How to Download More Data from Neighbors? A Metric for D2D Data Offloading Opportunity

Zehua Wang, *Student Member, IEEE,* Hamed Shah-Mansouri, *Member, IEEE,* Vincent W.S. Wong, *Fellow, IEEE*

Abstract—Mobile devices in close proximity can be connected in a device-to-device (D2D) manner to transfer digital objects (e.g., videos) to each other. By using D2D data offloading, mobile users can reduce the cost for data service from wireless cellular networks. However, due to users' mobility, the opportunity for a user to obtain his interested objects via D2D communication is transient. In this paper, we first propose an expected available duration (EAD) metric to evaluate the opportunity that an object can be downloaded by a user via D2D data offloading. The EAD metric takes into account the pairwise connectivity of users, social influence between users, diffusion of digital objects, and the time that users would like to wait for D2D data offloading. We then propose a distributed algorithm for a mobile device to determine the EAD of each object. Given a set of available objects in the neighborhood, a mobile device will first download the object that has the smallest EAD. We validate our model via trace-driven simulations. Results show that our proposed algorithm can effectively find the object that should be first downloaded. Comparing with existing schemes, our work can help users download more data via D2D data offloading.

Index Terms—Device-to-device (D2D) communications, data offloading, expected available duration (EAD), interest estimation, social influence.

1 Introduction

Mobile devices such as smartphones and tablets are increasingly common. They enable mobile users to download digital objects (*e.g.*, videos, photos) shared by their friends in online social networks (OSNs) (*e.g.*, Facebook, Twitter). Both the number of mobile devices and the average bulk of data that users consume are increasing rapidly. By 2020, the average mobile data traffic consumed by a smartphone and a tablet will be 4.4 and 7.1 GB per month, respectively [1].

To relieve the burden of wireless cellular networks, mobile data traffic can be delivered through other means to the users (*e.g.*, WiFi, device-to-device (D2D) communications). This is known as mobile data offloading. Several works have identified the benefits of WiFi data offloading [2]–[5]. The work in [2] showed that deferring the uploading tasks until WiFi access points are available can save the energy of smartphones. Lee *et al.* in [3] conducted experiments for WiFi data offloading. By jointly considering the power consumption and link capacity of wireless network interfaces, Ding *et al.* in [4] studied the criterion of downloading data from WiFi as well as the WiFi access point selection problem. With a budget of energy consumption and monetary cost, the download duration is minimized in [5] by allocating the data traffic demand to wireless cellular and WiFi networks.

However, mobile data traffic cannot always be offloaded to WiFi networks since the number of open-accessible WiFi access points is limited [4]. To fully exploit the benefits of data offloading, mobile data traffic can also be delivered via D2D networks. Specifically, mobile devices in close proximity can be connected via WiFi Direct [6] or Bluetooth in a D2D manner to disseminate digital objects between users. This is referred to as D2D data offloading. Mobile users (*e.g.*, classmates, colleagues) may be interested in the same digital objects [7]. For example, classmates who are friends in OSNs may be interested in the same set of photos or videos shared by their mutual friends. Han *et al.* in [8] proposed a scheme to epidemically disseminate the same digital objects to mobile users by properly choosing the initial bearers of digital objects. Wang *et al.* in [9] referred to the initial bearers as seeds and considered the connectivity between mobile users in the seeds selection problem. Lin *et al.* in [10] proposed a forwarding strategy by considering different interest between users. The connectivity between mobile users is an important issue in D2D data offloading. The pairwise contact and intercontact durations, also known as the pairwise contact and intercontact time, are commonly used to model the connectivity between mobile users [11]–[14]. The former is the connected duration between a given pair of mobile users. The latter is the duration between two successive connection periods between users of a given pair. The distributions of pairwise contact and intercontact durations may be different for different user pairs. The *aggregate distributions* of contact and intercontact durations are defined as the distributions of pairwise contact and intercontact durations, respectively, when all user pairs in the network are taken into account [11], [12]. It has been shown in [11] and [12] that the aggregate complementary cumulative distribution function (CCDF) of intercontact durations decays with power law in a long time range. The

- *Zehua Wang, Hamed Shah-Mansouri, and Vincent W.S. Wong are with the Department of Electrical and Computer Engineering, The University of British Columbia, 2332 Main Mall, Vancouver, V6T 1Z4, Canada. E-mail: {zwang; hshahmansour; vincentw}@ece.ubc.ca*

Manuscript received 1 May 2016; revised 9 Aug. 2016; accepted 15 Aug. 2016. Date of publication 26 Aug. 2016; date of current version xx xxx. 2016. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below. Digital Object Identifier no. 10.1109/TMC.2016.xxxxxx

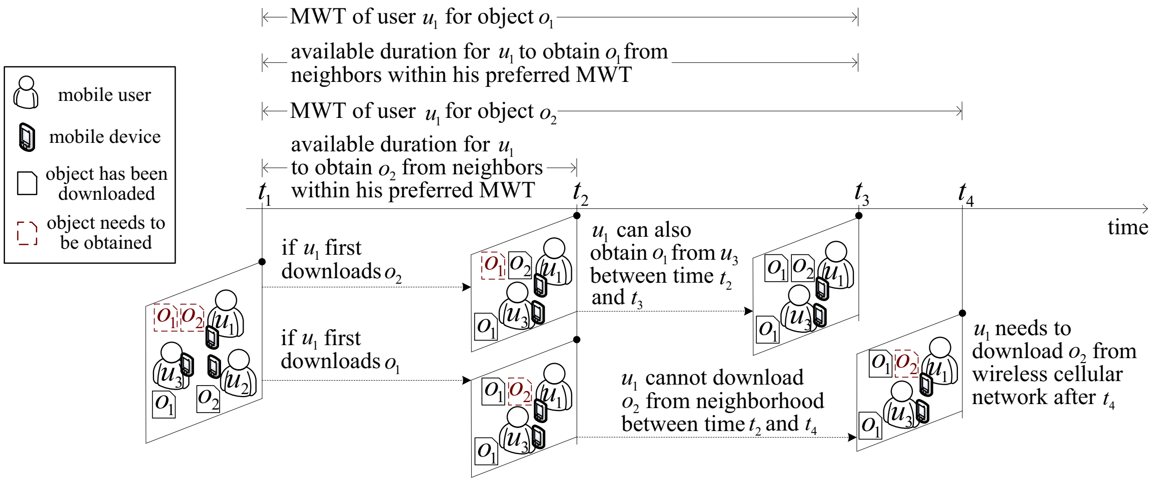


Fig. 1. An example of D2D data offloading.

work in [13] further showed that the aggregate CCDF of the intercontact durations features the dichotomy with a characteristic time. In particular, the aggregate CCDF first decays with power law before the characteristic time. It then decays exponentially after the characteristic time. Conan *et al.* in [15] showed that it is possible that the aggregate CCDF of intercontact durations decays by the power law while the intercontact duration of each individual user pair follows the exponential distribution. Furthermore, Cai *et al.* in [16] showed that when mobile users move in a finite area and pause with finite time, the intercontact duration of a given pair of users decays at least exponentially fast. Gao *et al.* in [17] further conducted the chi-square test [18] for the hypothesis that the intercontact duration of each individual user pair is exponentially distributed. They showed with several empirical data sets that over 85% of mobile user pairs in the data sets passed the test.

D2D data offloading can be enabled by running an application or background services on a mobile device. The following steps are required to perform D2D data offloading: 1) discovering mobile devices in the neighborhood, 2) determining whether the digital objects that the user is waiting to download are available in the neighborhood, and 3) choosing an available object to download and perform the D2D data transfer. These steps are necessary for D2D data offloading because: (a) D2D connections between mobile devices are stochastic, (b) a digital object may not be available on neighbors (c) a mobile device needs to choose an available digital object to send a data request and then receive the data from the neighbor.

Within the aforementioned steps to perform D2D data offloading, discovering the neighboring mobile devices can be accomplished by sending and receiving periodic hello messages [19]. The availability of digital objects on neighboring mobile devices can be determined by exchanging the uniform resource locators (URLs). However, when multiple digital objects are available on the neighbors of a mobile user, the user needs to decide which object should first be downloaded from his neighbors so that more mobile data traffic can be offloaded with D2D communications. This is because the time preferred by the user to wait for the opportunities of downloading an object from his neighbors is limited. When the waiting time exceeds the maximum waiting time (MWT) preferred by the user, the remaining data of the object, which has not been obtained via D2D data offloading, will be downloaded from the wireless cellular network [20]. Thus, choosing an appropriate object from the available objects on neighbors to first download is an important issue. Consider an example in Fig. 1. There are three mobile users u_1, u_2, and u_3. At time t_1, user u_1 wants to download digital objects o_1 and o_2 from its neighbors in order to reduce the cost of wireless cellular data service. User u_2, who has object o_2, is in the neighborhood of user u_1 from time t_1 to t_2. User u_3, who has object o_1, is a neighbor of user u_1 from time t_1 to t_4. Since the MWTs for objects o_1 and o_2 preferred by user u_1 end at time t_3 and t_4, respectively, the durations available for user u_1 to download objects o_1 and o_2 from neighboring devices within his preferred MWTs are t_3–t_1 and t_2–t_1, respectively. We assume that user u_1 can finish downloading object o_1 from user u_3 in the duration either from t_1 to t_2 or from t_2 to t_3. In this case, user u_1 should download object o_2 from user u_2 between time t_1 and t_2, as he can also download object o_1 from user u_3 between time t_2 and t_3 and obtain both objects o_1 and o_2 by D2D data offloading. Otherwise, if user u_1 downloads object o_1 from user u_3 between time t_1 and t_2, user u_1 has to download object o_2 from the cellular network after time t_4.

