Social-aware content dissemination through opportunistic D2D communications

Wenjuan Shao1,2 | Qingguo Shen1,3 | Liaoruo Huang1

1College of Communications Engineering, PLA Army Engineering University, Nanjing, China
2Zijin College, Nanjing University of Science and Technology, Nanjing, China
3National Mobile Communications Research Laboratory, Southeast University, Nanjing, China

Correspondence
Qingguo Shen, College of Communications Engineering, PLA Army Engineering University, Nanjing 210007, China; or National Mobile Communications Research Laboratory, Southeast University, Nanjing 211189, China.
Email: shenqg2@163.com

Funding information
National Natural Science Foundation of China, Grant/Award Number: 61271254; Research Project of Zijin College, Nanjing University of Science and Technology, Grant/Award Number: No. 2018ZKX0401004

Abstract
Last decade has witnessed rapid development in wireless network. The traffic demand is driven by enormous application data that have high-bandwidth requirement. From the operators perspective, device-to-device (D2D) communications are expected to provide an efficient way of offloading cellular traffic. Note that, in the high-density gathering cases, the redundant traffic in the cellular infrastructure can be drastically offloaded by disseminating the duplicated content through D2D communications. According to this observation, we propose a novel social-aware content dissemination scheme through D2D communications to effectively offload cellular traffic. We base our proposition on the inherent social aspects of user mobility, exploit three social structural metrics, namely, centrality, contact rate, and intercontact time. The contributions of this paper are two-fold. First, we integrate the carry-forward mechanism and multihop forwarding into our D2D content dissemination scheme to fully utilize the potential of D2D networks. Second, based on social pattern, we design a two-stage social-aware delay prediction scheme (SADP) to improve the D2D offloading utility gain. The two stages are decision-making stage and content dissemination stage. Based on real trace data set, we evaluate the proposed method through opportunistic D2D communications, which utilizes the aforementioned metrics to enhance offloading performance and D2D transmission utility gain. By comparing with other two well-known schemes, SA-Epidemic and SA-Prophet, it has been shown that our method is superior to them in terms of offloading ratio and message delay and can substantially enhance the content dissemination performance. Since SADP is designed based on a distributed insight, it ensures that it can be adaptive to mobile network. To that end, we also look forward to the future that SADP is a promising approach applied in the application scenario of 5G network.
1 | INTRODUCTION

Last decade has witnessed rapid development in wireless network. The traffic demand is driven by enormous application data, which have high bandwidth requirement. To alleviate the traditional cellular network from increasing capacity pressure, many effective technologies to improve the capacity and efficiency of wireless networks, eg, Bluetooth, Zigbee, Wi-Fi Direct, or device-to-device (D2D). Among these short-range technologies, D2D is a terminal-to-terminal network deployed by direct communication between mobile users without any need for extra infrastructure backbones. From the telecom operators’ perspective, D2D communications are expected to provide an efficient way of offloading cellular traffic, decreasing the cost (for power, location rent, deployment, and maintenance), increasing spectral efficiency, and providing robustness.

In recent years, much work has been done for D2D networks paradigm: data offloading, content dissemination, content sharing (eg, advertisement, sharing information between groups, and document distribution), coverage extension, and wearable devices networking. Data offloading is one of the most promising solutions to offload part of the traffic onto direct communication between wireless devices whenever possible. In the work of Han et al, opportunistic communications are exploited to facilitate information dissemination in the emerging mobile social networks and to reduce the amount of mobile data traffic. Mehmeti and Spyropoulos analyzed the case of tolerance data dissemination, which offers flexibility and potential gains to both the telecom operators and the users.

Task offloading is considered as a promising technology to meet the ever-increasing computation requests from a wide variety of mobile applications. The mobile device can offload those computationally intensive tasks to the nearby cloudlet instead of the centralized cloud with data centers. To reduce the energy consumption and computational cost, Guo et al proposed a data offloading and task allocation scheme for a cloudlet-assisted ad hoc mobile cloud. By using game theory, they jointly optimize the utilization of the resources from the ad hoc mobile cloud and the cloudlet under several quality-of-service (QoS) constraints. Based on network-assisted D2D collaboration, Pu et al developed an online task offloading algorithm, which can achieve superior energy consumption reduction and adapt to various of situations in terms of task type, user amount and task frequency.

Content dissemination is applied in a relatively localized scenario to utilize full potentiality of users reciprocity. Researchers originally focused on single-hop transmissions using D2D links. While one-hop D2D-based networks have been promising and energy efficient, multihop D2D-based networks, though demanded in some emerging applications, are not well studied. With the increasing networking and relaying capabilities of smart mobile devices, multihop D2D links become possible. Multihop D2D transmissions have been shown to be beneficial in QoS, energy consumption, and capacity.

In this paper, we propose a new social-aware content dissemination (SACD) scheme based on delay prediction through D2D communications. We utilize both direct D2D and carry-forward D2D to form multihop D2D communication, with the optimization goal to offload the potential mobile traffic. The former refers to the near field communication to transmit message when the nodes are in proximity. The latter refers to the situation when the nodes are not in proximity currently but are expected to meet in a certain duration. The relay nodes store and carry the message until it meets the receiver, then forward the message. With the emerging 5G networks, the base station can manage both cellular and D2D connections of mobile users. The telecom operators owing base stations control the system, schedule the transmission task, and collect each individual information (eg, location, power, and link status) to form global knowledge. Although interacting with the base stations periodically, each node can get global knowledge of other nodes. Via the control assistance by base stations, mobile users can dynamically and beneficially share the communication resources with each other.

The main contribution of this paper can be summarized in two-fold.

1. We integrate the carry-forward mechanism and multihop in the D2D content dissemination, to fully utilize the potential of D2D networks. At first, the content requesters are partly served by the eNB. Subsequently, the remaining requesters can thus be served by previous users who carry-forward the content, if they are within the D2D communication distance.

2. Base on social contact pattern, we use two-stage delay prediction to improve the D2D offloading ratio. In the first stage, a social-aware decision-making (SADM) scheme is used to arrange base-station-to-device (B2D) or D2D dissemination. In the second stage, based on the estimation of future social ability, the node with higher contact ability is authorized to schedule data dissemination task, aiming to maximize the offloading ratio.

This paper is organized as follows. Section 2 describes the ongoing related work. In Section 3, we introduce system model and the analytic expression. In Section 4, the two-stage social-aware delay prediction scheme (SADP) and its algorithm are illustrated. Experiments and results are provided in Section 5. Finally, Section 6 ends with the conclusion and future work.
2 | RELATED WORK

There are four key technical problems in designing the D2D communication underlaying cellular system: service and peer discovery, communication mode selection, spectrum resource allocation, and interference coordination and management. The concept of multihop D2D was firstly introduced in the context of delay tolerant networks, and many applicable scenarios of D2D networks have been proposed as well. Multihop D2D communication has drawn many researchers' attention in routing protocol and data dissemination.

Recent studies have shown that the integration of opportunistic networking with multihop D2D network can provide significant benefits for mobile delay tolerant services. Nunes et al designed a forwarding algorithm for D2D multihop networks based on detecting socialistic routing algorithms and evaluating their performance gain through a number of simulations and empirical measurements. Nunes et al designed a forwarding algorithm for D2D multihop networks based on detecting socialistic routing algorithms and evaluating their performance gain through a number of simulations and empirical measurements. Laha et al proposed a cluster-based multihop routing protocol to address the energy issue of increasing importance due to high energy costs of smart phones. Simulations demonstrated that it can save significant amounts of energy. A fast and energy efficient D2D multihop routing method was proposed in the work of Park. The algorithm used the geographic locations of nodes provided by the base station (eNB) in a centric mode. It could reduce end-to-end delay and energy by the cost of more progresses. In addition, it also devised a detour scheme to move around a routing hole based on geographic location information.

On the other hand, researchers also studied several relay selection and prioritization schemes to improve the potential capacity in opportunistic network. The derivation of throughput can be used to reject any excessive traffic in the admission control for real-time services. In the work of He et al, an opportunistic routing problem in multiple concurrent flows of multicast, subjected to resource constraint and route constraint, is analyzed, and a distributed algorithm for combining candidate node selection and speed distribution is put forward. The algorithm conducts flow speed distribution in an iteration mode, whether nodes are used as the candidate nodes of the flow is judged through speed distribution, and the network throughput is maximized on the premise that fairness is ensured. Niu and Cui introduced the concept of concurrent schedulable set to represent the constraints imposed by transmission conflicts and formulated a maximum concurrent flow linear programming problem, subjected to the transmission conflict constraints, to compute the optimal aggregate throughput in a given many-to-one wireless sensor network. In multiple concurrent flow wireless network, both of these research works take the advantage of concurrent schedule by selecting efficient relay nodes subjected to the constraints (transmission conflict, interference, and link capacity). This idea provides the insight for our scheme that, instead of relying on one next-hop node to forward a data packet, we can predetermine a set of candidate relays with a priority order and select the highest-priority relay that indeed receives the packet as the actual forwarder, based on the instantaneous availability.

