reality	UCSD	Infocom
μ_{co}	4.4688	6.8964	8.1247
σ_{co}^2	7.6412	18.4248	23.3918

TABLE 5
Notations

Notation	Explanation
t_c	Current time
t_e	The data expiration time
t_l	The time of the last contact happened before t_c
t_n	The time of the first contact to happen after t_c
t_{ls}	The starting time of the last on-period before t_c
t_{le}	The ending time of the last on-period starting before t_c
t_{ns}	The starting time of the first on-period after t_c
N_{on}	The total number of on-periods before t_c
N_{off}	The total number of off-periods before t_c
L_{on}	The total length of on-periods before t_c
C_{on}	The number of contacts happened during the past on-periods
λ	The average contact rate during the past on-periods

given time constraint based on notations shown in Table 5. It is easy to see that such metric is time-dependent. Since the metric of node i is calculated every time when a relay decides whether to forward data to i , we use C_i to indicate the metric of node i at time t_c without any loss of generality.

C_i is calculated in an accumulative manner, such that

$$C_i = \sum_{j \in \mathcal{N}} c_{ij}, \quad (2)$$

where \mathcal{N} denotes the set of nodes within the given scope for evaluating i 's centrality, and c_{ij} indicates the expected number of nodes that i can contact within the time constraint by contacting another node j . We define $c_{ii} \equiv 0$. We will focus on how to calculate c_{ij} in the rest of this section.

If we only consider direct contacts among nodes, c_{ij} is equivalent to the probability for node i to contact node j before t_e . We calculate this pairwise contact probability by exploiting the transient contact distribution of nodes i and j . The basic idea is to calculate this probability only based on contact process during on-periods.

The node contact capability is further improved by considering the indirect contacts among nodes. c_{ij} is then defined as the expected number of nodes that i can contact in j 's TCS. This expected number of nodes can be predicted based on the transient connectivity of node j . Particularly, by taking indirect contacts into account, a relay can decide whether to forward data to a node as long as they are within the same TCS. If a relay and the destination are within the same TCS, the data can be delivered to the destination immediately via multi-hop transmission within that TCS, and the data delivery ratio therefore can be improved.

5.2 Characterization of Transient Contact Patterns

Each node characterizes its transient contact patterns in real time according to formulations in Section 4. First, for each pair of nodes i and j , the parameters of their on-period and off-period are updated every time they directly contact each other. The update process is described in Algorithm 2 for the case that i contacts j at time t_c , based on notations shown in Table 5. The quantities N_{on} , N_{off} , C_{on} and L_{on} are initialized as 0 when the network starts. t_{ls} is initialized as 0 and t_{le} is initialized as a small positive real number to ensure that the first contact between i and j starts an on-period.

Algorithm 2. Updating On/Off-Periods (i, j, t_c)

```

1 if  $t_c - t_l \leq T_{on}$  then // In an on-period
2    $C_{on} \leftarrow C_{on} + 1$ 
3    $\lambda \leftarrow C_{on} / (L_{on} + t_c - t_{ls})$ 
4   if  $t_{ls} < t_{le}$  then // New on-period
5      $\mu_{old} \leftarrow \mu_{off}$ 
6      $\mu_{off} \leftarrow \frac{\mu_{on} \cdot N_{on} + t_l - t_{le}}{N_{off} + 1}$ 
7      $\sigma_{off}^2 \leftarrow \frac{N_{off}(\sigma_{off}^2 + \mu_{old}^2) + (t_l - t_{le})^2}{N_{off} + 1} - \mu_{off}^2$ 
8      $N_{off} \leftarrow N_{off} + 1$ 
9      $t_{ls} \leftarrow t_l$ 
10  if  $t_c - t_l > T_{on}$  then // In an off-period
11    if  $t_{ls} > t_{le}$  then // New off-period
12       $\mu_{old} \leftarrow \mu_{on}$ 
13       $\mu_{on} \leftarrow \frac{\mu_{on} \cdot N_{on} + t_l - t_{ls}}{N_{on} + 1}$ 
14       $\sigma_{on}^2 \leftarrow \frac{N_{on}(\sigma_{on}^2 + \mu_{old}^2) + (t_l - t_{ls})^2}{N_{on} + 1} - \mu_{on}^2$ 
15       $N_{on} \leftarrow N_{on} + 1$ 
16       $t_{le} \leftarrow t_l$ 
17       $L_{on} \leftarrow L_{on} + t_{le} - t_{ls}$ 
18     $t_l \leftarrow t_c$ 

```

Second, each node detects its TCS whenever it directly contacts another node, by broadcasting a detecting beacon message. In order to detect the TCS in a multihop range, such message is broadcasted among the nodes within the TCS, and each node having received the message acknowledges to the original sender. Transient connectivity is then updated by Gaussian curve fitting based on the recorded TCS sizes during different hours. The time needed for transmitting a beacon message is generally much shorter than the contact duration, and the transient connectivity hence can be accurately characterized. Since the broadcasting of beacon messages is only triggered by node contact events, and the sizes of beacon messages and acknowledgements are very small, such TCS detection only produces little data transmission overhead.

5.3 Pairwise Contact Probability

According to the formulation of transient contact distribution in Section 4.2, only the contact process during on-periods is stable and predictable, and the probability p_{ij} for node i to contact node j during $(t_c, t_e]$ is calculated based on the past contact records during the on-periods between node i and j . Since the time period $(t_c, t_e]$ in this paper is shorter than one day, node j can be contacted by node i during $(t_c, t_e]$ in the following two cases:

1. t_c is within an on-going on-period which continues after t_c , as illustrated in Fig. 12a.
2. t_c is within an on-going off-period, but the next on-period will start before t_e , as illustrated in Fig. 12b.

In general, a predictable contact only occurs if at least one on-period overlaps with $(t_c, t_e]$, and the occurrence probability is proportional to the overlapping length. In the first case, we estimate the remaining duration $(t_c, t_{le}]$ of the current on-period after t_c , and in the second case we predict the starting time t_{ns} of the next on-period after t_c . In contrast, node i and j will not contact during $(t_c, t_e]$ in the case illustrated by Fig. 12c, where $(t_c, t_e]$ is totally included in an off-period.