However, the D2D topology of mobile users and their interest are not known *a priori*. Thus, selecting one of the available objects to perform D2D data transfer is a difficult problem. The rarest first strategy [21], which is initially proposed for peer-to-peer (P2P) applications in the Internet to distribute files, may help mobile devices to make decisions. In the rarest first strategy, a computer first sends queries to determine the portion of a file that has been downloaded by the least number of computers in the Internet. That computer then downloads this portion of the file first. However, a mobile device using D2D data offloading cannot directly apply the existing rarest first strategy by first downloading the object that has been obtained by the least number of mobile devices. This is due to the following reasons. Without the network backbone, it is not efficient for a mobile device to transmit queries and replies via multiple hops in wireless domain. Even if the number of devices that have downloaded each object is known, the rarest first strategy may not work well for D2D data offloading because the mobile device can only download data from its neighbors which are changing over time. Indeed, the idea of the rarest first strategy can be extended to D2D data offloading in wireless domain. We let a mobile user first download the object (from his neighbors) which has the shortest available duration for the user to download before the end of the MWT preferred by the user for it. For user u_1 in the example in Fig. 1, the available durations of objects o_1 and o_2 in the neighborhood of u_1 within the preferred MWTs are t_3-t and $t_2-t, \forall t \in [t_1, t_2]$, respectively. Since $t_2-t < t_3-t, \forall t \in [t_1, t_2]$, user u_1 should download object o_2 from user u_2 during time t_1 to t_2.

However, evaluating the accurate available duration that a user can obtain an object from his neighbors within his preferred MWT is challenging in practice. First, the D2D connections between mobile devices are stochastic. Second, users usually have different interest on digital objects, which are not revealed until they aim to obtain the data of these objects from their neighbors or the cellular network. Moreover, information that a user is interested in an object is only known by his neighbors after they exchange the URLs of their interested objects. Furthermore, mobile users may be interested in an object at different time when it is diffused in the OSN and various MWTs may be preferred by these users for D2D data offloading.

In this paper, we propose the *expected available duration* (EAD) metric to evaluate the opportunity that an object can be obtained by a user via D2D data offloading. Specifically, for each digital object that a user is interested in and waiting to download from his neighbors, the EAD is defined as the expected length of time (evaluated by the user) that the object is available in the user's neighborhood before the end of his preferred MWT for the object. Our EAD metric takes into account the stochastic D2D connections between users, social influence to the users, and the diffusion process of digital objects in OSNs. We propose a distributed algorithm for a user to determine the EAD for each object that the user is interested in. When multiple objects are available in the neighborhood, the user will first download the object that has the smallest EAD. We assume that mobile devices have enough energy to participate in D2D data offloading. This assumption has also been made in [22]–[24]. Our major contributions are summarized as follows:

- We use a continuous-time Markov chain (CTMC) [25] to model the pairwise connectivity between each pair of mobile users. We show that by using CTMC model, we can obtain comparable results as using the power law to model the pairwise connectivity when fitting the aggregate CCDF curves of contact and intercontact durations given by the empirical data.
- We propose an interest estimation model, which takes social influence and Bayesian inference into account to estimate whether the digital object that a user is interested in also attracts the interest of other users.
- By considering the digital object diffusion and the stochastic connectivity between each user pair, we determine the availability that an object can be downloaded by a user from other users at a future time. We then propose an EAD metric for a user to evaluate the opportunity that he can download the object by D2D data offloading. We also propose to estimate the value of EAD in order to reduce the computational complexity.
- We validate our model by extensive trace-driven simulations. To the best of our knowledge, this paper is the first to propose a metric to evaluate the opportunity of D2D data offloading. Comparing with the existing scheduling schemes in the literature, using our proposed metric to schedule D2D data offloading can help mobile users download more data from neighbors.

The rest of this paper is as follows. In Section 2, we present our D2D data offloading model and an algorithm to obtain the parameters required to determine the EAD. In Section 3, we propose an algorithm to select the digital object that should first be downloaded from the neighborhood. Simulation results are presented in Section 4. Conclusion and future work are given in Section 5. The key notations and variables used in this paper are listed in Table 1.

2 D2D Data Offloading Model

In this section, we first introduce the pairwise connectivity model and the interest estimation model. Then, for each mobile user, we determine the pairwise data offloading availability of each object that the user is interested in. The EAD of each object for a user is obtained based on the MWT and the pairwise data offloading availability of the object.

2.1 Pairwise Connectivity Model

The pairwise connectivity model is used to model the stochastic D2D connection between a pair of mobile devices. Let $\mathcal{U} = \{1, \ldots, U\}$ denote the set of mobile users in the OSN. When two users are in close proximity, their mobile devices will be connected via Bluetooth or WiFi Direct in a D2D manner. We use the terms mobile users and mobile devices interchangeably. As mentioned earlier in Section 1, the pairwise contact and intercontact durations are commonly used to model the connectivity between mobile users. The aggregate CCDF of pairwise intercontact durations of all user pairs decays with power law [11], [12]. Meanwhile, it has been shown that the aggregate CCDF of intercontact durations may have the power law decay when the pairwise intercontact duration of each individual user pair follows the exponential distribution [15]. Furthermore, by conducting the chi-square test [18] with empirical data sets, the work in [17] showed that most of the mobile user pairs in these data sets satisfy the hypothesis that the intercontact duration of each individual user pair is exponentially distributed. Thus, in our work, we assume that the pairwise intercontact duration is exponentially distributed for each pair of mobile users. This assumption has also been made in several related works [17], [22], [24], and [26]. On the other hand, to obtain a tractable model, we assume that the

TABLE 1
List of key notations and variables used in this paper.

$\mathcal{A}^t_{i,k}$ (or $\mathcal{B}^t_{i,k}$)	User set maintained by user i for other users who are (or are not) interested in object k at time t
$a^{t,c}_{i,j}$ (or $a^{t,d}_{i,j}$)	Sum of the pairwise contact (or intercontact) durations between users i and j at time t
$C^t_{i,j}$	Binary random variable indicating if users i and j are connected ($C^t_{i,j}=1$) or not ($C^t_{i,j}=0$) at time t
$\boldsymbol{c}^t_i$	Connectivity profile of user i at time t, which is defined as $\boldsymbol{c}^t_i \triangleq (C^t_{i,1},\ldots,C^t_{i,i-1},C^t_{i,i+1},\ldots,C^t_{i,U})$
$I_{i,k}$	Binary random variable indicating if user i is interested in object k ($I_{i,k}=1$) or not ($I_{i,k}=0$)
$\mathcal{K}^t_{i,j}$	Set of URLs that have been sent from user i to user j at time t
l_k	URL of object k
$n^{t,c}_{i,j}$ (or $n^{t,d}_{i,j}$)	Number of times that users i and j have been connected (or disconnected) by time t
$\mathcal{O}$	Set of digital objects represented by index values $\{1,\ldots,O\}$
$\mathcal{O}^t_i$	Set of digital objects that user i is interested in at time t
$\mathcal{O}_i$	Set of digital objects defined as $\mathcal{O}_i \triangleq \lim_{t\to\infty}\mathcal{O}^t_i$
$p^t_{i,j,k}$	Estimate of user i at time t for the interest of user j in object k without Bayesian inference
$q^t_{i,j,k}$	Estimate of user i at time t for the interest of user j in object k with Bayesian inference
$\mathcal{Q}^t_i$	Set of digital objects that user i aims to download via D2D data offloading at time t
$R^t_{i,j,k}$	User i has been informed that user j is interested in object k at time t ($R^t_{i,j,k}=1$) or not ($R^t_{i,j,k}=0$)
$\boldsymbol{r}^t_{i,k}$	Interest record maintained by user i for object k at time t, defined as $\boldsymbol{r}^t_{i,k} \triangleq (R^t_{i,1,k},\ldots,R^t_{i,U,k})$
t_s	Time when time slot s begins
$\mathcal{U}$	Set of users represented by index values $\{1,\ldots,U\}$
$V_{i,k}(t)$	EAD evaluated by user i for object k at time t
$W_{i,k}(t)$	Approximated EAD of user i for object k at time t
x_k	Time when object k is initially posted in OSN
$y_{i,k}$	Realization of random variable $Y_{i,k}$ that denotes the time when user i is interested in object k
$Z^t_{i,j,k}(t')$	Binary random variable indicating if user i can download object $k\in\mathcal{Q}^t_i$ from user j at time $t'>t$
$z^t_{i,j,k}(t')$	Pairwise data offloading availability, which is defined as $\mathbb{P}(Z^t_{i,j,k}(t')=1 \mid \boldsymbol{r}^t_{i,k},\tau^t_{i,j},C^t_{i,j})$
$\delta_{i,k}$	MWT preferred by user i to download object k via D2D data offloading
$\tilde{\theta}_{i,j}$	Estimate of Jaccard coefficient $\theta_{i,j}\in[0,1]$ between users i and j
$\lambda_{i,j}$ (or $\mu_{i,j}$)	Exponential distribution parameter of the pairwise intercontact (or contact) duration between users i and j
$\widetilde{\lambda}_{i,j}$ (or $\widetilde{\mu}_{i,j}$)	Maximum likelihood estimation of $\lambda_{i,j}$ (or $\mu_{i,j}$)
$\rho_{i,k}$	Ratio between the time that object k is in the neighborhood of user i and the duration from current time to time $y_{i,k}+\delta_{i,k}$
$\tau^t_{i,j}$	Time of the most recent connection between users i and j by time t
ξ	Percentage of an object that has to be obtained before it can be shared with other users
$\xi^t_{i,k}$	Percentage of object k that has been downloaded by user i at time t

pairwise contact duration between a user pair follows an exponential distribution as well. This assumption has also been used in related works [27] and [28]. We will discuss in Section 2.3.1 that the insights of our work are also useful when other pairwise connectivity models are adopted.