Meanwhile, social networking is a new concept emerged in the recent years and provide a broader vision for D2D communications. While mobile networks topology changes dynamically, considering the stability of social network, D2D communications combined with social characteristic can enhance the routing performance. Iqbal and Giaccone proposed an interest based insertion policy for multihop caching to improve the performance of the overall network of caches based on the concept of social-distance. The main idea was to store only the contents, which appear to be of interest for the local user. Liu et al take the social relationships of the mobile users into the design of computation offloading scheme. With the objective to minimize the social group execution cost, they advocate game theoretic approach and propose a dynamic computation offloading scheme.

The previous works most related to ours are the works of Liu et al, Xu et al, and Coll-Perales et al. Liu et al studied the problem of multicopy multipath data dissemination based on a desired probability of satisfying delay constraint in mobile opportunistic D2D networks. Since the node participating in a D2D network are all social-aware, however, the proposed algorithm did not make use of the nodes social dimension to optimize the efficiency of D2D communication. Our scheme also investigates on multihop data dissemination within a delay probability constraint, but we lay the fundamental on nodes social contact patterns to predict transmission delay. Coll-Perales et al investigated the opportunistic forwarding to offload cellular traffic in multihop cellular networks using mobile relay. Their idea is similar to our work in exploiting the integration of opportunistic store-carry and forward mechanisms in cellular networks at the expense of some possible end-to-end transmission delays. However, ours differs from it. Besides, considering the social property, we focus on multihop downlink scenario with the goal to achieve maximum offloading effect, whereas they focused on two-hop uplink scenario with the aim to minimize the energy consumption. Similar to our work, Xu et al studied the
content delivery problem by combining both the social and physical layer information in D2D underlay networks. However, they aimed for joint peer discovery and resource allocation with power control in D2D communications by matching theory.

In summary, it can be observed that multihop routing protocol and data dissemination in D2D communication have been widely studied recently. However, recent investigations on social network show that there are social relationships, posing a strong impact on the efficiency of routing and data dissemination. Such an observation motivates us to exploit the encounter regularity in social network to predict future content dissimilation in physical network, as done in this paper.

3 | SYSTEM ANALYSIS

In this section, we formulate the delayed offloading problem and analytically derive the key metrics.

3.1 | Application scenarios

We build our application model under a real-world scenario data spot, eg, museum, concert, etc. For such highly dense areas, the majority of the traffic in cellular pertains to the download of popular content such as videos or mobile applications. In this case, Wi-Fi connections may be highly congested due to its unlicensed spectrum. Given the commonality of content, offloading it to the D2D tier can reduce the load on the cellular network’s infrastructure. An illustration of this proposed model is depicted in Figure 1.

In our model, we ignore the impact on the dissemination process of the interference from cellular network to D2D network. To be specific, the cellular interference has little impact on the delay-outage probability in D2D communication, which depends mainly on the contact interval between each user pairs. Hence, in our proposed scheme, we assume that the computation and decision-making on transmission delay will not be impacted by the cellular interference. This assumption has the rationality for its existence, whether in D2D underlay or overlay cellular network. For example, if D2D communications utilize licensed spectrum resources in D2D overlay LTE cellular networks, the D2D transmissions will be interfered by the cellular transmissions due to the reuse of the orthogonal frequency-division multiplexing time-frequency resources. However, the system can suppress the interference by use of various interference avoiding techniques: power control, multiantenna transmission techniques, and advanced network coding schemes.

From the aspect of dissemination path, two network layers exist over, which the content is disseminated. The infrastructure layer is the physical network, whereas the overlay layer is the social network. Once the users enter the radius scope of the cell, they register from the eNB. From then on, the social connections among users can be depicted from

---

**Figure 1**  Social-aware multihop content dissemination. UE, user equipment
the perspective of their encounter records, such metrics as centrality, contact rate, and intercontact time. Users’ social attributes steadily form and enrich over time. As is shown in the upper layer, they are esteemed to form a transient and localized social network. Each user equipment (UE) in the physical layer corresponds to a certain user in the social network layer. The connections between each user pairs in this layer reflect their social tie strength. Tighter social ties lead to higher probabilities to successful data transmission among UEs over the actual physical layer.

We consider the situation that a group of people temporally and spatially gathered in such location-sensitive zones. They request for the contents from content provider. Consider a cellular network with one base station (BS) and multiple users. The UEs can receive from the BS via the cellular network, or via D2D pairs using licensed spectrum resources. An example under study is depicted in Figure 1, where a group of tourists visit a museum with a time budget of several hours. Each user is willing to get the message either via the cellular network or from other carriers via D2D communication. The content provider (eg, web sites and services provider) receives these requests and sends the content to the requesters within the delay constraint.

During the sojourn, the users subscribe content requests related to the exhibitions at times. To offload the peak traffic in such a high-density data spot and alleviate the transmission pressure of the infrastructure, the telecom operators adopt a delay predicting opportunistic D2D content dissemination mechanism. Initially based on the delay-prediction based on social network, the eNB estimates each node’s delay probability. If a node’s delay probability is larger than a certain value, the eNB regards it as reachable through multihop D2D communications. Vice versa, the eNB will serve it with a cellular link directly. Subsequently, the remaining users will acquire the content from previous users who hold the content, if they are within the D2D communication distance. Let user 1 be the center user. At time slot 1 (referred as TS1), user 1 downloads the content via a cellular link, it looks up its adjacent tables and finds its D2D neighbors: user 2, user 3, and user 4. Based on the dissemination scheme, it makes transmission selections and apply the spectrum for D2D communications from the eNB, then schedules the transmission task. Noticing that at TS1, user 2, user 5, and user 6 are not in user 1’s direct D2D range. However, in the following slots, user 4 or user 3 will carry and forward to them, respectively, in a delayed manner. Thus, using group coalition, each user can get the request message before leaving the museum.

Without loss of generality, to motivate user to adopt mobile data offloading through D2D, there is also a pricing scheme that binds prices with cellular network with a discount for D2D communications.

### 3.2 Transmission utility analysis

Based on the above scenario, we propose a delay constraint analytical model \((M, D, p_0, T)\), where \(M\) is the content set, denoted as \((m_1, \ldots, m_k, \ldots, m_n)\), each content \(m_k\) with a assigned benefit value, \(v_k\). \(T\) is the corresponding delay constraint set of contents, denoted as \((\tau_1, \ldots, \tau_k, \ldots, \tau_n)\). \(D\) is a subset of users, denoted as \((D_1, \ldots, D_k, \ldots, D_n)\), with \(D_k\) being the group of users requesting for the content \(m_k\) and \(i\) is one of the users in \(D_k\). \(\Psi_i\) is the total path set from source \(s\) to destination \(i\) and \(\Psi_i^k\) is the \(k\)th path of \(\Psi_i\). \(P(D_{\Psi_i^k}(\tau_k))\) denotes the transmission probability within a given time constraint \(\tau_k\) through \(\Psi_i^k\), \(p_0\) denotes the lower bound of multihop transmission delay probability to ensure QoS requirement.

Inspired by the concept in the work of Liu et al.\(^7\) and Helmy et al.\(^29\) we introduce the definition of the delay-outage probability \(p_0\), a communication interruption probability based on multihop D2D transmission delay.

**Definition 1.** Delay-outage probability can be defined as

\[
P(D_{\Psi_i^k}(\tau_k)) > p_0.
\]  

\(\Psi_i^k\) is judged to be feasible if its transmission delay probability is greater than \(p_0\). Such a delay-outage probability acts as a decision-making factor for the operators to predetermine the transmission mode on B2D or D2D. To find the disseminating route that can provide maximum offloading under the combinational constraint of \(p_0\) and \(T\), we propose our routing algorithm based on transmission utility.

1. Transmission utility

In a multihop wireless network, the path cost and successful transmission probability of a content \(m_k\) sent from \(s\) to \(i\) are denoted as \(C_{D2D}\) and \(P(D_{\Psi_i^k}(\tau_k))\), respectively. In our delay tolerant scenario, when the source \(s\) sends out a content \(m_k\), the destination \(i\) will obtain the benefit \(v_k\) as long as it receives the desired content within the time constraint, and incur cost \(C_{D2D}\). Thus, its transmission utility is \(v_k - C_{D2D}\). Otherwise, the utility is \(0 - C_{D2D}\).\(^30,31\)
Since the successful probability is \( P(D_{\Psi_i}(\tau_k)) \) and the failure probability is \( 1 - P(D_{\Psi_i}(\tau_k)) \), the expected transmission utility is

\[
U_k = P\left(D_{\Psi_i}(\tau_k)\right) \cdot (v_k - C_{D2D}) + \left(1 - P\left(D_{\Psi_i}(\tau_k)\right)\right) \cdot (0 - C_{D2D}) \\
= P\left(D_{\Psi_i}(\tau_k)\right) \cdot v_k - C_{D2D}.
\]  

(2)

2. Link cost

According to Wu et al.,\(^3\) we assume that in both B2D and D2D wireless networks, the links cost is mainly determined by the consume of transmit energy, which is in proportion to the transmit distance.\(^3\) The link cost can be generated as following equation:

\[
c_k = \alpha(50 + \delta ||u, v||) \cdot |m_k| \quad |\Psi(u, v) \in \Psi_i^l|
\]  

(3)

\( |m_k| \) is the message size of \( m_k \), \( \delta, \gamma \) are the link cost coefficients. For D2D link, \( ||u, v|| \) is the distance between the intermediate node \( u \) and \( v \). For B2D link, \( \|u, v\| \) is the distance between the eNB and the destination node. \( \alpha \) is the amplification factor of energy consumption for B2D link, especially \( \alpha = 1 \) is for D2D link.