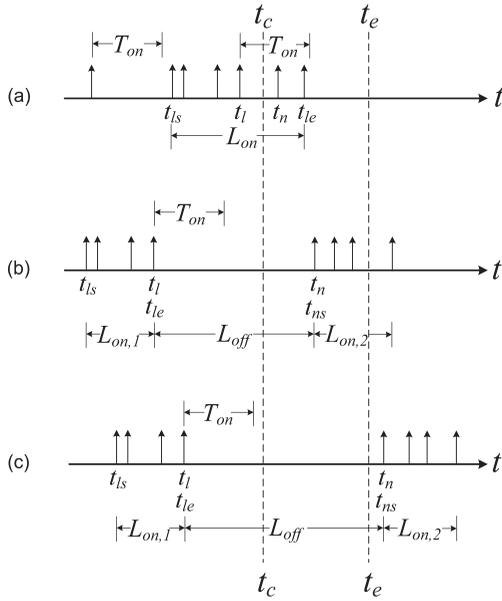


Fig. 12. Cases of predictable node contacts.

The three cases illustrated in Fig. 12 are complete and mutually exclusive for predictable contacts. Therefore, we have $p_{ij} = p_{ij}^{(1)} + p_{ij}^{(2)}$, where $p_{ij}^{(1)}$ and $p_{ij}^{(2)}$ are the contact occurrence probabilities for the cases 1 and 2, respectively. As a prerequisite, we define $p_c(t_1, t_2)$ as the probability that node i and node j contact during the time period $[t_1, t_2]$ within an on-period. Since the contact process between node i and j during on-periods is considered as stable and predictable, we assume that this contact process is a homogeneous Poisson process with the parameter λ , which is updated at real time according to Algorithm 2. As a result, we have

$$p_c(t_1, t_2) = 1 - e^{-\lambda(t_2 - t_1)}. \quad (3)$$

5.3.1 Case 1

For $\forall t_{le} \in (t_c, t_e]$, node i has the probability $p_c(t_c, t_{le})$ to contact node j . Therefore, due to the memoryless nature of Poisson process we have

$$p_{ij}^{(1)} = \int_0^T p_c(0, T) \cdot f_{on}(t) dt, \quad (4)$$

where $T = t_e - t_c$, and $f_{on}(t)$ is the Probability Density Function (PDF) of the distribution of the on-period length between node i and j , such that

$$f_{on}(t) = \frac{1}{\sqrt{2\pi}\sigma_{on}} \cdot e^{-\frac{(t+t_c-t_{ls}-\mu_{on})^2}{2\sigma_{on}^2}}. \quad (5)$$

By substituting (3) and (5) into (4), we have

$$\begin{aligned} p_{ij}^{(1)} &= \frac{1}{\sigma\sqrt{\pi}} \cdot \int_0^T (1 - e^{-\lambda t}) \cdot e^{-\frac{(t-\mu)^2}{\sigma^2}} dt \\ &= \frac{1}{2} \left(\operatorname{erf}\left(\frac{T-\mu}{\sigma}\right) + \operatorname{erf}\left(\frac{\mu}{\sigma}\right) - e^{\frac{1}{4}\lambda^2\sigma^2 - \lambda\mu} \right. \\ &\quad \left. \cdot \left(\operatorname{erf}\left(\frac{\lambda\sigma^2 + 2T - 2\mu}{2\sigma}\right) - \operatorname{erf}\left(\frac{\lambda\sigma^2 - 2\mu}{2\sigma}\right) \right) \right), \end{aligned} \quad (6)$$

where $\mu = \mu_{on} + t_{ls} - t_{cr}$, $\sigma = \sqrt{2}\sigma_{on}$, and $\operatorname{erf}(x)$ is the Gaussian error function [1].

5.3.2 Case 2

Similarly, we have

$$p_{ij}^{(2)} = \mathbb{P}(t_{le} < t_c) \cdot \int_0^T (1 - e^{-\lambda(T-t)}) \cdot f_{off}(t) dt,$$

where

$$f_{off}(t) = \frac{1}{\sqrt{2\pi}\sigma_{off}} e^{-\frac{(t+t_c-t_{le}-\mu_{off})^2}{2\sigma_{off}^2}}$$

and

$$\mathbb{P}(t_{le} < t_c) = \Phi\left(\frac{t_c - t_{ls} - \mu_{on}}{\sigma_{on}}\right)$$

with $\Phi(x)$ indicating the Cumulative Distribution Function (CDF) of standard normal distribution. Therefore,

$$\begin{aligned} p_{ij}^{(2)} &= \frac{\Phi\left(\frac{t_c - t_{ls} - \mu_{on}}{\sigma_{on}}\right)}{\sigma\sqrt{\pi}} \cdot \int_0^T (1 - e^{-\lambda(T-t)}) \cdot e^{-\frac{(t-\mu)^2}{\sigma^2}} dt \\ &= \frac{1}{2} \cdot \Phi\left(\frac{t_c - t_{ls} - \mu_{on}}{\sigma_{on}}\right) \cdot \left(\operatorname{erf}\left(\frac{T-\mu}{\sigma}\right) \right. \\ &\quad \left. + \operatorname{erf}\left(\frac{\mu}{\sigma}\right) - e^{\frac{1}{4}\lambda^2\sigma^2 - \lambda(T-\mu)} \cdot \left(\operatorname{erf}\left(\frac{\lambda\sigma^2 + 2\mu}{2\sigma}\right) \right. \right. \\ &\quad \left. \left. - \operatorname{erf}\left(\frac{\lambda\sigma^2 + 2\mu - 2T}{2\sigma}\right) \right) \right), \end{aligned} \quad (7)$$

where $\mu = \mu_{off} + t_{le} - t_{cr}$, $\sigma = \sqrt{2}\sigma_{off}$.

5.4 Exploiting Transient Connectivity

We improve the node contact capability by considering the indirect contacts among nodes within the same TCS, and we estimate c_{ij} by applying the transient connectivity of node j to the two cases analyzed in Section 5.3.