To model the pairwise connectivity, let binary random variable $C^t_{i,j}=1$ (or $C^t_{i,j}=0$) denote the event that users $i,j\in\mathcal{U}$ are connected (or disconnected) at time $t\geq 0$. In addition, let $\lambda_{i,j}$ (or $\mu_{i,j}$) denote the parameter of the exponential distribution for the pairwise intercontact (or contact) duration between users i and j. Then, the connectivity between users i and j follows a CTMC model [25, pp. 358]. Given the connection state $C^t_{i,j}$ between users i and j at current time t, the probability that they are connected at future time $t'\geq t$ is given by

$$\mathbb{P}(C^{t'}_{i,j}=1 \mid C^t_{i,j}) = \begin{cases} \frac{\lambda_{i,j}-\lambda_{i,j}e^{-(\lambda_{i,j}+\mu_{i,j})(t'-t)}}{\lambda_{i,j}+\mu_{i,j}}, & \text{if } C^t_{i,j}=0, \\ \frac{\lambda_{i,j}+\mu_{i,j}e^{-(\lambda_{i,j}+\mu_{i,j})(t'-t)}}{\lambda_{i,j}+\mu_{i,j}}, & \text{if } C^t_{i,j}=1. \end{cases} \quad (1)$$

Parameters $\lambda_{i,j}$ and $\mu_{i,j}$ in (1) can be obtained by maximum likelihood estimation (MLE). A mobile device can obtain the connection states with nearby devices by receiving the acknowledgements after sending hello messages periodically. Without loss of generality, we assume that users i and j have been connected $n^{t,c}_{i,j}$ times and disconnected $n^{t,d}_{i,j}$ times by time t. Let $a^{t,c}_{i,j}$ (or $a^{t,d}_{i,j}$) denote the sum of the pairwise contact (or intercontact) durations at time t. The parameters $\lambda_{i,j}$ and $\mu_{i,j}$ estimated by MLE are given by $\widetilde{\lambda}_{i,j}=n^{t,d}_{i,j}/a^{t,d}_{i,j}$ and $\widetilde{\mu}_{i,j}=n^{t,c}_{i,j}/a^{t,c}_{i,j}$. When the connection state between users i and j has changed many times in a sufficiently long time, the values of $\widetilde{\lambda}_{i,j}$ and $\widetilde{\mu}_{i,j}$ will converge to $\lambda_{i,j}$ and $\mu_{i,j}$, respectively. Given the current connection state between users i and j at time t, user i can determine the probability that it is connected with user j at time $t'\geq t$ by substituting $\widetilde{\lambda}_{i,j}$ and $\widetilde{\mu}_{i,j}$ for $\lambda_{i,j}$ and $\mu_{i,j}$ in (1), respectively. By comparing the aggregate CCDFs of the contact and intercontact durations obtained by simulations with the aggregate CCDFs given by the empirical results, we will show in Section 4.2.1 that the CTMC model achieves the comparable accuracy as the power law model in predicting the pairwise connectivity.

2.2 Distributed Interest Estimation Model

Mobile users will only download their interested digital objects. Thus, when a user evaluates the D2D data offloading opportunity for an object, the interest of other users should be taken into account. Predicting user's interest has been studied and used for personalized news recommendations [29], [30]. However, the existing interest estimation models designed for recommendation systems are not suitable for D2D data offloading due to the limited computation capability of mobile devices and the additional communication resource consumptions in the wireless cellular network. Thus, a lightweight interest estimation model that enables mobile users to estimate the interest of others in a distributed manner is required. We propose a distributed interest estimation model to let a user estimate the interest of other users in each object that he is interested in and waiting to download from neighbors. Notice that other interest estimation models may also be used in our work to predict the interest of mobile users. In this case, however, the insights of using EAD metric to evaluate D2D data offloading opportunity and using it to prioritize D2D data offloading tasks are still useful.

Our distributed interest estimation model contains two aspects. The first one is based on social influence [31], [32]. A user is more likely to be interested in a digital object if his friends are interested in it. This is because a user may talk about his interested digital objects or share them in OSNs.

Meanwhile, the effect of the social influence to a mobile user from his friends may be different. The second aspect is based on Bayesian inference. Given the event that a user is not interested in an object which has been diffused on the OSN for a long time, the user may not be interested in the object in the future. Thus, our interest estimation model is a dynamic model over time. We make the following assumptions for our interest estimation model. First, we assume that friends of a user can influence the user independently. This assumption implies that a user is interested in an object as long as he is influenced by one of his friends. Second, we assume that the effect of social influence between a pair of users is determined by the similarity of their interest. These two assumptions have also been adopted in related works [33] and [34]. Various measures for the similarity of interest can be found in [35]. In this paper, we adopt the *Jaccard coefficient*, which has been used in [29], [35], [36]. The similarity of interest is related with both social influence and Bayesian inference aspects in our interest estimation model. Thus, we first introduce the Jaccard coefficient. We then present our interest estimation model. We also propose a distributed algorithm for mobile devices to determine the parameters required in our model.

2.2.1 Similarity of Interest

We use $\mathcal{O} = \{1, \ldots, O\}$ to denote the set of digital objects. Let random variable $Y_{i,k}$ denote the time when user $i \in \mathcal{U}$ reveals his interest in object $k \in \mathcal{O}$ and aims to download it via D2D data offloading. Note that the realization $y_{i,k}$ of $Y_{i,k}$ is not known *a priori*. Let $\mathcal{O}_i^t = \{k \,|\, y_{i,k} \leq t\}$ denote the set of digital objects that user i is interested in at time t (*i.e.*, set of objects that user i has completely downloaded or is waiting to download from his neighbors at time t). For users $i, j \in \mathcal{U}$, let $\theta_{i,j}$ denote the similarity of their interest. We first define $\mathcal{O}_i \triangleq \lim_{t\to\infty} \mathcal{O}_i^t$. Then, according to the definition of Jaccard coefficient in [37, pp. 61], we have $\theta_{i,j} = \frac{|\mathcal{O}_i \cap \mathcal{O}_j|}{|\mathcal{O}_i \cup \mathcal{O}_j|}$. Although it is clear that $\theta_{i,j} = 1$ if $i = j$, the value of $\theta_{i,j}$ cannot be accurately determined if $i \neq j$. The works in [38] and [39] have shown that the interest of users are reflected by their activities in the past. Since the server of OSN maintains the browsing history of the digital objects that a user has requested in the past [40], the server can estimate the similarity of interest between users i and j. We denote $\tilde{\theta}_{i,j}$ as the estimate of the Jaccard coefficient $\theta_{i,j}$. Thus, $\tilde{\theta}_{i,j}$ can be calculated by the server of OSN based on the browsing history of users i and j. Let $\widetilde{\mathbf{\Theta}} \in \mathbb{R}^{U\times U}$ denote the matrix with $\tilde{\theta}_{i,j}$ as the element (i, j). Users in set $\mathcal{U}$ can thus obtain matrix $\widetilde{\mathbf{\Theta}}$ from the server of OSN.

2.2.2 Distributed Interest Estimation — The Social Influence Aspect

Let binary random variable $I_{i,k}$ denote whether user $i \in \mathcal{U}$ is interested in object $k \in \mathcal{O}$ ($I_{i,k} = 1$) or not ($I_{i,k} = 0$). According to the definition of $\mathcal{O}_i$, we have $I_{i,k} = 1$ if $y_{i,k} < \infty$. Otherwise, we have $I_{i,k} = 0$ for $y_{i,k} = \infty$. Let l_k represent the URL of object $k \in \mathcal{O}$. We denote $\tau_{i,j}^t$ as the time when users i and j are connected most recently before time t (*i.e.*, $\tau_{i,j}^t \leq t$). Therefore, the set of URLs of the digital objects that user i is interested in before time $\tau_{i,j}^t$ is $\mathcal{K}_{i,j}^t \triangleq \{l_k \,|\, y_{i,k} \leq \tau_{i,j}^t\}$. In our model, user j (user i) needs to obtain set $\mathcal{K}_{i,j}^t$ (set $\mathcal{K}_{j,i}^t \triangleq \{l_k \,|\, y_{j,k} \leq \tau_{i,j}^t\}$) when users i and j are connected. However, the communication overhead caused by exchanging sets $\mathcal{K}_{j,i}^t$ and $\mathcal{K}_{i,j}^t$ between users i and j may increase fast. In Section 2.4, we will propose a lightweight algorithm to update sets $\mathcal{K}_{j,i}^t$ and $\mathcal{K}_{i,j}^t$ in an incremental manner to reduce the communication overhead.

The independent cascade model in [33] considers that people influence each other independently with different effects. This model has been widely adopted in the study of social networks [41], [42]. The effect of social influence between a pair of users can be evaluated by the similarity of their interest [34]. Let $\mathcal{Q}_i^t$ denote the set of objects that user $i \in \mathcal{U}$ is waiting to download from his neighbors at time t. For each object $k \in \mathcal{Q}_i^t$, user i categorizes other users into the following sets:

$$\mathcal{A}_{i,k}^t = \{j \,|\, j \in \mathcal{U} \setminus \{i\}, l_k \in \mathcal{K}_{j,i}^t\}, \tag{2a}$$

$$\mathcal{B}_{i,k}^t = \{j \,|\, j \in \mathcal{U} \setminus \{i\}, l_k \notin \mathcal{K}_{j,i}^t\}. \tag{2b}$$

Specifically, set $\mathcal{A}_{i,k}^t$ (set $\mathcal{B}_{i,k}^t$) maintained by user i at time t contains the users who are (are not) interested in object k before the most recent contact with user i. We denote random variable $R_{i,j,k}^t$ to present whether user i is informed that user j is interested in object k at time t. Thus, $R_{i,j,k}^t = 1$ if $j \in \mathcal{A}_{i,k}^t$. Otherwise, $R_{i,j,k}^t = 0$. We refer to vector $\boldsymbol{r}_{i,k}^t \triangleq (R_{i,1,k}^t, \ldots, R_{i,U,k}^t)$ with element $R_{i,i,k}^t = 1$ as the *interest record* of object k on user i at time t. Let $p_{i,j,k}^t \in [0,1]$ denote the estimate of user i at time t for the interest of user j in object k. Specifically, the interest estimate $p_{i,j,k}^t$ is the conditional probability of $I_{j,k} = 1$, given the interest record $\boldsymbol{r}_{i,k}^t$. That is, $p_{i,j,k}^t \triangleq \mathbb{P}(I_{j,k} = 1 \,|\, \boldsymbol{r}_{i,k}^t)$. Thus, we have $p_{i,j,k}^t = 1, \forall j \in \mathcal{A}_{i,k}^t$. Considering the social influence, we have the following lemma for user $j \in \mathcal{B}_{i,k}^t$:

Lemma 1: In the considered independent cascade model, the interest estimate $p_{i,j,k}^t$ obtained by user i at time t for user $j \in \mathcal{B}_{i,k}^t$ with digital object k is

$$p_{i,j,k}^t = 1 - \prod_{u \in \mathcal{A}_{i,k}^t \cup \{i\}} (1 - \tilde{\theta}_{u,j}), \tag{3}$$

where $\tilde{\theta}_{u,j}$ is the element (u, j) in matrix $\widetilde{\mathbf{\Theta}}$.