3. Transmission utility via B2D link

If \( m_k \) is sent from core network to \( i \) through a B2D link, it will successfully reach \( i \) within a given time constraint \( \tau_k \) with a high probability. The transmission utility can be formulated as

\[
U_{B2D}^i = v_k - C_{B2D} \quad \forall i \in \mathbb{D}_k.
\]  

(4)

4. Transmission utility via D2D links

Suppose \( m_k \) is sent from \( s \) to \( i \) via D2D links for cooperative offloading. To simplify analysis, assume that the cost per D2D link is averagely denoted as \( c_{D2D} \). \( |\Psi_i^l| \) denotes the length of path \( \Psi_i^l \). Then, if \( m_k \) successfully reaches \( i \) through \( \Psi_i^l \) within \( \tau_k \), the average transmission utility can be formulated as

\[
U(\Psi_i^l) = P\left(D_{\Psi_i}(\tau_k)\right) \cdot v_k - c_{D2D} \cdot |\Psi_i^l|.
\]  

(5)

5. D2D transmission utility gain

D2D transmission utility gain is defined as the difference between the utilities via D2D link and B2D link

\[
\Delta U_{\Psi_i^l} = U(\Psi_i^l) - U_{B2D}^i.
\]  

(6)

Making Equation (4) and Equation (5) as the substitution in Equation (6), we get

\[
\Delta U_{\Psi_i^l} = C_{B2D} - c_{D2D} \cdot |\Psi_i^l| + v_k \cdot \left( P\left(D_{\Psi_i}(\tau_k)\right) - 1 \right).
\]  

(7)

If \( \Delta U_{\Psi_i^l} > 0, \Psi_i^l \) is deemed as a utility-feasible D2D path, by which the utilities via D2D link is guaranteed. We observed from Equation (7) that there is a linear growth in utility gain with transmission probability. In addition, the transmission utility gain can be ensured under the condition if the transmission probability exceeds a certain threshold. For example, if \( v_k = 1000, C_{B2D} = 800, c_{D2D} = 100 \), the transmission probability is set to 70\%, for two-hop, three-hop, and four-hop D2D path, the D2D transmission utility gain is 300, 200, 100, respectively.

3.3 Mathematical models

Based on the analytical model \((M, \mathbb{D}, p_0, T)\) presented in Section 3.2, we formulate the cooperative offloading problem to achieve the maximum D2D transmission utility gain, by means of finding an optimal routing path in all the feasible path set for each message, subject to a desired transmission delay probability. The optimization goal can be formulated as

\[
\max \Delta U_{\Psi_i^l} \forall k \in [1, n], i \in \mathbb{D}_k, \Psi_i^l \in \Psi_i
\]  

(8)

s.t.
\[
\Delta U_{\Psi_i^l} > 0 \quad \text{and} \quad \tau_{\Psi_i^l} \leq \tau_k \quad \text{and} \quad P\left(D_{\Psi_i}(\tau_k)\right) > p_0.
\]  

(9)

(10)

(11)
In Equation (8), the optimization goal is to achieve the maximum D2D transmission utility gain in all feasible D2D path for \( m_k \). Equation (9) denotes the constraint on the transmission utility gain for the feasible D2D dissemination path. Equation (10) is the transmission delay requirement on the D2D dissemination path. Equation (11) is QoS constraint on the transmission delay probability of D2D dissemination path.

The optimization goal then derives to find the optimal path in all feasible path set (denoted as \( \phi \)). If \( \phi \) is not empty, it turns to be the find out the path with maximum D2D transmission utility gain in \( \phi \). Else, the scheme will transfer to select the B2D link instead.

Using Equation (7), the constraint on transmission utility gain in Equation (8) turns to be the constraint on the transmission delay probability

\[
P(D_{\Psi l}(\tau_k)) > 1 - \frac{C_{B2D} - c_{D2D} \cdot |\Psi_l|}{v_k}.
\]  

(12)

With a certain threshold on transmission delay probability defined in Equation (12), the profit of multihop D2D links can be guaranteed. Such a threshold acts as a decision-making factor for the operators to predetermine the transmission mode on B2D or D2D. If D2D path \( \Psi_l \)'s delay probability is greater than it, the D2D path is judged to be feasible with respect to D2D link gain. Else, it is an unfeasible D2D path.

Cater to the joint constraint of QoS and transmission utility gain, transmission delay probability can be specified as follows:

\[
P(D_{\Psi l}(\tau_k)) > p_{\text{max}}.
\]  

(13)

where

\[
p_{\text{max}} = \max \left\{ p_0, 1 - \frac{C_{B2D} - c_{D2D} \cdot |\Psi_l|}{v_k} \right\}.
\]  

(14)

\( p_{\text{max}} \) is the utility-feasible delay probability. At the beginning of the dissemination, the operators can prejudge D2D transmission utility gain and select the transmission mode on B2D or D2D according to the following rules: If \( p(D_{\Psi l}(\tau_k)) > p_{\text{max}} \), \( \Psi_l \) is deemed as a utility-feasible D2D path, then the eNB will adopt the D2D mode through \( \Psi_l \). Otherwise, it is regarded as an utility-infeasible D2D path, and accordingly, the eNB will adopt B2D mode to transmit it to the destination directly.

Considering the positive correlation between transmission utility gain and transmission delay probability, selecting the optimized path with the maximum utility in all feasible paths can be implemented by estimating the delay probability of each path and choosing the maximum one among them.

### 3.4 Delay analysis

In the context illustrated in Figure 1, the total path transmission delay includes both the B2D transmission time from the eNB to the center user and the multihop D2D transmission time from the center user to the requester. To simplify the analysis on multihop transmission delay, the D2D communication rate between each user pair is assumed to be identical.

In our scenario, end-to-end delay mainly rests with the delay on the contact interval between each user-pairs, instead of delay on queuing up. Thus, the performance of our scheme is insensitive to distribution of message request intervals.

As a widely used mobility model used in the work of Wu et al and Gao et al, we consider the meeting process between each paired user \( (u, v) \) follows the Poisson distribution with a contact rate \( \lambda_{uv} \), thus the intercontact time \( X_{uv} \) between each paired user follows the exponential distribution, whose rate parameter is \( \lambda_{uv}^{-1} \). Moreover, considering the intermittently connectivity of the mobile network, we assume the sizes of requested contents are small-scale or middle-scale file.

**Definition 2.** Direct D2D and carry-forward D2D direct D2D: node \( v \) is in node \( u \)'s D2D communication radius of \( r \), denoted as \( ||u - v|| \leq r \).

Carry-forward D2D: node \( v \) is not in node \( u \)'s D2D communication radius of \( r \), denoted as \( ||u - v|| \geq r \), but node \( u \) will store and carry forward the content to node \( v \) in a delay constraint \( r \).

**Definition 3.** The transmission delay \( D_{uv} \) between node \( (u, v) \) is composed of D2D direct transmission delay \( (D_{uv}^D) \) and D2D carry-forward transmissions delay \( (D_{uv}^C) \), formulated as

\[
D_{uv} = D_{uv}^D + D_{uv}^C.
\]  

(15)
In particular, if in time slot \( t \), the node \( u \) happens to be adjacent to another user \( v \) and forms a direct D2D link, then \( D_{uv}^C = 0 \), \( D_{uv} = D_{uv}^B \).

Let \( \Psi_i^l \) denote the \( l \)th path of \( \Psi_i \), which is consisted of \( k \)-hop D2D links, denoted as: \( \Psi_i^l = \{1, \ldots, u, v, \ldots, k \} \).