In Section 4.3, the change of the TCS size of node j during different hours in a day is described by the Gaussian function

$$N_G(t) = A_G \cdot e^{-\frac{(t-\mu_G)^2}{\sigma_G^2}}.$$

Hence, for Case 1 in Section 5.3, (4) is rewritten as

$$c_{ij}^{(1)} = \int_0^T p_c(0, T) \cdot N_G(t) \cdot f_{on}(t) dt.$$

Since transient contact distribution and transient connectivity are represented in uniform forms of Gaussian function, and the set of Gaussian functions is closed under multiplication, we have $c_{ij}^{(1)} = A \cdot p_{ij}^{(1)}$, where $p_{ij}^{(1)}$ takes the similar form as in (6), with the only difference that

$$\sigma = \sqrt{\frac{\bar{\sigma}_{on}^2 \cdot \sigma_G^2}{\bar{\sigma}_{on}^2 + \sigma_G^2}}, \quad (8)$$

$$\mu = \frac{\sigma_G^2 \bar{\mu}_{on} + \bar{\sigma}_{on}^2 \mu_G}{\sigma_G^2 + \bar{\sigma}_{on}^2}, \quad (9)$$

and

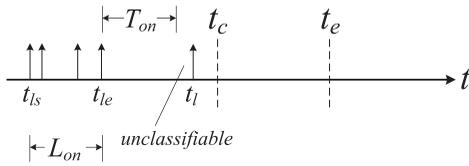


Fig. 13. Illustration of an unclassifiable contact.

$$A = A_G \cdot e^{-\frac{(\mu_G - \bar{\mu}_{on})^2}{\sigma_G^2 + \sigma_{on}^2}}, \quad (10)$$

where $\bar{\mu}_{on} = \mu_{on} + t_{ls} - t_c$, $\bar{\sigma}_{on} = \sqrt{2}\sigma_{on}$.

Similarly for Case 2 in Section 5.3, $c_{ij}^{(2)} = A \cdot p_{ij}^{(2)}$. Note that $p_{ij}^{(2)}$ takes the similar form as in (7), and the quantities $\bar{\mu}_{off} = \mu_{off} + t_{le} - t_c$ and $\bar{\sigma}_{off} = \sqrt{2}\sigma_{off}$ are used to substitute $\bar{\mu}_{on}$ and $\bar{\sigma}_{on}$ in (8), (9), and (10). Finally, we have $c_{ij} = c_{ij}^{(1)} + c_{ij}^{(2)}$.

5.5 Analysis on Prediction Error

The prediction error on node contact capability mainly comes from the node contact randomness in the following two perspectives, which may not follow the node contact patterns.

5.5.1 Unclassifiable Contacts

We cannot classify a contact into one of the two cases shown in Fig. 12, if it happens in the case shown in Fig. 13. Such a contact is the last one happened before t_c , but does not belong to the last recorded on-period. Hence, we cannot determine whether this contact starts a new on-period, or it just happens at random during the off-period starting at t_{le} . Especially when $t_c - t_{le}$ approaches μ_{off} , the calculation of contact probability for either case in Fig. 12 may produce nonnegligible prediction error.

5.5.2 Long Off-Periods

In practice, the lengths of some off-periods may exceed 24 hours. For example, two classmates may have no contact during weekend. The long off-period affects the prediction accuracy of Case 2 in Fig. 12, by postponing the next on-period to happen after t_e .

Nevertheless, the results in Section 4.2 show that the alternative appearances of on-period and off-period accurately fit the daily period, and over 80 percent of node contacts happen during on-periods. Therefore, the unclassifiable contacts and long off-periods are only occasionally found among the nodes with low contact frequency. Since in our approach data is mainly forwarded among the nodes with high contact frequency, the aforementioned prediction error can be effectively eliminated.

6 EXPLOITING TRANSIENT COMMUNITY STRUCTURE

When social community structure is exploited, node centrality is evaluated at various scopes for data forwarding. More specifically, the centrality of a mobile node is evaluated at the global scope before data is forwarded into the community of the destination, but is evaluated at the local scope of the destination community afterwards.³ As

3. The destination may belong to multiple communities simultaneously. However, the centrality of a node belonging to one of destination communities is only evaluated within the corresponding local community scope.

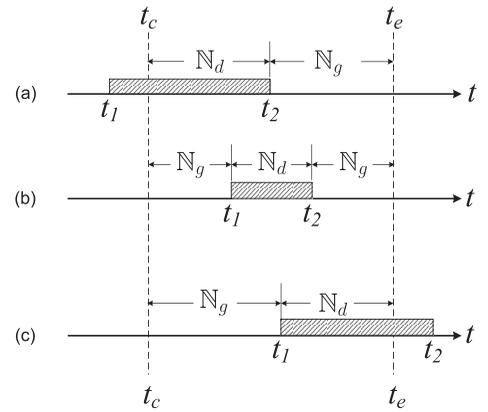


Fig. 14. Cases of overlapping between joint-period $[t_1, t_2]$ and the time constraint $(t_c, t_e]$ for data forwarding.

illustrated in Section 3, a node may be within the destination community only during specific time periods, i.e., its community relationship with the destination may change during the time period $(t_c, t_e]$ for data forwarding. As a result, the scope for evaluating node centrality during $(t_c, t_e]$ needs to be appropriately determined based on the characteristics of transient community structure.

We assume that a node i is within the destination community during the joint-period $[t_1, t_2]$ which overlaps with the time constraint $(t_c, t_e]$ for data forwarding. The three possible cases of such overlapping are illustrated in Fig. 14, where the set \mathbf{N}_g includes all nodes in the network and \mathbf{N}_d includes all nodes in the destination community at time t_1 . As illustrated in Fig. 14, letting $c_{ij}(t_1, t_2, \mathbf{N})$ denote the expected number of nodes in \mathbf{N} that i can contact during $[t_1, t_2]$ by contacting node j , the centrality of node i can be evaluated as

$$C_i = \frac{t_1 - t_c}{T} \cdot \frac{1}{|\mathbf{N}_g|} \cdot \sum_{j \in \mathbf{N}_g} c_{ij}(t_c, t_1, \mathbf{N}_g) + \frac{t_2 - t_1}{T} \cdot \frac{1}{|\mathbf{N}_d|} \cdot \sum_{j \in \mathbf{N}_d} c_{ij}(t_1, t_2, \mathbf{N}_d) + \frac{t_e - t_2}{T} \cdot \frac{1}{|\mathbf{N}_g|} \cdot \sum_{j \in \mathbf{N}_g} c_{ij}(t_2, t_e, \mathbf{N}_g), \quad (11)$$

where $T = t_e - t_c$, and $c_{ij}(t_1, t_2, \mathbf{N})$ is calculated as described in Section 5. Note that in (11), the centrality of node i is calculated in a normalized manner due to the difference in the sizes of \mathbf{N}_g and \mathbf{N}_d . Equation (11) corresponds to the case shown in Fig. 14b. When $t_1 < t_c$ or $t_2 > t_e$, the first or third term in (11) becomes zero accordingly.