Proof: Please refer to Appendix A. ■

For user i who would like to download object k from his neighbors, user i estimates the interest of other users by using the independent cascade model with the set of users who have revealed their interest in object k to him (*i.e.*, users in set $\mathcal{A}_{i,k}^t \cup \{i\}$). This is commonly used in related works to model social influence [32], [41], [43]. Note that mobile users can estimate the interest of other users according to their own interest records. This enables each mobile user to perform interest estimation in a distributed manner. Thus, the values of $p_{i,j,k}^t$ and $p_{i',j,k}^t$ determined by users $i, i' \in \mathcal{U}$ may be different for $\boldsymbol{r}_{i,k}^t \neq \boldsymbol{r}_{i',k}^t$.

2.2.3 Distributed Interest Estimation — The Bayesian Inference Aspect

As mentioned earlier, the first aspect in our interest estimation model considers the social influence. We now introduce the second aspect which considers Bayesian inference. The basic idea is that if a user is not interested in an object which

$$q^t_{i,j,k} = p^t_{i,j,k}, \qquad \text{if } \tau^t_{i,j} \le x_k, \tag{4a}$$

$$q^t_{i,j,k} < \frac{(1-f(\tau^t_{i,j}-x_k))(1-\prod_{u\in\mathcal{A}^t_{i,k}\cup\{i\}}(1-\tilde{\theta}_{u,j}))}{(1-f(\tau^t_{i,j}-x_k))(1-\prod_{u\in\mathcal{A}^t_{i,k}\cup\{i\}}(1-\tilde{\theta}_{u,j}))+2\prod_{u\in\mathcal{A}^t_{i,k}\cup\{i\}}(1-\tilde{\theta}_{u,j})}, \qquad \text{if } \tau^t_{i,j} > x_k. \tag{4b}$$

has been diffused among users or over OSNs for a long time, then probably the user is not interested in the object. Information diffusion models have been studied in [44] and [45]. It has been shown that the time when a user shares his interested information to the OSN follows log-normal distribution $\ln\mathcal{N}(\mu=3.91, \sigma^2=6.86)$ after the information is initially posted [44]. Moreover, a user may not share an online video to the OSN until he has watched a part of the video that has been buffered. We thus consider that a mobile user shares his interested object when he has obtained ξ ($0\% < \xi \le 100\%$) or a higher percentage of the object, where ξ is a given constant. We use x_k to denote the time when object k is initially posted to the OSN. In our work, x_k is referred to as the *diffusion start time* of object k. It can be known by the users who are interested in object k, since x_k is recorded by servers of the OSN and can be conveyed to the users along with the URL of object k. Let $q^t_{i,j,k} \in [0,1]$ denote the estimate of user i at time t for the interest of user j in object k when both social influence and Bayesian inference are considered. Specifically, $q^t_{i,j,k}$ is defined as the conditional probability of $I_{j,k}=1$, given the interest record $\boldsymbol{r}^t_{i,k}$ and time $\tau^t_{i,j}$, *i.e.*, $q^t_{i,j,k} \triangleq \mathbb{P}(I_{j,k}=1 \mid \boldsymbol{r}^t_{i,k}, \tau^t_{i,j})$. We thus have $q^t_{i,j,k} = 1, \forall j \in \mathcal{A}^t_{i,k}$. For user $j \in \mathcal{B}^t_{i,k}$, we have the following theorem:

Theorem 1: According to the independent cascade model and information diffusion model, the estimate $q^t_{i,j,k}$ obtained by user i at time t for the interest of user $j \in \mathcal{B}^t_{i,k}$ in object k satisfies the equality in (4a) or the inequality in (4b) where $f(\varphi) = \mathrm{erf}(\frac{\ln\varphi - 3.91}{\sqrt{13.72}})$ and $\mathrm{erf}(\cdot)$ is the error function.

Proof: Please refer to Appendix B. ∎

We discuss the insight of Theorem 1 as follows. For user $i \in \mathcal{U}$ and digital object $k \in \mathcal{Q}^t_i$ at time t, we consider a user $j \in \mathcal{U}\backslash\{i\}$ who is not interested in object k at time t. Thus, URL $l_k \notin \mathcal{K}^t_{j,i}$. User i categorizes user j in set $\mathcal{B}^t_{i,k}$. If the most recent connection between users i and j is before the diffusion start time of object k, (*i.e.*, $\tau^t_{i,j} \le x_k$), then only social influence is considered (*i.e.*, $p^t_{i,j,k} = q^t_{i,j,k}$). If object k has been diffused for a long time but user j is still not interested in object k when users i and j are connected at time $\tau^t_{i,j}$ (*i.e.*, $\varphi = \tau^t_{i,j} - x_k$ is large), we have $\lim_{\varphi\to\infty} f(\varphi) = 1$. Hence, for an object k which has been diffused for a sufficient length of time and user j is not interested in it, the estimate obtained by user i that user j is interested in k approaches 0. To obtain a tractable $q^t_{i,j,k}$, we use the right-hand-side of (4b) to approximate $q^t_{i,j,k}$ when $\tau^t_{i,j} > x_k$. We will show in Section 4.2.2 that the interest estimates given in (4) obtains a much smaller estimation error compared with the interest estimates given in (3).

2.3 Pairwise Data Offloading Availability and the EAD Metric

2.3.1 Pairwise Data Offloading Availability

To define the pairwise data offloading availability, let random variable $Z^t_{i,j,k}(t') = 1$ (or $Z^t_{i,j,k}(t') = 0$) denote the event that object $k \in \mathcal{Q}^t_i$ can (or cannot) be downloaded by user i from user j at time $t' \ge t$. Without loss of generality, we also consider that user j can transmit data of object k to other users if ξ or a higher percentage of object k has been obtained by user j. Besides, two users who have partially downloaded the same object are assumed to have non-overlapped portions to transfer to each other. Although this assumption can be relaxed by applying network coding technique on D2D data offloading [46], it is beyond the scope of this paper. The pairwise data offloading availability for user i to download object k from user j at time $t' \ge t$ is defined as $z^t_{i,j,k}(t') \triangleq \mathbb{P}(Z^t_{i,j,k}(t')=1 \mid \boldsymbol{r}^t_{i,k}, \tau^t_{i,j}, C^t_{i,j})$. We assume that the stochastic D2D connections between mobile users are independent from both the interest of users and the diffusion of digital objects. We have the following theorem:

Theorem 2: For user i, given the interest record $\boldsymbol{r}^t_{i,k}$, the time $\tau^t_{i,j}$ when users i and j contacted most recently, and connection state $C^t_{i,j}$ between users i and j at time t, the pairwise data offloading availability for user i to download object k from user j at time $t' \ge t$ is

$$z^t_{i,j,k}(t') = \begin{cases} \frac{1+f(t'-x_k)}{2}\mathbb{P}(C^{t'}_{i,j}=1 \mid C^t_{i,j}), & \text{if } j \in \mathcal{A}^t_{i,k}, \\ \frac{q^t_{i,j,k}(1+f(t'-x_k))}{2}\mathbb{P}(C^{t'}_{i,j}=1 \mid C^t_{i,j}), & \text{if } j \in \mathcal{B}^t_{i,k}, \end{cases} \tag{5}$$

where $\mathbb{P}(C^{t'}_{i,j}=1 \mid C^t_{i,j})$ is given by (1).

Proof: Please refer to Appendix C. ∎

When user i is connected with others, he can keep a record of the neighboring users who have already downloaded ξ or a higher percentage of object k. When user i notices that user j has downloaded ξ or a higher percentage of object k, the pairwise data offloading availability for user i to download data from user j at time t' becomes $z^t_{i,j,k}(t') = \mathbb{P}(C^{t'}_{i,j}=1 \mid C^t_{i,j})$ by following similar steps in the proof of Theorem 2. Note that $\mathcal{A}^t_{i,k}$ in (5) is the set of users who have informed user i that they are interested in object k at time t. Therefore, the set of users who have downloaded ξ or a higher percentage of object k is a subset of $\mathcal{A}^t_{i,k}$. The case of $z^t_{i,j,k}(t') = \mathbb{P}(C^{t'}_{i,j}=1 \mid C^t_{i,j})$ is thus a special case of $j \in \mathcal{A}^t_{i,k}$ in (5), where the accuracy of the pairwise D2D data offloading availability can be improved. It can be seen from (5) that other pairwise connectivity models may also be used in our work. Specifically, given the current connection state between users i and j at time t (*i.e.*, $C^t_{i,j}$), as long as a pairwise connectivity model returns the probability that users i and j are connected at future time $t' \ge t$, it can be directly used in our work.