The total transmission delay of path set \( \Psi_i^l \) (denoted as \( D(\Psi_i^l) \)) equals to the sum delay of \( \tau_{B2D} \) and sum delay of each hop in the path set \( \sum_{(u,v) \in \Psi_i^l}(D_{uv}^D + D_{uv}^C) \) (denoted as \( D_{uv} \)). Then, \( D(\Psi_i^l) \) can be evolved to

\[
D(\Psi_i^l) = \sum_{(u,v) \in \Psi_i^l} D_{uv} + \tau_{B2D}.
\]  

For each hops delay \( D_{uv} \) in path set \( \Psi_i^l \), based on Equation (15), we can get

\[
D(\Psi_i^l) = \left( \sum_{(u,v) \in \Psi_i^l} (D_{uv}^D + D_{uv}^C) \right) + \tau_{B2D}.
\]

For a \( k \)-hop D2D path set \( \Psi_i^l \), based on our aforementioned assumption on identical D2D communication rate, the D2D direct transmission delay \( D_{uv} \) between each paired user can be deemed to be identical. Subsequently, \( \Psi_i^l \)'s total direct D2D transmission delay can be uniformly represented as \( kD_{uv}^D \). We then make it as the substitution of \( \sum_{(u,v) \in \Psi_i^l} D_{uv}^D \) in Equation (17). Therefore, the transmission delay \( D(\Psi_i^l) \) of the path set \( \Psi_i^l \) can be inferred as

\[
D(\Psi_i^l) = \left( \sum_{(u,v) \in \Psi_i^l} D_{uv}^C \right) + kD_{uv}^D + \tau_{B2D}.
\]

Let \( \{\lambda_1, \ldots, \lambda_{uv}, \lambda_{vuv}, \ldots, \lambda_{(k-1)uv}\} \) denote the contact rate of adjacent nodes in \( \Psi_i^l \). Because the contact process of each paired user is independent and identically distributed (iid), the intercontact time variable \( X_{uv} \) is also iid. The transmission delay of path \( u \to v \to w \) is \( X_{uv} = X_{uv} + X_{vw} \). It can be inferred that its delay probability density functions (PDF): \( p_{X_{uv}}(z) \) can be calculated as the convolution of \( p_{X_{uv}}(x) \) and \( p_{X_{vw}}(y) \)

\[
p_{X_{uv}}(z) = p_{X_{uv}}(x) \otimes p_{X_{vw}}(y).
\]

For \( \Psi_i^l \), its overall carry-forward D2D delay \( D^C(\Psi_i^l) \) is the sum of each hop's carry-forward delay \( X_{uv} | (u, v) \in \Psi_i^l \), we can get the following equation:

\[
D^C(\Psi_i^l) = \sum_{(u,v) \in \Psi_i^l} D_{uv}^C \]

\[
= X_{12} + \cdots + X_{uv} + \cdots + X_{(k-1)uv} | (u, v) \in \Psi_i^l.
\]

Each variable \( X_{uv} | (u, v) \in \Psi_i^l \) is an iid exponential random variable with PDF: \( p_{X_{uv}}(x) = \lambda_{uv} e^{-\lambda_{uv} x} \). Then, \( D^C(\Psi_i^l) \) is a hyperexponential random variable consisting of \( k \) exponential random variables. According to Ross, \( p_{D^C(\Psi_i^l)}(x) \) can be calculated as hyperconvolution of total adjacent nodes in \( \Psi_i^l \)

\[
p_{D^C(\Psi_i^l)}(x) = p_{X_{12}}(x) \otimes \cdots \otimes p_{X_{uv}}(x) \otimes \cdots \otimes p_{X_{(k-1)uv}}(x).
\]

Accordingly, we can compute \( p_{D^C(\Psi_i^l)}(x) \) by using the following equation:

\[
p_{D^C(\Psi_i^l)}(x) = \sum_{i=1}^{k-1} C_{ik} \lambda_{k(i+1)} e^{-\lambda_{k(i+1)} x}
\]

\[
x \geq 0,
\]

where

\[
C_{ik} = \prod_{j=1, j \neq i}^{k} \frac{\lambda_{j(i+1)}}{\lambda_{j(i+1)} - \lambda_{i(i+1)}}.
\]
Definition 4. For a $k$-hop D2D path $\Psi_i^l$, its cumulative distribution function (CDF) of transmission delay $P_{\Psi_i^l}(\tau_k)$ denotes the transmission delay probability before a given time constraint $\tau_k$

$$P_{\Psi_i^l}(\tau_k) = P\left(D\left(\Psi_i^l\right) \leq \tau_k\right).$$

(24)

Using Equation (18) as the substitution of $D(\Psi_i^l)$ in Equation (24), we can get

$$P_{\Psi_i^l}(\tau_k) = P\left(\left(\sum_{(u,v) \in \Psi_i^l} DC_{uv} + kD_{av} + \tau_{B2D}\right) \leq \tau_k\right).$$

(25)

$$P_{\Psi_i^l}(\tau_k) = P\left(\sum_{(u,v) \in \Psi_i^l} DC_{uv} \leq \tau_k - kD_{av} - \tau_{B2D}\right).$$

Since $\sum_{(u,v) \in \Psi_i^l} DC_{uv}$ equals to the overall carry-forward D2D delay $DC(\Psi_i^l)$ mentioned in Equation (18), the formula in Equation (25) can be derived into the following equation:

$$P_{\Psi_i^l}(\tau_k) = P\left((DC(\Psi_i^l) + kD_{av} + \tau_{B2D}) < \tau_k\right).$$

(26)

Then, $P_{\Psi_i^l}(\tau_k)$ can be calculated by the integral over $p_{DC(\Psi_i^l)}(x)$ with the integral region from 0 to $(\tau_k - kD_{av} - \tau_{B2D})$. $P_{\Psi_i^l}(\tau_k)$ can be derived as

$$P_{\Psi_i^l}(\tau_k) = \int_{0}^{\tau_k - kD_{av} - \tau_{B2D}} p_{DC(\Psi_i^l)}(x)dx.$$  

(27)

By the combinational use of Equation (13) and Equation (27), our scheme can estimate the path’s transmission delay probability to decide whether it is a utility-feasible D2D path. This will be later discussed in the Section 4.

3.5 Online social attributes

We borrow the concept of encounter information in opportunistic network. History encounter information can predict future encounter opportunities. In the proposed scheme, we exploit three social structural metrics, namely, centrality, contact rate, intercontact time in content dissemination. In practical scenario, due to the mobility, each metric is time varying. Designed for practical applications, our scheme calculates these metrics for individuals in an online manner. That is to say, as soon as the user enters the cell scope and registers from the eNB, the scheme forms each metric of the user and detect online social contact pattern in a distributed way. Each time when two users encounter within the D2D communication distance, they detect each other via D2D discovery. Accordingly, their social metrics can be derived and updated locally. In addition, through the signal channel at each slot, UEs inform the eNB of the update on metrics in real time.

3.5.1 Centrality

Considering high-degree nodes play an influential role in maintaining the network’s connectivity, the selection and removal of a few nodes has long been a promising approach in mobile social networks. Centrality measurement is an important concept in social networks, which indicates the relative importance of nodes in a network. Various centrality measures have been proposed as an important concept in the literature, such as degree centrality, closeness centrality, etc, and utilized to select relay nodes in a delay tolerant network and opportunistic network. Numerical results show that the utilization of centrality measures has significantly improved the network performance in terms of delivery ratio, delivery delay, etc.

Motivated by the center users influence on rapid content-spreading ability in social network, in our collaborative offloading scenario, we extend the previous works by using degree centrality to choose the optimal relay in the first dissemination stage, in order to reduce the overall delay.
First, we solve the formulation of the D2D contact records. A D2D contact pair from user $i$ to another user can be modeled as a collection of three tuples. Each tuple $X_i$ has three elements, denoted as

$$X_i = \{(j, t_1, t_2) | j \in (1, \ldots, n), j \neq i, t_1 > t_0, t_2 > t_0\}.$$  

(28)

Supposed that the element $j$ represents the other user who has encounter records with user $i$. The elements $t_1$ and $t_2$ are the time stamps when the connection is up or down. $t_0$ is the initial time stamp. Once a D2D pair is established, the contact tuple is created with set to current time stamp and set to infinity. Once this pair is disconnected, in this contact triple is updated to current time stamp. For example, the tuple $(j = 1, t_1 = 3, t_2 = \infty)$ demonstrates that user $i$ connects user1 in TS3 and still sustains the connection up-to-now. At TS7, user $i$ disconnects user1, this tuple turns into $(j = 1, t_1 = 3, t_2 = 7)$, representing that user $i$ connects user1 in TS3 and disconnects user1 in TS7.

Second, we proposed a modified distributed centrality, called local degree centrality. Hui et al. have already proved that the degree measurement for previous unit time slot can approximate the centrality degree of nodes in the network. Thus, we consider the degree centrality equals to how many unique nodes a node has met in the previous slot window ($\Delta t$).

**Definition 5.** User $i$'s local degree centrality is defined as

$$C_i = \sum_{j=1, j \neq i}^{n} j | if \exists (j, t_1, t_2) : (t_1, t_2) \cap [t_0, t_0 + \Delta t].$$  

(29)

To keep up with the dynamic topology and also to balance the load on computation, a proper slot window is selected for 10 minutes, which is in accord with BUBBLE Rap. That means, since a user enters the coverage of the eNB, its centrality is recalculated with a frequency of 10 minutes. In particular, for the initial period, in which the observing window is smaller than 10 minutes, its actual observing window is computed as the slot window.

Figure 2 and Figure 3 reveal the distribution of nodes' average centrality in Infocom06. We observe very wide heterogeneity in each node centrality. This clearly shows that there are a small number of nodes, which have extremely high relaying ability, and a large number of nodes that have moderate or low centrality values, across all experiments. These observations motivate us to select the node with the maximum centrality as a good offloading carrier.

### 3.5.2 Contact rate and intercontact time

The contact rate $\lambda_{ij}$ is formulated as the average contact times between the paired user $(i, j)$ during user $i$ sojourn in the data spot. Similarly, $T_{ij}$ is formulated as the average intercontact duration between the paired user $(i, j)$ in the same observation window.
The contact rate and the intercontact time form determining factor for the transmission delay. According to previous analysis on real trace data, we draw the conclusion that the contact rate is nonuniformly distributed, particularly the contact rate is extremely high for some specific user pairs. These observations motivate us to leverage it to predict D2D transmission delay.