In our approach, each node in the network periodically detects its community members using the distributed community detection methods proposed in [19], so as to maintain the node set \mathbf{N}_d at real time. Based on the formulation of transient community structure in Section 4.4, t_1 is fixed according to the recorded information about joint-periods with the destination in the past, and t_2 is a random variable follows normal distribution described by the PDF

$$f_{co}(t) = \frac{1}{\sqrt{2\pi}\sigma_{co}} \cdot e^{-\frac{(t - (t_1 + \mu_{co}))^2}{2\sigma_{co}^2}},$$

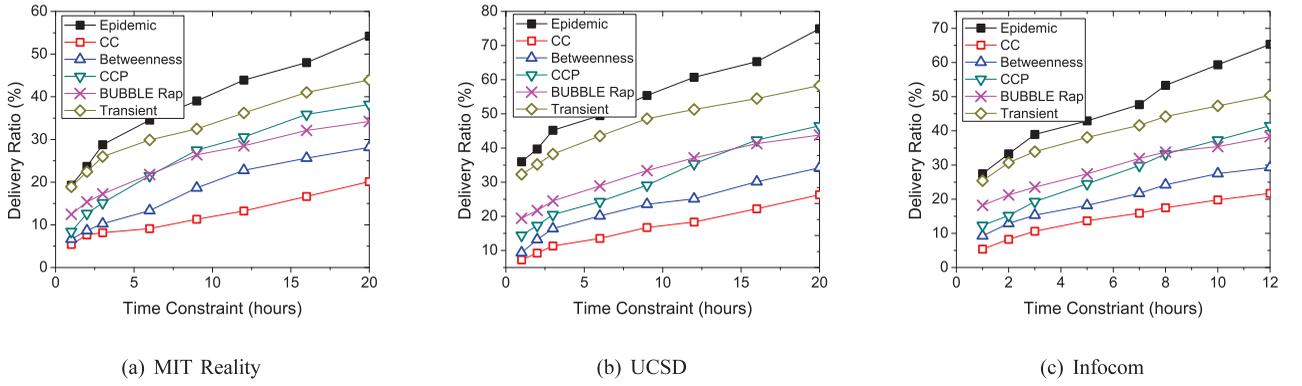


Fig. 15. Delivery ratio with the compare-and-forward strategy.

and hence the centrality of node i can be calculated as

$$\begin{aligned}
 C_i = & \frac{t_1 - t_c}{T} \cdot \frac{1}{|\mathbf{N}_g|} \cdot \sum_{j \in \mathbf{N}_g} c_{ij}(t_c, t_1, \mathbf{N}_g) \\
 & + \int_{t_1}^{t_c} f_{co}(t) \cdot \frac{t - t_1}{T} \cdot \frac{1}{|\mathbf{N}_d|} \cdot \sum_{j \in \mathbf{N}_d} c_{ij}(t_1, t, \mathbf{N}_d) dt \quad (12) \\
 & + \int_{t_1}^{t_c} f_{co}(t) \cdot \frac{t_e - t}{T} \cdot \frac{1}{|\mathbf{N}_g|} \cdot \sum_{j \in \mathbf{N}_g} c_{ij}(t, t_e, \mathbf{N}_g) dt.
 \end{aligned}$$

The integrations in (12) generally can only be solved using numeric methods, but the computation can be simplified when the overlapping between $[t_1, t_2]$ and $(t_c, t_e]$ falls into the first and third cases shown in Fig. 14.

7 PERFORMANCE EVALUATION

In this section, we compare the performance of our data forwarding approach with existing data forwarding schemes.

7.1 Simulation Setup

Our evaluations are performed on the DTN traces described in Section 4.1. We first randomize the data generation time, and then randomly pick data sources and destinations among the nodes with nonzero contacts within the time constraint.⁴ This method ensures nonnegligible delivery ratio during the short time constraint, and is applied to all the forwarding schemes for fair comparison. The transient social contact patterns are characterized in real time as described in Section 5.2, and each node detects its community structure periodically using the modularity method [19] in a distributed manner.

We evaluate the performance of our approach in terms of data delivery ratio and forwarding cost. The forwarding cost is measured by the number of data copies created in the network. Data delivery delay is not considered as long as the data can be delivered on time, and the cost for distributed community detection is negligible as suggested in [18]. Each experiment is repeated 500 times for statistical convergence.

We compare our data forwarding metric with the following existing metrics:

4. The data source/destination contact at least one another node in the network during the time constraint for data forwarding.

- **Contact counts (CC).** It is calculated cumulatively since the network starts. It is used in [3].
- **Betweenness.** The social importance of a node facilitating communication among others. It was used in SimBet [9] and BUBBLE Rap [18] for social-based data forwarding.
- **Cumulative contact probability [17].** the probability of a node contacting others based on its cumulative contact rates.

These data forwarding metrics are applied to the following forwarding strategies. The performance of Epidemic routing is also evaluated as a basic benchmark.

- **Compare-and-forward.** It has been used in FRESH [10], where a relay forwards data to the node whose forwarding metric is higher than itself.
- **Delegation forwarding [12].** each relay records the highest value of the data forwarding metric it has ever seen, and only forward data to the nodes whose forwarding metric is higher than the recorded highest value.
- **Spray.** The maximum number of data copies in the network is fixed as K . The source forwards data to K nodes, such that for $\forall j \in [1, K]$ the $(j + 1)$ th node has a higher forwarding metric than the j th node. Each relay then follows the Compare-and-Forward strategy, and deletes its local data copy after having forwarded data to another node. This strategy can generally be considered as an extension of Spray-and-Focus [28].

We also compare the performance of our scheme with BUBBLE Rap [18] in which social community structure is characterized based on the cumulative network knowledge.

7.2 Performance Comparisons

We first evaluate the data forwarding performance of our approach with the Compare-and-Forward strategy. For fairness, indirect contacts are exploited in all schemes for comparison, such that data is delivered from a relay to the destination as long as they indirectly contact each other.