2.3.2 EAD Metric

Users who are interested in the same digital object may have different MWT, *i.e.*, the time that they prefer to wait to download the object via D2D data offloading before downloading it from the wireless cellular network. Let $\delta_{i,k}$ denote the MWT of user $i \in \mathcal{U}$ for object $k \in \mathcal{O}_i$. When user i reveals his interest in object k at time $y_{i,k}$, the MWT $\delta_{i,k}$

is also given by the user. After trying to download object k via D2D data offloading from time $y_{i,k}$ to $y_{i,k}+\delta_{i,k}$, the remaining part of object k, which has not yet been obtained by user i from neighboring devices, will be downloaded by user i from wireless cellular network.

Now, we first define $z^t_{i,k}(t') \triangleq 1-\prod_{j\in\mathcal{U}\backslash\{i\}}\left(1-z^t_{i,j,k}(t')\right)$, which is referred to as the neighborhood data availability of object k for user i at time $t' \geq t$. Note that $z^t_{i,k}(t') \in [0,1]$. The value of $z^t_{i,k}(t')$ is the probability that user i can download object k from at least one neighbor at time t'. We then define the EAD of object k for user i at time t as follows:

$$V_{i,k}\left(t\right) \triangleq \int_t^{y_{i,k}+\delta_{i,k}} z^t_{i,k}(\tau)\,\mathrm{d}\tau. \tag{6}$$

In the next subsection, we propose a lightweight URL exchanging algorithm to facilitate each user obtaining the set of URLs for the objects that are of interest to other users. Thus, user i can determine the value of $V_{i,k}\left(t\right)$ for object k at time t in a distributed manner.

Algorithm 1: The algorithm that user $i \in \mathcal{U}$ uses to obtain $\mathcal{K}^t_{j,i}$ from user $j \in \mathcal{U}\backslash\{i\}$.

```
1  Initialize t := 0, τ^t_{i,j} := 0, O^t_i := ∅, and K^t_{j,i} := ∅.
2  Start to increase t according to system clock.
3  Loop
4      for k ∈ {κ | y_{i,κ} = t} do
5          Include object k into set O^t_i.
6      if user j becomes a new neighbor of user i at time t
       then
7          if τ^t_{i,j} = 0 then
8              Send set L^t_{i,j} := {l_k | k ∈ O^t_i} to user j.
9              Receive set L^t_{j,i} from user j.
10             K^t_{j,i} := L^t_j, τ^t_{i,j} := t.
11         else
12             Send set L^t_{i,j} := {l_k | τ^t_{i,j} < y_{i,k} ≤ t, k ∈ O^t_i} to
               user j.
13             Receive set L^t_{j,i} from user j.
14             K^t_{j,i} := K^t_{j,i} ∪ L^t_{j,i}, τ^t_{i,j} := t.
```

2.4 URL Exchanging Algorithm

In Algorithm 1, we present a lightweight algorithm that let mobile user $i \in \mathcal{U}$ obtain the set of URLs $\mathcal{K}^t_{j,i}$ from user $j \in \mathcal{U}\backslash\{i\}$. For each instance of time t, those objects that user i is interested in and aims to download by D2D data offloading are included in set $\mathcal{O}^t_i$ (Lines 4–5). If user i is connected with user j for the first time (Line 7), set $\mathcal{L}^t_{i,j}$, which contains the URLs of the objects that user i is interested in, is sent to user j (Line 8). Similarly, user j sends set $\mathcal{L}^t_{j,i}$ to user i. This set contains the URLs of the objects that user j is interested in. After receiving $\mathcal{L}^t_{j,i}$ (Line 9), user i updates $\mathcal{K}^t_{j,i}$ to $\mathcal{L}^t_{j,i}$ and the time of the most recent contact with user j is updated to time t (Line 10). If user i is connected with user j again at time t (Line 11), the set of URLs of the objects that user i is interested in since his last contact with user j is assigned to $\mathcal{L}^t_{i,j}$ and sent to user j (Line 12). Similarly, set $\mathcal{L}^t_{j,i}$ is sent by user j and received by user i (Line 13). Set $\mathcal{K}^t_{j,i}$ and the value of $\tau^t_{i,j}$ are then updated by user i (Line 14).

Consider users 1 and 2 in Fig. 2 as an example. At time t_1, user 1 is not interested in any object and user 2 has already obtained objects 1 and 2. That is, we have $y_{2,1} < t_1$, $y_{2,2} < t_1$, $\mathcal{O}^{t_1}_1 = \emptyset$, and $\mathcal{O}^{t_1}_2 = \{1,2\}$. We assume that users 1 and 2 are not neighbors at time t_1. We thus have $\tau^{t_1}_{1,2} = 0$ and $\mathcal{K}^{t_1}_{1,2} = \mathcal{K}^{t_1}_{2,1} = \emptyset$. At time t_2, user 2 is interested in object 3 (*i.e.*, $y_{2,3} = t_2$), thus $\mathcal{O}^{t_2}_2 = \{1,2,3\}$. Similarly, when user 1 is interested in objects 1 and 2 at time t_3 (*i.e.*, $y_{1,1} = y_{1,2} = t_3$), we have $\mathcal{O}^{t_3}_1 = \{1,2\}$. Users 1 and 2 are connected at time t_4. Since $\tau^{t_4}_{1,2} = 0$, user 1 sends the set of URLs $\mathcal{L}^{t_4}_{1,2} = \{l_1, l_2\}$ to user 2 due to $\mathcal{O}^{t_4}_1 = \{1,2\}$. Since $\mathcal{O}^{t_4}_2 = \{1,2,3\}$, the set of URLs sent from user 2 to user 1 is $\mathcal{L}^{t_4}_{2,1} = \{l_1,l_2,l_3\}$. When $\mathcal{L}^{t_4}_{2,1}$ is received by user 1, $\mathcal{K}^{t_4}_{2,1}$ on user 1 is updated to $\{l_1,l_2,l_3\}$. Then, $\tau^{t_4}_{1,2}$ is updated to time t_4. Meanwhile, user 2 also receives $\mathcal{L}^{t_4}_{1,2} = \{l_1,l_2\}$ from user 1, so he updates $\mathcal{K}^{t_4}_{1,2} = \{l_1,l_2\}$ and $\tau^{t_4}_{1,2} = t_4$ by conducting similar procedures as user 1. Users 1 and 2 are disconnected at time t_6. After a period of time, user 1 is interested in object 3 at time t_{11} (*i.e.*, $y_{1,3} = t_{11}$). Thus, $\mathcal{O}^{t_{11}}_i = \{1,2,3\}$. When users 1 and 2 are connected again at time t_{12}, since $\tau^{t_{12}}_{1,2} = t_4 \neq 0$, user 1 incrementally sends set $\mathcal{L}^{t_{12}}_{1,2} = \{l_3\}$ to user 2. Since user 2 is not interested in any object from time t_4 to t_{12}, so $\mathcal{L}^{t_{12}}_{2,1} = \emptyset$ is sent from user 2 to user 1. Therefore, we have $\mathcal{K}^{t_{12}}_{2,1} = \{l_1,l_2,l_3\}$. Similarly, user 2 updates $\mathcal{K}^{t_{12}}_{1,2}$ by $\{l_1,l_2\} \cup \mathcal{L}^{t_{12}}_{1,2} = \{l_1,l_2,l_3\}$. Then, users 1 and 2 update the value of $\tau^{t_{12}}_{1,2}$ to time t_{12}.

3 Selecting a Digital Object to First Download in a Neighborhood

User $i \in \mathcal{U}$ may be interested in multiple digital objects. Even if multiple digital objects in set $\mathcal{Q}^t_i$ are available in the neighborhood of user i at time t, user i can obtain the data for only one of them at any instance of time. In practice, we consider a time-slotted system where the time slots are denoted by index values $s = 1, 2, \cdots$. In each time slot, user i selects a digital object available in his neighborhood to download. Let t_s denote the beginning of time slot s. We assume that user i reveals his interest in an object at the beginning of time slots. We also assume that the D2D connections change at the beginning of time slots. We propose that user i in time slot s first downloads the data of object $k \in \mathcal{Q}^{t_s}_i$ which has the smallest EAD (*i.e.*, smallest value of $V_{i,k}(t_s)$ determined by substituting t_s for t in (6)). We denote $\xi^{t_s}_{i,k}$ ($0\% \leq \xi^{t_s}_{i,k} \leq 100\%$) as the percentage of object $k \in \mathcal{O}$ that has been obtained by user $i \in \mathcal{U}$ at the beginning of time slot s. Hence, user i selects an object to download in time slot s by solving the following problem:

$$\begin{aligned} &\underset{k\in\mathcal{Q}^{t_s}_i}{\text{argmin}} && V_{i,k}\left(t_s\right) \\ &\text{subject to} && \left\{j \,\middle|\, C^{t_s}_{i,j} = 1, \xi^{t_s}_{j,k} \geq \xi\right\} \neq \emptyset, \end{aligned} \tag{7}$$

where the constraint ensures that the object that user i chooses to download in time slot s must be downloaded by at least one neighbor who has received at least ξ percentage of the object. In particular, user i can obtain the value of $\xi^{t_s}_{j,k}$ from neighboring user j by considering that users i and j have to be connected in time slot s (*i.e.*, $C^{t_s}_{i,j}=1$) in order to perform D2D data transfer. For a mobile user, his D2D link capacity to neighbors may be different and changing over

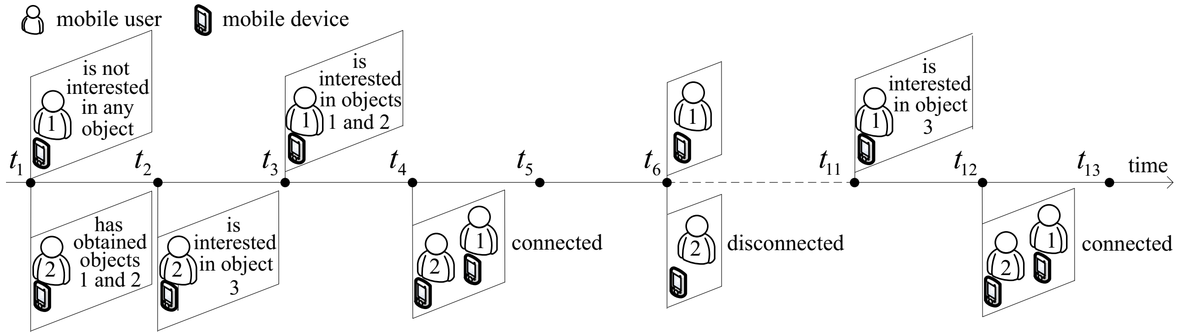


Fig. 2. An example for Algorithms 1 and 2.

time. Taking the variation of D2D link capacity into account to select an object to download is an extension of the current model, which requires a stochastic optimization framework and is beyond the scope of this paper.