Both metrics are maintained by individual user locally according to up-to-date encounter information. When two nodes meet each other, they also update the records locally on how many time they met each other and how long since last time they met each other. In this way, each node maintains individual information of contact rate and intercontact duration of other nodes.

We take the process of maintenance as an example. Each node maintains its local intercontact time vector in real time, e.g., \( \langle T_{11}, \ldots, T_{ij}, \ldots, T_{in} \rangle \). When they encounter each other, they first exchange their intercontact time vector, then update the vector based on the up-to-date contact. In each slot, each node informs the eNB of its interval vector. The eNB maintains a global directed graph \( G(V, E(t_0 + t), E(t_0 + 2t), \ldots, E(t_0 + nt)) \), in which \( V \) is the set of vertices representing requesters for \( mk \), and \( E(t) \) is a time-varying set of edges referring to proximity D2D contacts between a pair of entities in \( V \). Its weight \( T_{ij} < i, j > \in E(t) \) equals to the intercontact time between the user pair \((i, j)\). In each slot, each node informs the eNB of its interval vector. In this way, the eNB maintains a global adjacent matrix \( M \) of all nodes’ intercontact time, denoted as

\[
\begin{bmatrix}
0 & T_{12} & \cdots & T_{1n} \\
T_{21} & 0 & \cdots & T_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
T_{n1} & T_{n2} & \cdots & 0
\end{bmatrix}
\]

(30)

Since the intercontact time between the user pair \((i, j)\) and \((j, i)\) is identical, we get \( T_{ij} = T_{ji} \). Thus, \( M \) is symmetric. Then, the minimum delay between two nodes can be calculated by performing Dijkstra algorithm on \( M \).

4 | SOCIAL-AWARE DELAY PREDICT CONTENT DISSEMINATION

Since D2D communication occurs between individual users, the connectivity among users can be intermittent. As a result, the communication may drop due to users’ mobility, which can impact the users’ QoS while also reducing the efficiency of traffic offload. In contrast, the mobile nodes, whose carriers are human beings, are potentially of social attributes. The social relations between individuals tend to be stable over time. Previous trace analysis has proved that social network remains stable and regular while physical network dynamics at times. Such social ties can be utilized to achieve a stable transmission link via D2D communication. In addition, social attributes have been verified to promote routing efficiency effectively in previous research. Our scheme lays the fundament on nodes’ social contact patterns, to predict transmission delay.
In this section, assisted with social relationship, we present a two-stage scheme, which is called SADP scheme, to improve the D2D transmission utility: SADM stage and SACD stage.

At the SADM stage, we arrange the dissemination task according to social contact rate, estimate D2D transmission utility gain for each path, and make a wait or stop decision on D2D communication. If a node is suitable for the stop decision on D2D communication, the scheme will immediately arrange the dissemination in B2D communication instead of hopeless waiting. Else, it will send to it through the center user via D2D communication.

At the SACD stage, we take nodes’ future social ability into consideration and exploit the node with higher contact ability to schedule data dissemination, with the aim to maximize the D2D transmission link utility gain.

4.1 | SADM stage

Assume the eNB interacts with the nodes and schedules the dissemination through each control channel. To centralize the information on the contacts between nodes, each node must periodically (e.g., every time slot) inform the eNB of its latest contact information. In this way, the eNB maintains global information of all the node.

Once the nodes request for the content $m_k$, the eNB gathers these requests and construct a content-specific requester set, then schedules the dissemination from a global perspective, restrained the dissemination within the set in order to avoid message flooding.

At the beginning of the dissemination, the eNB chooses the node with the maximum centrality to act as the best carrier. Then, based on history contact record, the eNB estimates each paths D2D transmission utility gain and forms the feasible D2D path set the by comparing transmission delay probability with feasible delay probability $p_{\text{max}}$. If the feasible path set is not empty, it will proceed with dissemination through multihop D2D links in the following SACD stage. Otherwise, the eNB will send the content to it through a cellular link directly.

Let us take Figure 4 as an example. At TS1, once received, the content requests from the nodes, the eNB responds to the carrier with the highest centrality (e.g., UE1), sends the content $m_k$ to UE1 with a “B2D” acknowledgment label (ACK=“B2D”) representing the cellular transmission mode, and informs it of the requester set for $\text{REQ}(m_k)$=(UE2, UE3, UE4, UE5, UE6). In the meanwhile, it responses to the remaining requesters with a “D2D” acknowledgment label (ACK=“D2D”) denoting the D2D transmission mode and a notification of the carrier set (Noti($m_k$)=UE1). If a new request for $m_k$ arrives at TS1 or later, the eNB will response to it with a similar ACK/Noti label in the same way.

4.2 | SACD stage

In the SACD stage, the content dissemination scheduling process abides by the following basic rules. Due to the complex network topology, a mobile node can meet multiple users in D2D proximity distance simultaneously. However, due to the occupation of D2D interface, a D2D user can transmit data to at most one receiver and receive data from one transmitter at the same time.47

Due to the multihop dynamic transmission network, the schedule of transmission task is complex and undeterministic. From a global view, the estimation of transmission delay and probability over an opportunistic path is difficult and this
cooperative offloading problem is NP-hard. In practice, to tackle this problem, we apply the heuristic algorithm in a distributed way to keep up with the time-varying topology.

Based on local information, each carrier individually decides and select the optimal node that possess the best transmission utility gain. In our scheme, the transmission utility gain refers to the ability to forward the message to more nodes in the future with the least influence on overall delay. Specifically, at each time interval, the messages are sorted decreasingly according to their size. For each given message, to meet the delay probability restriction, we sort the requesters from neighbors in decreasing order of their future-delay abilities. Then, the sender selects the one with the smallest value as its receiver. With our disseminating mechanism, the devices always select the links that will potentially contribute most to the transmission utility gain.

To extend the previous hypothesis, we assume that initially, UE1~6 request for the content simultaneously. The D2D content dissemination progress in each time slot is shown in Figure 5. The dashed lines denote the encounter record between a paired user, particularly the red dashed lines denote the trace of content dissemination. At TS0, according to SADP, the eNB makes the dissemination decision to UE1 via B2D link while to others via D2D links. From then on, UE1 undertakes the role of the content carrier, keeping on storing and carrying the content, seeking for the D2D forwarding opportunity.

Figure 6 demonstrates the D2D content dissemination process at TS1 in detail. The carrier (UE1) encounters some requesters (UE2, UE3, and UE4) simultaneously in a D2D peer discovery radius. After they exchange local encounter
information, the carrier predicts that UE4 is earlier to encounter other nodes than UE2 and UE3. Thus, in TS1, the carrier makes the dissemination choice to the selected requester (UE2). In fact, as is proved in future slot, UE3 will meet UE6 later in TS3, and UE2 will so far as to meet none, with the comparison that UE4 will meet UE5 earlier in TS2. After the decision, the carrier informs the eNB of it and apply for the arrangement of the D2D communication. Then, the eNB allocates the spectrum resource, arranges the dissemination for the D2D communication, so the carrier sends the content to the selected requester.

Furthermore, to loss of generality, we also consider one special case that when a new added requester for the same content arrives at a certain time slot, the eNB will add the node into the requester set immediately while informing it of all the carriers. Through this way, in the next slot, the new added node will become the D2D offloading target if it happens to encounter the carriers.

According to Equation (8), the optimization goal is to achieve maximum link utility gain in a certain delay budget, then it can convert to be implemented in a heuristic way: In each slot, for a given message, the scheme selects the node in $R$, which brings minimum delay to the carrier set $S$ in the future. Let us take Figure 7 as an example. If in slot $t$ we choose node $j$, this additional delay can be measured as the expectation value of the delay $D_{jk}$ from node $j$ to the remaining set $j \in R \setminus \{j\}$. By means of graph theory, $D_{jk}$ can be calculated as the short distance between $j$ and $k$ by performing Dijkstra algorithm on $G(V,E)$. According to the prediction on the delay from node $j \in R$ to the remaining set $k \in R \setminus \{j\}$, denoted as $D_{j \rightarrow R \setminus \{j\}}$, the carrier selects the node, which has the least expected delay value

$$j^* = \arg \min_{j \in R, k \in R \setminus \{j\}} (D_{j \rightarrow R \setminus \{j\}}).$$

(31)

Considering each encounter pair that node $j$ forms with the node $k$ in remaining set $R \setminus \{j\}$, $D_{j \rightarrow R \setminus \{j\}}$ can be represented as the expected value of all the delay between each encounter pair

$$D_{j \rightarrow R \setminus \{j\}} = E(D_{jk})|j \in R, k \in R \setminus \{j\}).$$

(32)

Supposed that the each encounter pair is of equal probability, Equation (31) can turn into the formula as follows, based on which the optimal D2D pair is chosen

$$j^* = \arg \min \left( \frac{1}{|R| - 1} \sum_{k \in R \setminus \{j\}} D_{jk} |j \in R, k \in R \setminus \{j\} \right).$$

(33)

Usually, the carrier will discover multi-tie D2D links simultaneously under dynamic environment. It will choose the optimal link, which provides the maximum link utility to the owner set. Since the scheme is based on such a distributed insight, it ensures that the dissemination scheme is locally optimal instead of globally optimal. In this way, it can be adaptive to dynamic network topology.
The pseudo codes of SADP with two stages is shown as Algorithm 1.