The results are shown in Figs. 15 and 16. Generally speaking, our approach achieves higher data delivery ratio. When the time constraint (T) is short, the transient contact characteristics differ more from their cumulative characteristics, and the advantage of our approach is larger. In the MIT Reality trace, when $T < 3$ hours, our approach

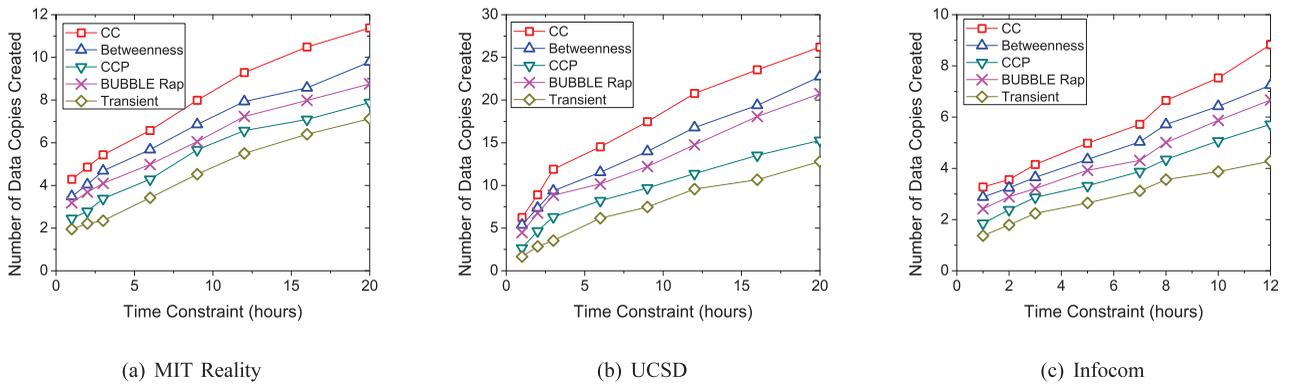
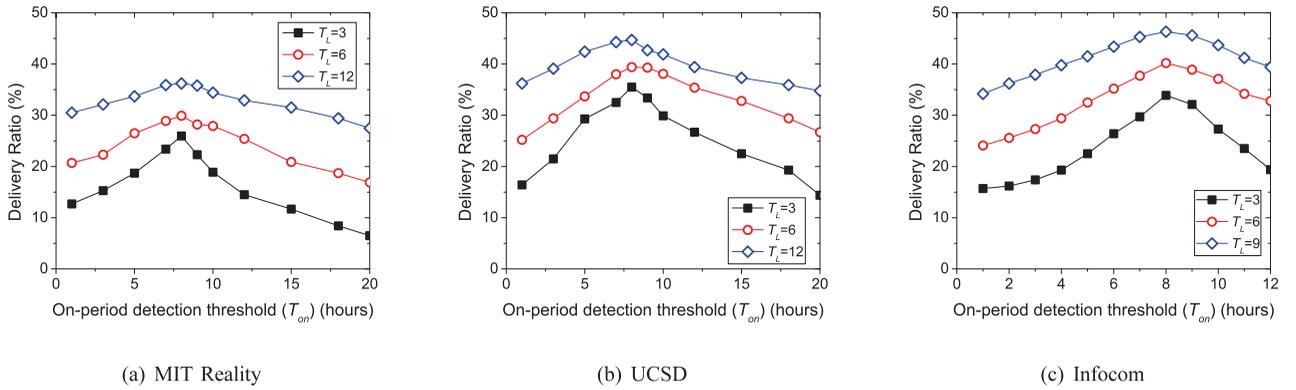


Fig. 16. Forwarding cost with the compare-and-forward strategy.

Fig. 17. Impact of different values of T_{on} .

achieves similar delivery ratio with that of Epidemic, and performs over 100 percent better than CCP. The performance of other forwarding metrics, including CC and Betweenness, is lower because they only consider cumulative effects. Similar results can also be found for the UCSD and Infocom traces. On the other hand, when T becomes longer, the transient and cumulative node contact characteristics become consistent, and the advantage of our approach decreases accordingly. Nevertheless, with a long time period of 12 hours, our approach still performs at least 20 percent better in all traces.

We notice that BUBBLE Rap outperforms other schemes due to the exploitation of social community structures, but its performance is still 20-30 percent lower than that of our approach. The reasons for such performance difference are two-fold. First, BUBBLE Rap uses Betweenness as the forwarding metric, which is unable to accurately estimate the pairwise node contact probability. Second, BUBBLE Rap considers the social community structure in the network as static, and hence may calculate node centrality based on an incorrect scope when the community structure changes over time.

Fig. 16 shows that the forwarding cost of our approach is generally more than 20 percent lower than that of other schemes. When T becomes longer, the data source and relays have higher chance to contact other nodes, and the forwarding cost increases accordingly in all schemes.

7.3 Impact of Transient Contact Distribution

In our approach, transient contact distribution is exploited to estimate the pairwise node contact probability, and

therefore has impacts on the performance of data forwarding.

7.3.1 Different Values of T_{on}

We first evaluate the impact of the on-period detection threshold T_{on} . In Section 4.2, T_{on} is fixed as 8 hours, which is validated by Fig. 17 as optimal. The delivery ratio decreases when T_{on} deviates from 8 hours, and this decrease is related to the time constraint T_L . The delivery ratio is more sensitive to the change of T_{on} when T_L is small. As shown in Fig. 17, in case of $T_L = 3$ hours, the delivery ratio decreases by 50-75 percent when the value of T_{on} changes to the smallest (1 hour) or the largest (20 hours). When T_L increases to 12 hours, the delivery ratio only decreases by 15-25 percent in similar cases.

7.3.2 Excluding Transient Connectivity

To evaluate the stand-alone impact of transient contact distribution on data forwarding, we exclude the consideration of transient connectivity and indirect contacts from our approach. More specifically, we calculate the data forwarding metric C_i of node i only based on its pairwise contact probabilities p_{ij} calculated in Section 5.3. We compare the performance of our approach to the CCP forwarding metric with the Delegation forwarding strategy.