The objective function of problem (7), which is given by (6), is not in closed-form. Since user i needs to solve problem (7) for each time slot, the computational complexity is high. To reduce the computational complexity, user i can approximate the value of $V_{i,k}(t_s)$ for time slot $s > 1$ by using $V_{i,k}(t_{\widehat{s}})$ that has been determined in an earlier time slot $\widehat{s} < s$ for object $k \in \mathcal{Q}_i^{t_{\widehat{s}}} \cap \mathcal{Q}_i^{t_s}$. We first define $\boldsymbol{c}_i^t \triangleq (C_{i,1}^t, \ldots, C_{i,i-1}^t, C_{i,i+1}^t, \ldots, C_{i,U}^t)$. Vector $\boldsymbol{c}_i^t$ is referred to as connectivity profile of user i at time t. We consider the case that connectivity profile of user i remains unchanged in time slots $\widehat{s}, \widehat{s}+1, \ldots, s$. We have $\boldsymbol{r}_{i,k}^{t_s} = \cdots = \boldsymbol{r}_{i,k}^{t_{\widehat{s}+1}} = \boldsymbol{r}_{i,k}^{t_{\widehat{s}}}$ as long as $\boldsymbol{c}_i^{t_s} = \cdots = \boldsymbol{c}_i^{t_{\widehat{s}+1}} = \boldsymbol{c}_i^{t_{\widehat{s}}}$ due to $\mathcal{K}_{j,i}^{t_s} = \cdots = \mathcal{K}_{j,i}^{t_{\widehat{s}+1}} = \mathcal{K}_{j,i}^{t_{\widehat{s}}}$ for each user $j \in \mathcal{U} \backslash \{i\}$. Thus, we have $\mathcal{A}_{i,k}^{t_s} = \cdots = \mathcal{A}_{i,k}^{t_{\widehat{s}+1}} = \mathcal{A}_{i,k}^{t_{\widehat{s}}}$ and $\mathcal{B}_{i,k}^{t_s} = \cdots = \mathcal{B}_{i,k}^{t_{\widehat{s}+1}} = \mathcal{B}_{i,k}^{t_{\widehat{s}}}$ for each object $k \in \mathcal{Q}_i^{t_{\widehat{s}}} \cap \mathcal{Q}_i^{t_s}$ according to (2). Now, we determine $z_{i,j,k}^{t_s}(\tau)$ for time slot s and $z_{i,j,k}^{t_{\widehat{s}}}(\tau)$ for time slot $\widehat{s}$ at $\tau \in (t_s, y_{i,k} + \delta_{i,k})$ by first substituting τ for t' in (5) and then substituting t_s and $t_{\widehat{s}}$ for t in the obtained result, respectively. We find that the difference between $z_{i,j,k}^{t_s}(\tau)$ and $z_{i,j,k}^{t_{\widehat{s}}}(\tau)$ is caused by the difference between $\mathbb{P}(C_{i,j}^\tau = 1 \,|\, C_{i,j}^{t_s})$ and $\mathbb{P}(C_{i,j}^\tau = 1 \,|\, C_{i,j}^{t_{\widehat{s}}})$ for both cases of j in (5) due to $q_{i,j,k}^{t_s} = \cdots = q_{i,j,k}^{t_{\widehat{s}+1}} = q_{i,j,k}^{t_{\widehat{s}}}$.

Let $\alpha_{i,j,k}(\tau, t_{\widehat{s}})$ denote the first order partial derivative of $z_{i,j,k}^{t_{\widehat{s}}}(\tau)$ at $\tau \in (t_s, y_{i,k} + \delta_{i,k})$ with respect to (*w.r.t.*) $t_{\widehat{s}}$ for $j \in \mathcal{A}_{i,k}^{t_{\widehat{s}}}$. That is, we consider the case $z_{i,j,k}^{t_{\widehat{s}}}(\tau) = \frac{1+f(\tau - x_k)}{2} \mathbb{P}(C_{i,j}^\tau = 1 \,|\, C_{i,j}^{t_{\widehat{s}}})$ in (5). By considering (1), $\alpha_{i,j,k}(\tau, t_{\widehat{s}})$ is given by

$$
\begin{aligned}
&\alpha_{i,j,k}(\tau, t_{\widehat{s}}) \\
&= \frac{\partial\, z_{i,j,k}^{t_{\widehat{s}}}(\tau)}{\partial\, t_{\widehat{s}}} \\
&= \begin{cases} -\frac{1+f(\tau-x_k)}{2} \lambda_{i,j} e^{-(\lambda_{i,j}+\mu_{i,j})(\tau - t_{\widehat{s}})}, & \text{if } C_{i,j}^{t_{\widehat{s}}} = 0, \\ \frac{1+f(\tau-x_k)}{2} \mu_{i,j} e^{-(\lambda_{i,j}+\mu_{i,j})(\tau - t_{\widehat{s}})}, & \text{if } C_{i,j}^{t_{\widehat{s}}} = 1. \end{cases}
\end{aligned} \tag{8}
$$

The first order partial derivative of $z_{i,j,k}^{t_{\widehat{s}}}(\tau)$ at $\tau \in (t_s, y_{i,k} + \delta_{i,k})$ *w.r.t.* $t_{\widehat{s}}$ for $j \in \mathcal{B}_{i,k}^{t_{\widehat{s}}}$ is denoted and given by $\beta_{i,j,k}(\tau, t_{\widehat{s}}) = q_{i,j,k}^{t_{\widehat{s}}} \alpha_{i,j,k}(\tau, t_{\widehat{s}})$. By considering the Taylor series, we have $z_{i,j,k}^{t_s}(\tau) = \sum_{m=0}^\infty \frac{\partial^m z_{i,j,k}^{t_{\widehat{s}}}(\tau)}{\partial (t_{\widehat{s}})^m} \frac{(t_s - t_{\widehat{s}})^m}{m!}$. Thus, we have

$$
\begin{aligned}
&z_{i,j,k}^{t_s}(\tau) - z_{i,j,k}^{t_{\widehat{s}}}(\tau) \\
&= (t_s - t_{\widehat{s}}) \\
&\times \begin{cases} \alpha_{i,j,k}(\tau, t_{\widehat{s}}) + \sum_{m=1}^\infty \frac{\partial^m \alpha_{i,j,k}(\tau, t_{\widehat{s}})}{\partial (t_{\widehat{s}})^m} \frac{(t_s - t_{\widehat{s}})^m}{(m+1)!}, & \text{if } j \in \mathcal{A}_{i,k}^{t_{\widehat{s}}}, \\ \beta_{i,j,k}(\tau, t_{\widehat{s}}) + \sum_{m=1}^\infty \frac{\partial^m \beta_{i,j,k}(\tau, t_{\widehat{s}})}{\partial (t_{\widehat{s}})^m} \frac{(t_s - t_{\widehat{s}})^m}{(m+1)!}, & \text{if } j \in \mathcal{B}_{i,k}^{t_{\widehat{s}}}. \end{cases}
\end{aligned} \tag{9}
$$

We find that the connection profile of user i remains unchanged for only short time intervals (*i.e.*, $t_s - t_{\widehat{s}}$ is small). Moreover, as we will show in Section 4.2.1, the values of $\lambda_{i,j}$ and $\mu_{i,j}$ in (8) have the order of magnitude of 10^{-4} and 10^{-3}, respectively. Thus, $\sum_{m=1}^\infty \frac{\partial^m \alpha_{i,j,k}(\tau, t_{\widehat{s}})}{\partial (t_{\widehat{s}})^m} \frac{(t_s - t_{\widehat{s}})^m}{(m+1)!}$ and $\sum_{m=1}^\infty \frac{\partial^m \beta_{i,j,k}(\tau, t_{\widehat{s}})}{\partial (t_{\widehat{s}})^m} \frac{(t_s - t_{\widehat{s}})^m}{(m+1)!}$ in (9) are much smaller than $\alpha_{i,j,k}(\tau, t_{\widehat{s}})$ and $\beta_{i,j,k}(\tau, t_{\widehat{s}})$, respectively. Therefore, we approximate the difference between $z_{i,j,k}^{t_s}(\tau)$ and $z_{i,j,k}^{t_{\widehat{s}}}(\tau)$ by

$$
z_{i,j,k}^{t_s}(\tau) - z_{i,j,k}^{t_{\widehat{s}}}(\tau) \approx (t_s - t_{\widehat{s}}) \times \begin{cases} \alpha_{i,j,k}(\tau, t_{\widehat{s}}), & \text{if } j \in \mathcal{A}_{i,k}^{t_{\widehat{s}}}, \\ \beta_{i,j,k}(\tau, t_{\widehat{s}}), & \text{if } j \in \mathcal{B}_{i,k}^{t_{\widehat{s}}}. \end{cases} \tag{10}
$$