**Algorithm 1** Social-aware delay predict scheme

```
Input: G(\(D_k, E\)), \(D_k\), \(\tau_k\)
Output: Offloaded list \(F\)
1 Initialize;
2 \(S = \emptyset, F = \emptyset, R = \emptyset, \phi = \emptyset\);
3 for \(t = t_0; t \leq \tau_k, t + +\) do
4     The eNB maintains encounter information and collects requests of each node;
5     The eNB collects requests of each node;
6     Choose the center user: UE\(_{Center}\);
7     Transmit the content from eNB to UE\(_{Center}\);
8     Move UE\(_{Center}\) from \(D_k\) to \(S\);
9     Find all possible paths from UE\(_{Center}\) to the requesters;
10    WFT(\(G(D_k, E\)), UE\(_{Center}\));
11   for \(i \in D_k\) do
12       Predict \(\Psi'_i\)'s delay probability \(P(D_{\Psi'_i}(\tau_k))\);
13       Calculate \(\Psi'_i\)'s D2D transmission utility gain \(\Delta U_{\Psi'_i}\);
14       if \(\Delta U_{\Psi'_i} > 0\) then
15           if \(P(D_{\Psi'_i}(\tau_k)) > \rho_0\) then
16               add \(i\) to \(R\);
17           else
18               Send the content to \(i\) directly via cellular link;
19       end
20   end
21 if \(\phi \neq \emptyset\) then
22   for \(i \in S\) do
23     for \(j \in R\) do
24       if \((i, j) \in \phi\) and \(\| i, j \| \leq r\) then
25         for \(k \in R\) and \(k \neq j\) do
26             \(D_{j,k} = \text{Dijkstra}(G(D_k, E), j, k)\);
27             \(E(D_{j,k}) = \frac{1}{|R|-1} \sum_{k \in R \setminus \{j\}} D_{j,k}\);
28         end
29       end
30       Calculate the delay from node \(j\) to the remaining set \(D_{j \rightarrow R \setminus \{j\}}\);
31       Choose the optimal D2D pair: \((i, j^*) | j^* = \arg \min (D_{j \rightarrow R \setminus \{j\}})\);
32       Node \(i\) transmits the message to \(j\) via D2D link;
33       Add \(j\) to \(S\) and \(F\);
34     Remove \(j\) from \(R\);
35   end
36 return \(F\);
```

In the beginning, we initialize the set \(S, F, R,\) and \(\phi\) to empty. Thereinto, set \(S\) is used to store the content carriers. Set \(F\) is used to store the offloaded nodes. Set \(R\) is used to store the content requesters. Set \(\phi\) is used to store the D2D utility-feasible path set.

At the first step, we start the SADM stage (line 4–19). Then, the algorithm chooses the center user with the max centrality degree. After finding the center user, we perform the width-first traversal algorithm (WFT) to find all possible paths from the center user to the requesters (line 10). The complexity of WFT algorithm is \(O(|D_k|^2)\). For each node \(i\) in \(D_k\), according to Equation (13) and Equation (27), we predict each paths delay probability and D2D transmission utility gain, then form the D2D utility-feasible path set (line 11–19). This procedure of SADM has the complexity of \(O(|D_k| \cdot |\Psi_i|)\). Assumed \(|\Psi_i| \leq |D_k|\), the complexity of SADM stage is \(O(|D_k|^2)\).

The second step is the SACD stage (line 20–31). After performing the Dijkstra algorithm on each user pair \((j, k)\), we find the shortest path between this user pair (line 25). Then, based on the rule defined in Equation (33), the optimal D2D
link \((i, j)\) is chosen for the content carrier \(i\) (line 28). Since Dijkstra algorithm has the complexity of \(O(|D_k| + |V| \log |V|)\), the complexity of the SACD stage is \(O(|D_k|(|D_k| + |V| \log |V|))\).

Thus, the total complexity of the algorithm is \(O(|D_k|^2 + |D_k| |V| \log |V|)\). In our next step research, we will be working on optimizing the algorithms and try to obtain the optimal solution with lower complexity.

Besides, to evaluate the offloading efficiency of SADP, for each simulation conducted, we use the following four metrics:

1. **Successful Rate**: If a node’s multihop delay probability subject to a delay constraint is greater than the feasible delay probability, it is regarded as a successful node, which can be reachable through multihop D2D communications. Successful rate is then defined as the ratio of successful nodes to the total nodes

\[
\text{SuccessfulRate} = \frac{\sum_{k=1}^{n} \sum_{i \in D_k} \text{success}(i)}{\sum_{k=1}^{n} |D_k|}
\]

\[|\text{if } P(D_{\psi_k}(\tau_k)) > p_0 \Rightarrow \text{success}(i) = 1; \text{ else success}(i) = 0.\] (34)

2. **Offloaded Ratio**: The proportion of offloaded contents to the total set, is defined as

\[
\text{OffloadedRatio} = \frac{\sum_{k=1}^{n} \sum_{i \in D_k} x_i^k |m_k|}{\sum_{k=1}^{n} \sum_{i \in D_k} |m_k|}.
\]

(35)

For each offloaded content, its offloading traffic can be measured by its content size \(|m_k|\). The numerator in Equation (35) denotes the total offloading traffic through D2D communication. Then, it is averaged by the denominator, the sum traffic either through D2D or cellular link.

3. **Average Hop Count**: For all the offloaded content, the number of hops, counted from the source, that a content can travel before reaching the destination, is divided by total number of delivered content.

\[
\text{AverageHopCount} = \frac{\sum_{k=1}^{n} \sum_{i \in D_k} x_i^k \cdot \text{hop}(\psi_i^k)}{\sum_{k=1}^{n} \sum_{i \in D_k} x_i^k} |x_i^k = 1.
\]

(36)

Therefore, \(x_i^k = 1\) denotes that node \(i\) gets the message \(k\) through D2D link. Otherwise, \(x_i^k = 0\). \(\psi_i^k\) denotes the dissemination path for content \(m_k\) from eNB to user \(i\). Thereinto, \(\text{hop}(\psi_i^k)\) denotes the hop count of \(\psi_i^k\).

4. **Average Transmission Delay**: For all the offloaded content, the transmission time, which is calculated from the source to destination, is divided by total number of delivered content.

\[
\text{AverageTransmissionDelay} = \frac{\sum_{k=1}^{n} \sum_{i \in D_k} x_i^k D(\psi_i^k)}{\sum_{k=1}^{n} \sum_{i \in D_k} x_i^k} |x_i^k = 1.
\]

(37)

From previous theoretical analysis, the total transmission delay \(D(\psi_i^k)\) equals to the sum delay of cellular delay and D2D delay in each hop. In practical simulation, we calculate the time interval from the content disseminating time stamp to the content received time stamp.

## 5 PERFORMANCE EVALUATION

In this section, we evaluate the performance of our scheme by using one, a time-slotted simulator. We deploy real trace into the simulated scenario and calculate the aforementioned metrics to determine content dissemination performance.

We conduct extensive simulations to evaluate the performances of the proposed algorithms. The compared algorithms, the traces that we used, the simulation settings, and the results are presented as follows.
TABLE 1  Main simulation parameters

| Parameter            | Value          |
|----------------------|----------------|
| TTI                  | 0.1 s          |
| \( \tau \)           | 100 s~36000 s  |
| TTL                  | 300 minutes    |
| Messages size        | 500 kB~1 MB    |
| \( p_0 \)            | 0.7            |
| Message request interval | 20 s~30 s    |
| Request ratio        | 10%, ... , 100%|
| D2D transmission speed | 2 Mbps        |
| Cellular transmission speed | 100 Mbps |
| Buffer size          | 50 MB          |

Abbreviation: D2D, device-to-device.

5.1 | Simulation settings

In our simulation, the coverage of the cellular link is 300 m, the coverage of the D2D link is 20 m. We assume that mobile devices are randomly distributed in one cell served by an eNB node, which can establish cellular link to other nodes any time during the simulation. At randomly selected time stamp in each simulation run, nodes send content requests. Each mobile user can get the content via D2D or B2D link. The performance results are averaged for 50 runs. For simplicity, we assume each device has equal storage capacities and the rates of D2D links are identical.

Conforming to the message scale aforementioned in Definition 3, messages with different sizes follow uniform distribution ranging from 500 k to 1 M. The request ratio among the mobile users ranges from 10% to 100% with an increasing step of 10%. Message request interval follows uniform distribution and is randomly selected between 20 s and 30 s.

The communication distances of D2D and B2D are randomly selected, following uniform distribution ranging from 15 m to 25 m and 150 m to 250 m, respectively. The value of amplification factor for B2D link energy consume is \( \alpha = 30 \). The value of link cost coefficients is \( \delta = \frac{1}{20} \), \( \gamma = 2 \).