The results are shown in Fig. 18. When data forwarding is only based on the direct contacts among nodes, the delivery ratio of both our approach and CCP decrease accordingly, due to the reduction on the nodes' capability of contacting others. The amount of such reduction depends on the characteristics of TC in different traces, which have been investigated in Section 4.3. The delivery

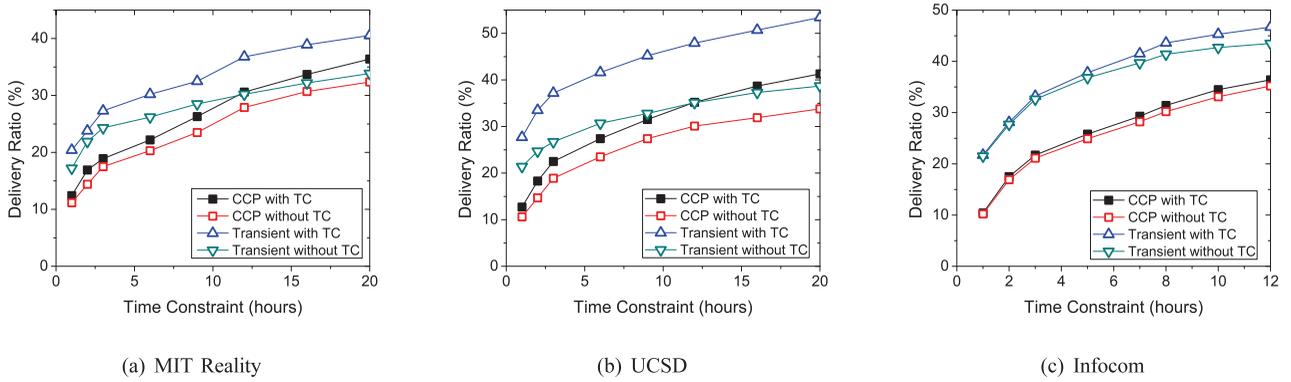


Fig. 18. Data forwarding performance with the Delegation forwarding strategy.

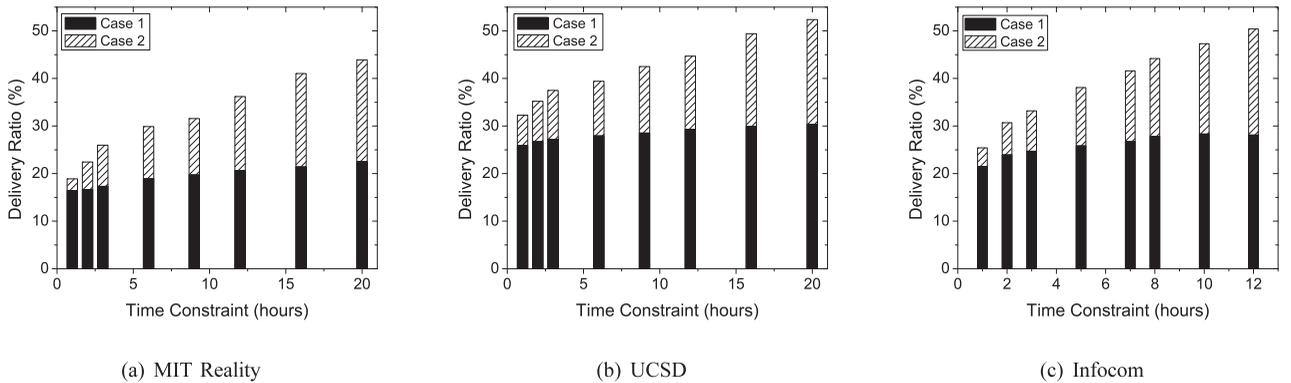


Fig. 19. Impacts of the two cases for predicting contacts on data delivery ratio.

ratio in the UCSD trace is reduced by up to 30 percent when TC is excluded, and this result is consistent with the large values of TCS size in the UCSD trace. Comparatively, the exclusion of TC only reduces the delivery ratio in the Infocom trace by less than 10 percent, due to the short contact duration in the Infocom trace.

Nevertheless, Fig. 18 shows that our approach still outperforms CCP by 30-50 percent when $T < 6$ hours. This result shows that our approach provides more accurate estimation on pairwise node contacts, compared to CCP which is based on the cumulative node contacts.

7.3.3 Two Cases for Predicting Contacts

In Section 5.3, future node contacts are predicted in two cases, and we evaluate the impact of each case on the performance of data forwarding. Whenever a relay contacts the destination, we classify this contact into one of the two cases.

The results are shown in Fig. 19. When T is short (< 3 hours), over 70 percent of successful data delivery benefits from contacts of Case 1, which indicates that an ongoing on-period will continue in the future. The accuracy of predicting the future on-period occurrences, which corresponds to Case 2, is generally low during such a short time period. When the time constraint increases, the prediction accuracy of Case 2 can be significantly improved, and therefore the importance of the two cases on successful data delivery is gradually balanced.

7.4 Impact of Transient Community Structure

In this section, we evaluate the impact of exploiting transient community structure on data forwarding. When

the scope of evaluating node centrality is determined by the static community structure, Fig. 20 shows that the delivery ratio of our approach decreases by up to 20 percent. This decrease is negligible when T is small, but quickly becomes larger when T increases. The major reason is that data is mainly forwarded within the local community when T is small, and hence node centrality is only evaluated with the local community scope.

We further evaluate the impact of different community detection methods, including both k -clique and modularity methods proposed in [19], on data forwarding using the MIT Reality trace. For the k -clique method, we set $k = 4$ as suggested by [18] and set the community detection threshold as 20 percent of the highest pairwise contact rate. For the modularity method, we use the weighted modularity which is extended from [26]. The results are shown in Fig. 21. In general, the two community detection methods lead to similar level of forwarding cost, but the modularity method is more effective in improving the data delivery ratio. Such difference is negligible when T is shorter, but can be up to 15 percent when T is large.

We also evaluate the impact of exploiting transient community structure on different data forwarding strategies as shown in Fig. 22. We set $K = 5$ for the Spray strategy. In general, Delegation has the best cost-effectiveness when transient community structure is exploited, because it achieves similar delivery ratio with that of Compare-and-Forward and has much lower forwarding cost. Comparatively, Spray is less effective because the allowed K relays cannot be satisfied in most cases due to the limited node contact capability.