That is, we have $z_{i,j,k}^{t_s}(\tau) - z_{i,j,k}^{t_{\widehat{s}}}(\tau) \ll 1$. This means if $\boldsymbol{c}_i^{t_s} = \cdots = \boldsymbol{c}_i^{t_{\widehat{s}+1}} = \boldsymbol{c}_i^{t_{\widehat{s}}}$, the value of the integrand in (6) for object $k \in \mathcal{Q}_i^{t_{\widehat{s}}} \cap \mathcal{Q}_i^{t_s}$ changes slightly from $t_{\widehat{s}}$ to t_s. Note that the above analysis also applies when we substitute $\widetilde{s} \in \{\widehat{s}+1, \widehat{s}+2, \ldots, s-1\}$ for $\widehat{s}$. Without loss of generality, when $V_{i,k}(t_{\widehat{s}})$ is evaluated in time slot $\widehat{s}$ by substituting $t_{\widehat{s}}$ for t in (6), we define $\rho_{i,k} \triangleq V_{i,k}(t_{\widehat{s}}) / (y_{i,k} + \delta_{i,k} - t_{\widehat{s}})$ for object $k \in \mathcal{Q}_i^{t_{\widehat{s}}}$. Note that we have $0 < \rho_{i,k} < 1$. The value of $\rho_{i,k}$ represents the ratio between the time that object k is in the neighborhood of user i and the duration from current time to time $y_{i,k} + \delta_{i,k}$. We then approximate the value of $V_{i,k}(t_s)$ in time slot $s > \widehat{s}$ by $(y_{i,k} + \delta_{i,k} - t_s)\rho_{i,k}$ for object $k \in \mathcal{Q}_i^{t_s} \cap \mathcal{Q}_i^{t_{\widehat{s}}}$ if $\boldsymbol{c}_i^{t_s} = \cdots = \boldsymbol{c}_i^{t_{\widehat{s}+1}} = \boldsymbol{c}_i^{t_{\widehat{s}}}$. We define

$$
W_{i,k}(t_s) \triangleq \begin{cases} (y_{i,k} + \delta_{i,k} - t_s)\rho_{i,k}, & \text{if } k \in \mathcal{Q}_i^{t_s} \cap \mathcal{Q}_i^{t_{\widehat{s}}}, \boldsymbol{c}_i^{t_s} = \cdots = \boldsymbol{c}_i^{t_{\widehat{s}+1}} = \boldsymbol{c}_i^{t_{\widehat{s}}}, \\ V_{i,k}(t_s), & \text{otherwise.} \end{cases} \tag{11}
$$

Now, user i solves the following problem in time slot s instead:

$$
\begin{aligned}
&\underset{k \in \mathcal{Q}_i^{t_s}}{\text{argmin}} && W_{i,k}(t_s) \\
&\text{subject to} && \left\{ j \,|\, C_{i,j}^{t_s} = 1, \xi_{j,k}^{t_s} \geq \xi \right\} \neq \emptyset.
\end{aligned} \tag{12}
$$

For user i in time slot s, the objective function of problem (7) is a special case of the objective function of problem

Algorithm 2: D2D data offloading algorithm on mobile device i.

1 **Initialize** t, $s:=0$, $\mathcal{Q}_i^{t_s}:=\emptyset$, $\widehat{s}:=0$, $\mathcal{Q}_i^{t_{\widehat{s}}}:=\emptyset$, and $\boldsymbol{c}_i^{t_{\widehat{s}}}:=\mathbf{0}$.
2 Start to increase t according to system clock.
3 **Loop** for each time slot
4 $\quad s:=s+1$, $\mathcal{Q}_i^{t_s}:=\mathcal{Q}_i^{t_{s-1}}$.
5 $\quad$ **for** $j\in\mathcal{U}\backslash\{i\}$ **do**
6 $\quad\quad$ **if** user i detects user j in the neighborhood **then**
7 $\quad\quad\quad C_{i,j}^{t_s}:=1$.
8 $\quad\quad\quad$ Obtain $\mathcal{K}_{j,i}^{t_s}$ and $\tau_{i,j}^{t_s}$ from Algorithm 1 at time $t=t_s$.
9 $\quad\quad$ **else**
10 $\quad\quad\quad C_{i,j}^{t_s}:=0$.
11 $\quad$ **for** $k\in\{\kappa\,|\,y_{i,\kappa}=t_s\}$ **do**
12 $\quad\quad$ **Initialize** $\delta_{i,k}$ by the MWT preferred by user i for object k.
13 $\quad\quad \mathcal{Q}_i^{t_s}:=\mathcal{Q}_i^{t_s}\cup\{k\}$.
14 $\quad$ **if** $\boldsymbol{c}_i^{t_s}\neq\boldsymbol{c}_i^{t_{\widehat{s}}}$ **then**
15 $\quad\quad$ **for** $k\in\mathcal{Q}_i^{t_s}$ **do**
16 $\quad\quad\quad$ Determine $W_{i,k}\left(t_s\right):=V_{i,k}\left(t_s\right)$ by substituting t_s for t in (6), $\rho_{i,k}:=\frac{V_{i,k}(t_s)}{y_{i,k}+\delta_{i,k}-t_s}$.
17 $\quad\quad \widehat{s}:=s$, $\boldsymbol{c}_i^{t_{\widehat{s}}}:=\boldsymbol{c}_i^{t_s}$, $\mathcal{Q}_i^{t_{\widehat{s}}}:=\mathcal{Q}_i^{t_s}$.
18 $\quad$ **else**
19 $\quad\quad$ **for** $k\in\{\kappa\,|\,y_{i,\kappa}=t_s\}$ **do**
20 $\quad\quad\quad$ Determine $W_{i,k}\left(t_s\right):=V_{i,k}\left(t_s\right)$ by substituting t_s for t in (6), $\rho_{i,k}:=\frac{V_{i,k}(t_s)}{y_{i,k}+\delta_{i,k}-t_s}$.
21 $\quad\quad$ **for** $k\in\mathcal{Q}_i^{t_{\widehat{s}}}$ **do**
22 $\quad\quad\quad$ Determine $W_{i,k}\left(t_s\right):=\left(y_{i,k}+\delta_{i,k}-t_s\right)\rho_{i,k}$.
23 $\quad\quad \mathcal{Q}_i^{t_{\widehat{s}}}:=\mathcal{Q}_i^{t_s}$.
24 $\quad$ Search $k^\star:=\operatorname{argmin}_k W_{i,k}\left(t_s\right)$ which can be downloaded from the neighborhood.
25 $\quad$ Download object $k^\star$ via D2D data offloading in time slot s.
26 $\quad$ **if** $k^\star$ is completely downloaded **then**
27 $\quad\quad \mathcal{Q}_i^{t_s}:=\mathcal{Q}_i^{t_s}\backslash\{k^\star\}$.
28 $\quad$ **for** $k\in\{\kappa\,|\,t_s\leq y_{i,\kappa}+\delta_{i,\kappa}<t_{s+1}\}$ **do**
29 $\quad\quad \mathcal{Q}_i^{t_s}:=\mathcal{Q}_i^{t_s}\backslash\{k\}$ and download the remaining data of object k from the wireless cellular network.

(12) when either an object attracts the interest of user i or the connectivity profile $\boldsymbol{c}_i^{t_s}$ is changed in time slot s. When multiple time slots are considered, solving problem (12) is simpler because problem (7) requires user i to evaluate (6) for each object $k\in\mathcal{Q}_i^{t_s}$ in each time slot s. In fact, the connectivity profile changes in a much larger scale compared with a time slot. Thus, solving problem (12) has less computational complexity than solving problem (7). Our D2D data offloading algorithm for mobile device i is given in Algorithm 2. For each time slot s, the digital objects that the user is waiting to download via D2D data offloading are those that have not been completely downloaded in the previous time slot (Line 4). Mobile device i updates the connectivity profile by determining the connection state with each user $j\in\mathcal{U}\backslash\{i\}$ (Lines 6–9). Then, the objects that attract the interest of user i in time slot s are initialized and included in set $\mathcal{Q}_i^{t_s}$ (Lines 11–13). If the connectivity profile of user i changes (Line 14), the value $\rho_{i,k}$ is calculated after evaluating $V_{i,k}(t_s)$ and assigning $V_{i,k}(t_s)$ to $W_{i,k}(t_s)$ for each object in set $\mathcal{Q}_i^{t_s}$ (Lines 15–16). After that, the time slot in which the connectivity profile is changed and the new connectivity profile are saved in $\widehat{s}$ and $\boldsymbol{c}_i^{t_{\widehat{s}}}$, respectively. In addition, the set of objects whose EADs have been calculated by user i is saved in $\mathcal{Q}_i^{t_{\widehat{s}}}$ (Line 17). On the other hand, if the connectivity profile of user i does not change (Line 18), $W_{i,k}\left(t_s\right)$ is determined by calculating $V_{i,k}(t_s)$ only for the objects that attract the interest of user i in time slot s (Lines 19–20) and approximated for the objects in set $\mathcal{Q}_i^{t_{\widehat{s}}}$ by $\left(y_{i,k}+\delta_{i,k}-t_s\right)\rho_{i,k}$ (Lines 21–22). The set $\mathcal{Q}_i^{t_{\widehat{s}}}$ is then updated by set $\mathcal{Q}_i^{t_s}$ (Line 23). Mobile device i selects an available object $k^\star$ with the smallest value of $W_{i,k}\left(t_s\right)$ in the neighborhood to download in the current time slot s (Lines 24–25). The object $k^\star$ is removed from set $\mathcal{Q}_i^{t_s}$ if it has been completely downloaded (Lines 26–27). Mobile device i also removes each object which has not been completely downloaded within its MWT from set $\mathcal{Q}_i^{t_s}$. It then downloads the remaining data of these objects from the cellular network (Lines 28–29).