The D2D transmission speed is set to 2 Mbps based on Bluetooth’s transmission rate and cellular transmission speed is set to 100 Mbps based on 4G network’s transmission rate, so as to simulate the real network environment. The granularities of contact duration and intercontact duration are at minute level. Due to the timeliness of different contents, ie, some are time sensitive highly required being shared on the spot, some are time insensitive as long as they can be disseminated within the sojourning time. To make an overall consideration on the two scenes, we use diverse granularity of delay budget. When in a time-sensitive scene, the granularity of delay budget is restricted to 100 s, whereas in a time-insensitive scene, it is restricted to 3600 s (1 hour).

In our experiment, we perform the algorithm at each selected TTI and the results are averaged for 50 runs. The detailed simulation parameters are shown in Table 1.

5.2 | Real-traces used

We use the encounter data set Infocom06 gathered by the Haggle Project, which is consist of 98 mobile devices (0-97). The devices were distributed while participants were attending the Infocom 2006. Participants belong to different social communities (depending on their country, research topic, etc). The duration of the data set covers three days. Contacts were recorded among mobile devices equipped with Bluetooth while moving in workshop area. Among the existing numerous traces, we choose Infocom06 data set for the following reasons. First, the contact duration is relatively much closer to the delay constraint compared with other data sets. Second, Infocom06 works in an intensive data spot such as office environments, conference environments, which fits quite well the working scenario mentioned in Section 3. The details of the trace is summarized in Table 2. Besides, because participants in Infocom06 are in the same academic affiliation, there are stable social correlations between these mobile devices.

In the experiment, we use the initial half part ([1 s, 170 000 s]) of the trace for the mobile network to warm up. Data taken from the remaining part([170 001 s,34000 s]) is used to run data transmission.
TABLE 2  The statistics of the real trace

| Notations            | Value     |
|----------------------|-----------|
| Trace                | Infocom06 |
| Network type         | Bluetooth |
| Duration (days)      | 3         |
| Number of experimental devices | 98         |
| Contacts             | 191,336   |
| Average contacts/pair/day | 6.7       |
| Granularity (seconds)| 120       |

5.3 | Evaluation results

5.3.1 | The influence of centrality

After estimating each nodes’ transmission delay via D2D link, the eNB selects the node with maximum centrality degree as the offloading carrier among the requesters.

To demonstrate the impact of social property on data dissemination, we also simulate nonsocial-aware distributed dissemination scheme (N-SADP), which is used for comparison. In N-SADP, for each round of the simulation, we randomly selected a node to send messages to the remaining requesters.

Both offloaded ratio and average transmission delay are affected by the centrality. First of all, the centrality degree represents the ability to meeting other nodes in per time, which means the carry and forward chance to offload content. For per hour centrality degree, the center node is 53 and randomly selected nodes is 37 separately. This gives rise to the advantage for SADP in offloaded ratio. Second, in our scheme, end-to-end delay mainly rests with the contact interval between each user pairs. In terms of contact interval, from 1-hour trace analysis, the average contact interval of the center node and randomly selected nodes is 492 s and 551 s for each, which bring about a smaller transmission delay in SADP compared with N-SADP.

As is shown in Figure 8 and Figure 9, the simulation result demonstrates that, at first, when the delay constraint is within the initial 900 s, the offloaded ratio of N-SADP slightly exceeds that of SADP and the delay of N-SADP is less than that of SADP. This is because the centrality has the nature of statistical character and its meeting ability cannot be fully reflected in small observation window. When probing into the real trace, we discover that, in such small zone (900 s), the probability the center node meets other nodes is 2.04%, which is far less than the average probability (22.44%) the random selected node meets other nodes. Nevertheless, from the long-term vision, the meeting ability for center user will be potentially reflected. The later simulation result demonstrates the trend that a looser delay constraint brings out the superiority in performance for our social-aware method. As is illustrated from 1200 s and go on, with delay constraint...

![Figure 8](image-url)  Influence of centrality on offloaded ratio. N-SADP; nonsocial-aware distributed dissemination scheme; SADP, social-aware delay prediction scheme
steadily increasing, the overwhelming superiority of SADP is embodied in a stable value of 8% (in offloaded ratio) and 100 s (in transmission delay) to that of N-SADP.

5.3.2 The influence of delay constraint and delay-outage probability

To simplify the calculation, in our experiment, we typically assume one to three hops communications, respectively, to evaluate the D2D link utility gain of multihop D2D communications.

Figure 10 shows the influence of delay-outage probability on D2D transmission utility gain. To be normalized, each utility gain is divided by each messages benefit. It can be observed that, as the delay-outage probability increases, the transmission utility gain speeds up with a linear tendency correspondingly. Specially, when \( p_0 \) increases from 0.6 to 0.7, it triggers a sharp incline in the transmission utility gain. We also vary the content benefit from 500 to 1000 to evaluate the impact of benefit on transmission utility. As we can see, for diverse benefit value, the utility gain of one-hop D2D
transmission always outperforms that of two-hop D2D transmission; so does two-hop to thre-hop. We can infer that when path hop increases, the D2D transmission utility gain will decrease due to the generation of more D2D link cost.

Figure 11 demonstrates the actual delay-outage probability compared with the desired delay-outage probability under different delay constraint. A tight delay constraint leads to the sharp differences in them, for example, when $\tau = 1$ or $\tau = 2$. However, for a middle delay constraint, for example, when $\tau = 5$ or $\tau = 6$, the actual delay-outage probability fits the desired delay-outage probability quite well. Furthermore, a much looser delay constraint may have little impact on the actual delay-outage probability. When $\tau$ is equal to 9 and above, the actual delay-outage probability will converge to an upper bound whatever the desired delay-outage probability is. From these observation, the gap between the delay-outage probability disparity is affected by delay constraint. The estimation of the delay-outage probability can work effectively under the selection of a proper delay threshold.

5.3.3 | Algorithms in comparison

To evaluate the performance of SADP based on real trace, we use other two popular schemes as benchmarks: SA-Epidemic, where the carrier offloads the data to any encountered node, and SA-Prophet, where the carrier offloads the data to the node with higher delivery probability.

As is enclosed in Figure 12 and Figure 14, the obvious fact is that, even though the number of hops obtained by the proposed method is larger than those obtained by SA-Prophet and SA-Epidemic, our method is superior to them in terms of the offloading ratio. It is because that given the same delay constraint, in SADP, the content can be disseminated through more hops.
In addition, as is shown in Figure 13 and Figure 14, we observed some fluctuations in the performance of SADP in some small zone, for example, the tiny decline in average transmission delay from 698 s to 677 s when the delay constraint varies from 2200 s to 2400 s in Figure 13, and the tiny decline in average hop count from 6.67 to 6.37 when the delay constraint varies from 1200 s to 1400 s in Figure 14. They are caused by contingencies due to that the selection of the relay node is based on statistical method. However, in general, there are asymptotic uptrends both in average transmission delay and average hop count when the delay constraint rises.

According to the general tendency, our method can effectively reduce end-to-end delay considerably compared with the other two methods at the cost of more hop counts. The SADP can fully utilize the potential of D2D networks by the way of providing multihop D2D communications.

6 | CONCLUSION AND FUTURE WORK

As a result, social-aware delay predicting is effective for improving the offloading ratio of content dissemination. In the initial stage, judging on social contact regularity, it can distinguish different communication types (ie, B2D and D2D) suitable for each request node within the delay constraint in a central way. For the second stage, it operates autonomously, ie, each carrier predicts the transmission delay of encounter user among the requesters and selects the user with minimum
delay to arrange D2D communications in a distributed way. Through numeric results compared with traditional method, we find that our proposed scheme can not only improve the performance of offloading but also D2D transmission utility gain in a cellular network. In particular, our scheme is not technology specific and also applies to the mobile network case, in which users use D2D over Wi-Fi bands via Wi-Fi Direct.

In the next step research, we mainly focus on following aspects:

1. Since the influence of the centrality was analyzed in the proposed method, nevertheless, we did not take community into consideration. If there are communities, the requester can be divided into several clusters. Content can thus be disseminated through multicenter user to improve the offloading ratio and alleviate single center user's overload. The impact of community in social network will also be further essentially required to be discussed.

2. In our experiment, we use the real trace to evaluate the performance of SADP, in which the intercontact time approximately follows exponent distribution. However, for synthetic traces, further analyses and simulations carried out to prove the performance are also expected.

3. The application scenario of SADP is highly expectant. According to SADP, the telecom operators can apply a specialized center user in RAN to act as a data sink. We will further explore the application of SADP in a 5G network to accelerate the content dissemination process and reserve cellular bandwidth via D2D link.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China under Grant 61271254 and by the Research Project of Zijin College, Nanjing University of Science and Technology under Grant 2018ZRKX0401004.

ORCID

Wenjuan Shao https://orcid.org/0000-0003-3789-4717
Liaoruo Huang https://orcid.org/0000-0002-0393-7558

REFERENCES

1. Mehmeti F, Spyropoulos T. Is it worth to be patient? Analysis and optimization of delayed mobile data offloading. Paper presented at: 33rd Annual IEEE International Conference on Computer Communications (INFOCOM); 2014; Toronto, Canada.

2. Mao G, Zhang Z, Anderson B. Cooperative content dissemination and offloading in heterogeneous mobile networks. IEEE Trans Veh Technol. 2016;65:6573-6587.