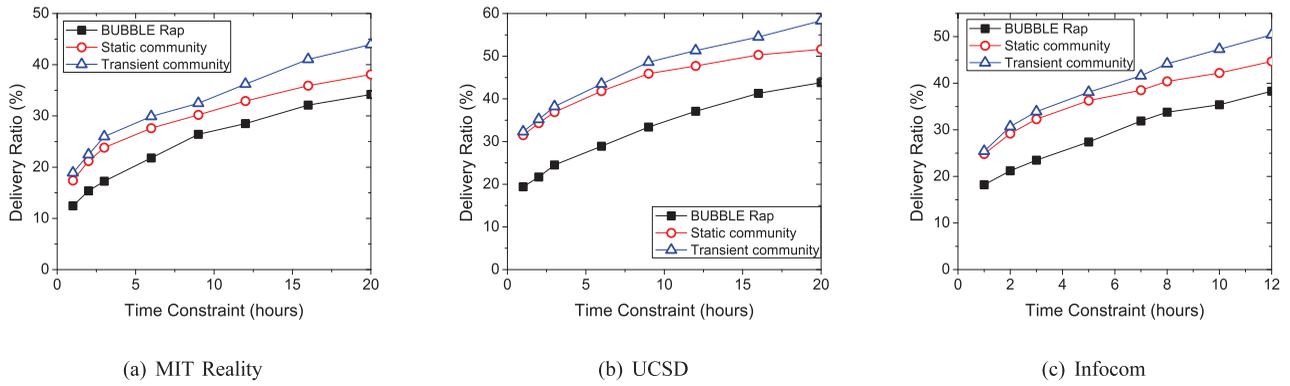


Fig. 20. Impact of exploiting transient community structure on data delivery ratio.

8 CONCLUSIONS

In this paper, we propose effective forwarding metrics to improve the performance of data forwarding in DTNs, by exploiting the transient social contact patterns. We formulate these patterns based on experimental observations from realistic DTN traces, and exploit these patterns for more accurate prediction on the node contact capability. Through extensive trace-driven experiments, we show that our approach significantly improves the data delivery ratio, while keeping similar forwarding cost with existing schemes.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers whose insightful comments helped improve the presentation of this paper significantly. This work was supported in part by Network Science CTA under grant W911NF-09-2-0053.

REFERENCES

- [1] M. Abramovitz and I. Stegun, *Handbook of Mathematical Functions*. Dover, 1972.
- [2] A. Balasubramanian, B. Levine, and A. Venkataramani, "DTN Routing as a Resource Allocation Problem," *Proc. SIGCOMM*, 2007.
- [3] J. Burgess, B. Gallagher, D. Jensen, and B. Levine, "MaxProp: Routing for Vehicle-Based Disruption-Tolerant Networks," *Proc. IEEE INFOCOM*, 2006.
- [4] H. Cai and D.Y. Eun, "Crossing over the Bounded Domain: From Exponential to Power-Law Inter-Meeting Time in MANET," *Proc. ACM MobiCom*, pp. 159-170, 2007.
- [5] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of Human Mobility on Opportunistic Forwarding Algorithms," *IEEE Trans. Mobile Computing*, vol. 6, no. 6, pp. 606-620, June 2007.
- [6] A. Chaintreau, A. Mtibaa, L. Massoulie, and C. Diot, "The Diameter of Opportunistic Mobile Networks," *Proc. ACM CoNEXT*, 2007.
- [7] S.Y. Chan, P. Hui, and K. Xu, "Community Detection of Time-Varying Mobile Social Networks," *Complex Sciences*, vol. 4, pp. 1154-1159, 2009.
- [8] P. Costa, C. Mascolo, M. Musolesi, and G. Picco, "Socially Aware Routing for Publish-Subscribe in Delay-Tolerant Mobile Ad Hoc Networks," *IEEE J. Selected Areas in Comm.*, vol. 26, no. 5, pp. 748-760, June 2008.
- [9] E. Daly and M. Haahr, "Social Network Analysis for Routing in Disconnected Delay-Tolerant MANETs," *Proc. ACM MobiHoc*, 2007.
- [10] H. Dubois-Ferriere, M. Grossglauser, and M. Vetterli, "Age Matters: Efficient Route Discovery in Mobile Ad Hoc Networks Using Encounter Ages," *Proc. ACM MobiHoc*, pp. 257-266, 2003.
- [11] N. Eagle and A. Pentland, "Reality Mining: Sensing Complex Social Systems," *Personal and Ubiquitous Computing*, vol. 10, no. 4, pp. 255-268, 2006.
- [12] V. Erramilli, A. Chaintreau, M. Crovella, and C. Diot, "Delegation Forwarding," *Proc. ACM MobiHoc*, 2008.
- [13] K. Fall, "A Delay-Tolerant Network Architecture for Challenged Internets," *Proc. SIGCOMM*, pp. 27-34, 2003.
- [14] L. Freeman, "A Set of Measures of Centrality Based on Betweenness," *Sociometry*, vol. 40, no. 1, pp. 35-41, 1977.
- [15] W. Gao and G. Cao, "On Exploiting Transient Contact Patterns for Data Forwarding in Delay Tolerant Networks," *Proc. IEEE Int'l 18th Network Protocols Conf. (ICNP)*, 2010.
- [16] W. Gao and G. Cao, "User-Centric Data Dissemination in Disruption Tolerant Networks," *Proc. IEEE INFOCOM*, 2011.
- [17] W. Gao, Q. Li, B. Zhao, and G. Cao, "Multicasting in Delay Tolerant Networks: A Social Network Perspective," *Proc. ACM MobiHoc*, pp. 299-308, 2009.
- [18] P. Hui, J. Crowcroft, and E. Yoneki, "Bubble Rap: Social-Based Forwarding in Delay Tolerant Networks," *Proc. ACM MobiHoc*, 2008.
- [19] P. Hui, E. Yoneki, S. Chan, and J. Crowcroft, "Distributed Community Detection in Delay Tolerant Networks," *Proc. ACM Int'l Workshop Mobility in the Evolving Internet Architecture (MobiArch '07)*, 2007.
- [20] T. Karagiannis, J.-Y. Boudec, and M. Vojnovic, "Power Law and Exponential Decay of Inter Contact Times Between Mobile Devices," *Proc. ACM MobiCom*, pp. 183-194, 2007.

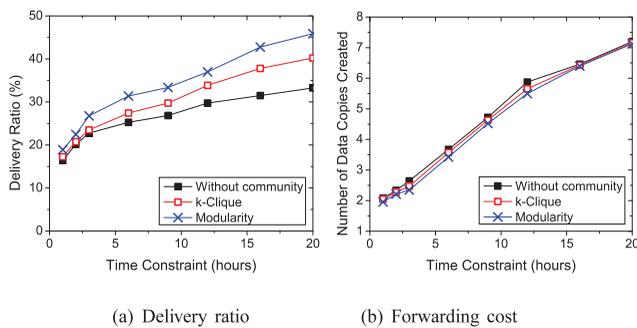


Fig. 21. Impacts of different community detection methods on the performance of data forwarding.