3. Han B, Hui P, Kumar VSA, Marathe MV, Shao J, Srinivasan A. Mobile data offloading through opportunistic communications and social participation. IEEE Trans Comput. 2012;11(5):821-834.

4. Mehmeti F, Spyropoulos T. Performance analysis of "on-the-spot" mobile data offloading. Paper presented at: IEEE Global Communications Conference (GLOBECOM); 2014; Austin, TX.

5. Guo X, Liu L, Chang Z, Ristaniemi T. Data offloading and task allocation for cloudlet-assisted ad hoc mobile clouds. Wirel Netw. 2018;24(1):79-88.

6. Pu L, Chen X, Xu J, Fu X. D2D fogging: an energy-efficient and incentive-aware task offloading framework via network-assisted D2D collaboration. IEEE J Sel Areas Commun. 2016;34(12):3887-3901.

7. Liu Y, Bashar AMAE, Li F, Wang Y, Liu K. Multi-copy data dissemination with probabilistic delay constraint in mobile opportunistic device-to-device networks. Paper presented at: 17th International Symposium on a World of Wireless, Mobile and Multimedia Networks; 2016; Coimbra, Portugal.

8. Xu C, Gao C, Zhou Z, Chang Z, Jia Y. Social network-based content delivery in device-to-device underlay cellular networks using matching theory. IEEE Access. 2017;5:924-937.

9. Laha A, Cao X, Shen W, Tian X, Cheng Y. An energy efficient routing protocol for device-to-device based multihop smartphone networks. Paper presented at: IEEE International Conference on Communications (ICC); 2015; London, UK.

10. Gozalvez J, Coll-Perales B. Experimental evaluation of multihop cellular networks using mobile relays [accepted from open call]. IEEE Commun Mag. 2016;51(7):122-129.

11. Fodor G, Dahlman E, Mildh G, et al. Design aspects of network assisted device-to-device communications. IEEE Commun Mag. 2012;50(3):170-177.

12. Li Y, Wu T, Hui P, Jin D, Chen S. Social-aware D2D communications: qualitative insights and quantitative analysis. IEEE Commun Mag. 2014;52(6):150-158.

13. Iqbal J, Giaccone P. Interest-based cooperative caching in multie wireless networks. Paper presented at: IEEE Global Communications Conference (GLOBECOM); 2013; Atlanta, GA.
14. Coll-Peralles B, Gozalvez J, Lazaro O, Sepulcre M. Opportunistic multi-hop cellular networking for energy-efficient provision of mobile delay tolerant services. *IEEE Veh Technol Mag*. 2015;10(2):93-101.

15. Nunes JD, Celest C, De Melo POSV, Loureiro AAF. GROUPS-NET: group meetings aware routing in multi-hop D2D networks. *Comput Netw*. 2017;127:94-108.

16. Coll-Peralles B, Gozalvez J, Friderikos V. Context-based opportunistic forwarding in multi-hop cellular networks using mobile relays. Paper presented at: 2nd High performance Mobile Opportunistic Systems (HP-MOSys); 2013; Barcelona, Spain.

17. Park J. Fast and energy efficient multihop D2D routing scheme. *Int J Distrib Sens Netw*. 2016;12(5):1-9.

18. He SM, Zhang DF, Xie K, Zhang J, Qiao H. Opportunistic routing for multiple multicast flows in wireless mesh networks. *Int J Wirel Mob Comput*. 2014;7(1):84-93.

19. Niu X, Cui L. Throughput capacity of opportunistic routing in wireless sensor networks. *Int J Distrib Sens Netw*. 2010;6(1):125-138.

20. Liu L, Chang Z, Guo X. Socially-aware dynamic computation offloading scheme for fog computing system with energy harvesting devices. *IEEE Internet Things J*. 2018;5:1869-1879.

21. Bao X, Lin Y, Lee U, Rimac I, Choudhury RR. DataSpotting: exploiting naturally clustered mobile devices to offload cellular traffic. Paper presented at: 32nd Annual IEEE International Conference on Computer Communications (INFOCOM); 2013; Turin, Italy.

22. Xing B, Seada K, Venkatasubramanian N. An experimental study on wi-fi ad-hoc mode for mobile device-to-device video delivery. Paper presented at: IEEE Conference on Computer Communications Workshops (INFOCOM); 2009; Rio De Janeiro, Brazil.

23. Zhang Y, Pan E, Song L, Saad W, Dawy Z, Han Z. Social network aware device-to-device communication in wireless networks. *IEEE Trans Wirel Commun*. 2015;14(1):177-190.

24. Cha M, Kwak H, Rodriguez P, Ahn YY, Moon S. I tube, you tube, everybody tubes: analyzing the world’s largest user generated content video system. Paper presented at: 7th ACM SIGCOMM Conference on Internet Measurement (IMC); 2007; San Diego, CA.

25. Song L, Niyato D, Han Z, Hossain I. *Wireless Device-to-Device Communications and Networks*. Cambridge, UK: Cambridge University Press; 2015.

26. Yu CH, Tirkkonen O, Doppler K, Ribeiro C. Power optimization of device-to-device communication underlaying cellular communication. Paper presented at: IEEE International Conference on Communications (ICC); 2009; Dresden, Germany.

27. Janis P, Koivunen V, Ribeiro CB, Doppler K, Hugl K. Interference-avoiding MIMO schemes for device-to-device radio underlaying cellular networks. Paper presented at: 21st International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC); 2010; Istanbul, Turkey.

28. Doppler K, Xiao M. Innovative concepts in peer-to-peer and network coding. Winner+/celtic Deliverable Celtic/cp5-026 D. https://pdfs.semanticscholar.org/ae10/2ddfdce8d398a747a2b5034773258e1827e3.pdf

29. Helmy A, Musavian L, Le-Ngoc T. Energy-efficient power allocation for multicarrier systems with delay-outage probability constraints. Paper presented at: IEEE International Conference on Communications (ICC); 2013; Oulu, Finland.

30. Lu M, Wu J. Social welfare based routing in ad hoc networks. Paper presented at: International Conference on Parallel Processing (ICPP); 2006; Columbus, OH.

31. Wu J, Lu M, Li F. Utility-based opportunistic routing in multi-hop wireless networks. Paper presented at: 8th International Conference on Distributed Computing Systems (ICDCS); 2008; Beijing, China.

32. Liu B, Cao Y, Wang W, Jiang T. Energy budget aware device-to-device cooperation for mobile videos. Paper presented at: IEEE Global Communications Conference (GLOBECOM); 2016; Washington, DC.

33. Wu J, Xiao M, Huang L. Homing spread: community home-based multi-copy routing in mobile social networks. Paper presented at: 32nd Annual IEEE International Conference on Computer Communications (INFOCOM); 2013; Turin, Italy.

34. Gao W, Li Q, Zhao B, Cao G. Multicasting in delay tolerant networks: a social network perspective. Paper presented at: 10th ACM International Symposium on Mobile ad hoc Networking and Computing (MobiHoc); 2009; New Orleans, LA.

35. Ross SM. *Introduction to Probability Models*. 11th ed. Oxford, UK: Academic Press; 2014.

36. Albert R, Jeong H, Barabási ÁL. Error and attack tolerance of complex networks. *Nature*. 2000;406:378-382.

37. Wasserman S, Faust K. *Social Network Analysis: Methods and Application*. Cambridge, UK: Cambridge University Press; 1994.

38. Magaña N, Francisco A, Pereira P, Correia M. Betweenness centrality in delay tolerant networks: a survey. *Ad Hoc Netw*. 2015;33:284-305.

39. Huang Y, Dong Y, Zhang S, Wu G. A centrality estimation method based on hidden Markov model in social delay tolerant networks. Paper presented at: Wireless and Optical Communication Conference (WOCC); 2013; Chongqing, China.

40. Jiang L, Li F, Tian C, Zhu L, Wang Y. Closeness-based routing with temporal constraint for mobile social delay tolerant networks. Paper presented at: IEEE Global Communications Conference (GLOBECOM); 2015; San Diego, CA.

41. Patel J, Shah H. FCER routing using temporal closeness centrality for delay tolerant network. Paper presented at: Nirma University International Conference on Engineering(NUICONE); 2014; Ahmedabad, India.

42. Zhang H, Liu K. *A Routing Mechanism Based On Social Networks and Betweenness Centrality in Delay-Tolerant Networks*. Vol 7; 2016.

43. Hui P, Crowcroft J, Yoneki E. Bubble rap: social-based forwarding in delay-tolerant networks. *IEEE Trans Mob Comput*. 2011;10(11):1576-1589.

44. Xia F, Liu L, Li J, Ma J, Vasilakos AV. Socially aware networking: a survey. *IEEE Syst J*. 2015;9(3):904-921.

45. Li F, Wu J. LocalCom: a community-based epidemic forwarding scheme in disruption-tolerant networks. Paper presented at: 6th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON); 2009; Rome, Italy.
How to cite this article: Shao W, Shen Q, Huang L. Social-aware content dissemination through opportunistic D2D communications. *Trans Emerging Tel Tech*. 2019;30:e3542. [https://doi.org/10.1002/ett.3542](https://doi.org/10.1002/ett.3542)