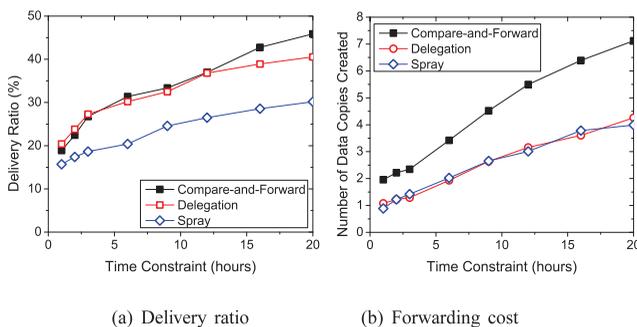


Fig. 22. Impacts of exploiting transient community structure on different data forwarding strategies.

- [21] Q. Li, S. Zhu, and G. Cao, "Routing in Socially Selfish Delay Tolerant Networks," *Proc. IEEE INFOCOM*, 2010.
- [22] A. Lindgren, A. Doria, and O. Schelen, "Probabilistic Routing in Intermittently Connected Networks," *ACM SIGMOBILE Mobile Computing and Comm. Rev.*, vol. 7, no. 3, pp. 19-20, 2003.
- [23] M. McNett and G. Voelker, "Access and Mobility of Wireless PDA Users," *ACM SIGMOBILE Mobile Computing and Comm. Rev.*, vol. 9, no. 2, pp. 40-55, 2005.
- [24] S.C. Nelson, M. Bakht, R. Kravets, and A. Harris, "Encounter-Based Routing in Dtns," *ACM SIGMOBILE Mobile Computing and Comm. Rev.*, vol. 13, no. 1, pp. 56-59, 2009.
- [25] M. Newman, "Analysis of Weighted Networks," *Physical Rev. E*, vol. 70, no. 5, p. 056131, 2004.
- [26] M. Newman and M. Girvan, "Finding and Evaluating Community Structure in Networks," *Physical Rev. E*, vol. 69, no. 2, p. 026113, 2004.
- [27] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, "Uncovering the Overlapping Community Structure of Complex Networks in Nature and Society," *Nature*, vol. 435, no. 7043, pp. 814-818, 2005.
- [28] T. Spyropoulos, K. Psounis, and C. Raghavendra, "Spray-and-Focus: Efficient Mobility-Assisted Routing for Heterogeneous and Correlated Mobility," *Proc. Pervasive Computing and Comm. Workshops*, 2007.
- [29] T. Spyropoulos, K. Psounis, and C. Raghavendra, "Efficient Routing in Intermittently Connected Mobile Networks: The Single-Copy Case," *IEEE/ACM Trans. Networking*, vol. 16, no. 1, pp. 63-76, Feb. 2008.
- [30] A. Vahdat and D. Becker, "Epidemic Routing for Partially Connected Ad Hoc Networks," Technical Report CS-200006, Duke Univ., 2000.
- [31] Q. Yuan, I. Cardei, and J. Wu, "Predict and Relay: An Efficient Routing in Disruption-Tolerant Networks," *Proc. ACM MobiHoc*, pp. 95-104, 2009.
- [32] J. Zhao and G. Cao, "VADD: Vehicle-Assisted Data Delivery in Vehicular Ad Hoc Networks," *Proc. IEEE INFOCOM*, 2006.



Wei Gao received the BE degree in electrical engineering from the University of Science and Technology of China in 2005 and the PhD degree in computer science from the Pennsylvania State University in 2012. He is currently an assistant professor in the Department of Electrical Engineering and Computer Science at the University of Tennessee, Knoxville. His research interests include wireless and mobile network systems, mobile social networks, cyber-physical systems, and pervasive and mobile computing. He is a member of the IEEE.



Guohong Cao received the BS degree from Xian Jiaotong University, China, and the MS and PhD degrees in computer science from the Ohio State University in 1997 and 1999, respectively. Since then, he has been with the Department of Computer Science and Engineering at the Pennsylvania State University, where he is currently a professor. His research interests include wireless networks and mobile computing. He has published more than 150 papers in the

areas of wireless sensor networks, wireless network security, vehicular ad hoc networks, cache management, data access and dissemination, and distributed fault-tolerant computing. He has served on the editorial board of the *IEEE Transactions on Mobile Computing* and *IEEE Transactions on Wireless Communications*, and has served on the organizing and technical program committees of many conferences. He was a recipient of the US National Science Foundation CAREER award in 2001. He is a fellow of the IEEE and the IEEE Computer Society.



Tom La Porta received the BSEE and MSEE degrees from the Cooper Union, New York, and the PhD degree in electrical engineering from Columbia University, New York. He is a distinguished professor in the Department of Computer Science and Engineering, The Pennsylvania State University (Penn State), where he is also the director of the Networking and Security Research Center. Prior to joining Penn State, he was the director of the Mobile Networking

Research Department, Bell Laboratories (Bell Labs), where he led various projects in wireless and mobile networking. He is the founding editor-in-chief of the *IEEE Transactions on Mobile Computing*. His research interests include mobility management, signaling and control for wireless networks, mobile data systems, and protocol design. He has published more than 100 technical papers and is the holder of 35 patents. He received a Thomas Alva Edison Patent Award. He is a fellow of the IEEE, the IEEE Computer Society, and Bell Labs.



Jiawei Han is a professor in the Department of Computer Science, University of Illinois, Urbana-Champaign. His research interests include data mining, data warehousing, stream data mining, spatiotemporal and multimedia data mining, biological data mining, information network analysis, text and web mining, and software bug mining, with more than 350 conference and journal publications. He has chaired or served on many program committees of international conferences and workshops. He also served or is serving on the editorial boards of *Data Mining and Knowledge Discovery*, the *IEEE Transactions on Knowledge and Data Engineering*, the *Journal of Computer Science and Technology*, and the *Journal of Intelligent Information Systems*. Currently, he serves as the founding editor-in-chief of the *ACM Transactions on Knowledge Discovery from Data* and on the board of directors for the executive committee of the ACM Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD). He has received many awards and recognitions, including the ACM SIGKDD Innovation Award in 2004 and the IEEE Computer Society Technical Achievement Award in 2005. He is a fellow of the IEEE, the IEEE Computer Society, and the ACM.

► For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.