Let us consider user 1 in Fig. 2 as an example. Since user 1 has no neighbor and is not interested in any object in time slot 1, we have $\mathcal{Q}_1^{t_1}=\emptyset$, $\widehat{s}=0$, and $\boldsymbol{c}_1^{t_0}=0$. User 1 updates $s=2$ and $\mathcal{Q}_1^{t_2}=\emptyset$ in time slot 2. Then, user 1 sets $C_{1,2}^{t_2}=0$ as he has no neighbor in time slot 2. That is, he has $\boldsymbol{c}_1^{t_2}=0$. User 1 is not interested in any object in time slot 2 (*i.e.*, $\{\kappa\,|\,y_{1,\kappa}=t_2\}=\emptyset$). Meanwhile, as $\boldsymbol{c}_1^{t_2}=\boldsymbol{c}_1^{t_0}$, user 1 sets $\mathcal{Q}_1^{t_0}=\emptyset$ and no other operation is required. At time slot 3, user 1 is interested in objects 1 and 2 (*i.e.*, $y_{1,1}=y_{1,2}=t_3$) and initializes $\delta_{1,1}$ and $\delta_{1,2}$ by his preferred MWTs. He then includes both objects in set $\mathcal{Q}_1^{t_3}$. We assume $\delta_{1,1}=t_6-t_3$ and $\delta_{1,2}=t_{11}-t_3$. Since $\boldsymbol{c}_1^{t_3}=\boldsymbol{c}_1^{t_0}$ and $y_{1,2}=y_{1,3}=t_3$, both $W_{1,1}(t_3)$ and $W_{1,2}(t_3)$ are determined by evaluating $V_{1,1}(t_3)$ and $V_{1,2}(t_3)$, respectively. Then, $\rho_{1,1}$ and $\rho_{1,2}$ are calculated. At time slot 4, user 1 has $\mathcal{Q}_1^{t_4}=\{1,2\}$. Then, $C_{1,2}^{t_4}=1$ is detected (*i.e.*, $\boldsymbol{c}_i^{t_4}=1$). User 1 thus obtains $\mathcal{K}_{2,1}^{t_4}$ and $\tau_{1,2}^{t_4}$ by using Algorithm 1. Since $\boldsymbol{c}_i^{t_4}\neq\boldsymbol{c}_i^{t_0}$, $W_{1,1}(t_4)$ and $W_{1,2}(t_4)$ are calculated by evaluating $V_{1,1}(t_4)$ and $V_{1,2}(t_4)$, respectively, and $\rho_{1,1}$ and $\rho_{1,2}$ are updated. User 1 then updates $\widehat{s}=4$, $\boldsymbol{c}_1^{t_4}=1$, and $\mathcal{Q}_1^{t_4}=\{1,2\}$. We assume $W_{1,1}(t_4)<W_{1,2}(t_4)$, so user 1 downloads object 1 from user 2 in time slot 4. We also assume that object 2 is not completely downloaded. We thus have $\mathcal{Q}_i^{t_5}=\{1,2\}$ at the beginning of time slot 5. Due to $C_{1,2}^{t_5}=1$, we have $\boldsymbol{c}_i^{t_5}=\boldsymbol{c}_i^{t_4}$. The values of $W_{1,1}^{t_5}$ and $W_{1,2}^{t_5}$ are approximated by $(t_6-t_5)\rho_{1,1}$ and $(t_{11}-t_5)\rho_{1,2}$, respectively. Given $W_{1,1}^{t_5}<W_{1,2}^{t_5}$, user 1 continues to download object 1 from user 2 in time slot 5. Now, we consider that object 1 is completely downloaded at the end of time slot 5. Set $\mathcal{Q}_1^{t_5}$ is then updated to $\{1,2\}\setminus\{1\}=\{2\}$. Hence, user 1 has $\mathcal{Q}_1^{t_6}=\{2\}$ in time slot 6. As $C_{1,2}^{t_6}=0$ is detected, we have $\boldsymbol{c}_1^{t_6}\neq\boldsymbol{c}_1^{t_4}$. The value of $W_{1,2}(t_6)$ is determined by evaluating $V_{1,2}(t_6)$ and $\rho_{1,2}$ is updated. User 1 updates $\widehat{s}=6$, $\boldsymbol{c}_1^{t_6}=0$, and $\mathcal{Q}_1^{t_6}=\{2\}$. User 1 downloads object 2 from the cellular network at the end of time slot 11 since its MWT for object 2 is reached. He then updates $\mathcal{Q}_1^{t_{11}}=\emptyset$.

4 Trace-Driven Simulation Results

We use the real-world traces in Cambridge/Haggle dataset [47] to reproduce the topology of D2D connections between mobile devices. We first introduce the dataset and

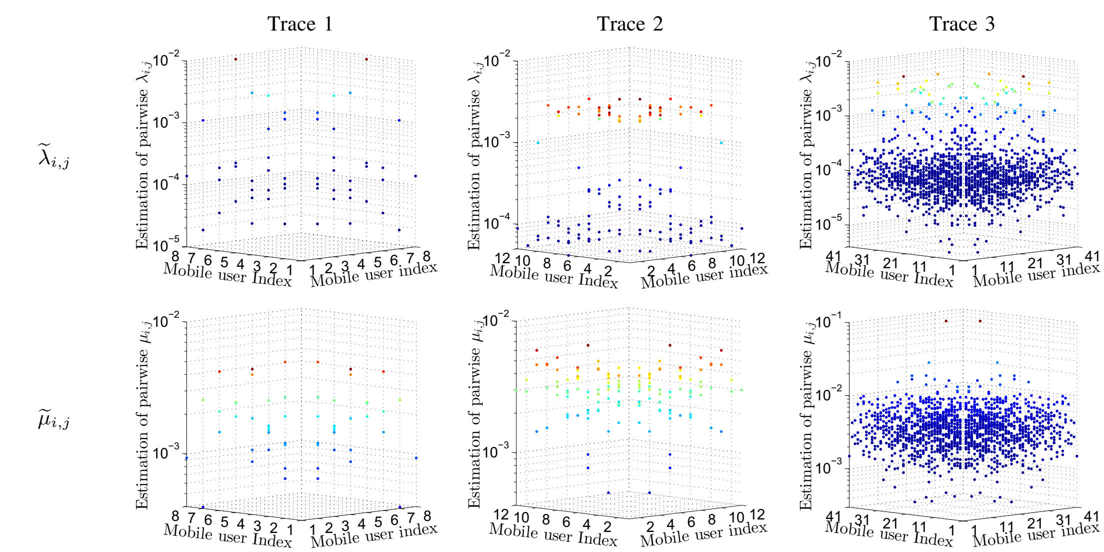


Fig. 3. The values of $\widetilde{\lambda}_{i,j}$ and $\widetilde{\mu}_{i,j}$ are obtained by MLE based on connectivity traces in [47]. The value of $\widetilde{\lambda}_{i,j}$ is almost with the order of magnitude of 10^{-4} and the value of $\widetilde{\mu}_{i,j}$ is mainly with the order of magnitude of 10^{-3}.

the scheme used to create data flows. We then present trace-driven simulation results to validate our system model. The performance of Algorithm 2 is also presented by running trace-driven simulations and comparing with the existing schemes in the literature.

4.1 Dataset in Traces and Data Flow Creation Scheme

We use three real-world traces "Intel" (Trace 1), "Cambridge" (Trace 2), and "Infocom" (Trace 3) in Cambridge/Haggle dataset in [47] for our simulations. Traces 1–3 are recorded by 8, 12, and 41 mobile iMotes using Bluetooth with $30\,\mathrm{m}$ radio range, respectively. Although these iMotes are not smartphones or tablets, the connection states recorded in these traces can be used to reproduce the dynamic topology of mobile users. The interval of each iMote sending a beacon (*i.e.*, hello message) is 120 ± 12 sec.

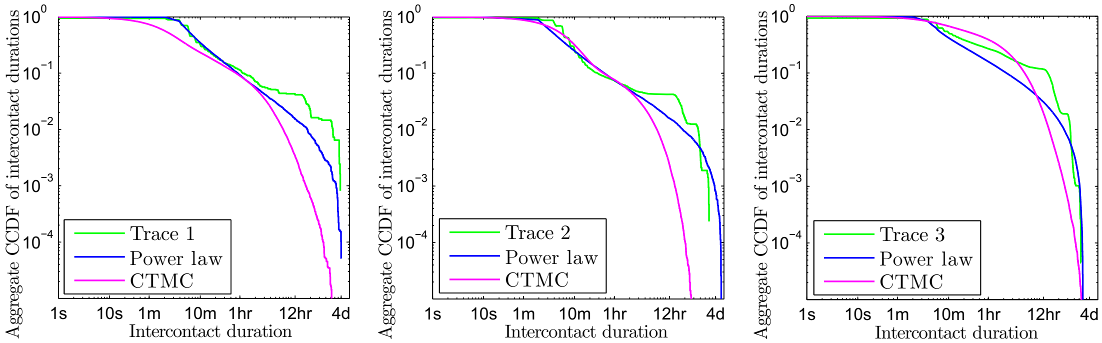
The connectivity between mobile users is assumed to be symmetric in our work. However, the connect and disconnect events in traces were recorded by each iMote individually. Thus, we consider that a pair of iMotes were connected (or disconnected) as long as one of them detected a connect (or disconnect) event. In the real-world traces, an iMote has recorded a connect event with a zero contact duration when it was connected with another iMote for a short period of time such that the iMote failed to receive two or more consecutive beacons. Thus, for a record with the zero contact duration, we assume the actual contact duration is uniformly distributed on $[0\,\mathrm{sec}, 120\,\mathrm{sec}]$. We concatenate the contact and intercontact durations recorded by each pair of iMotes in a chronological order to reproduce the connect and disconnect events for both of them. We then run trace-driven simulations with the D2D topologies reproduced by all iMotes pairs in each trace. Since all traces are recorded indoors, our simulation results are for indoor environments only. However, our proposed algorithm can also be used by outdoor users. The radio range of mobile iMotes is $30\,\mathrm{m}$. In our trace-drive simulations, the length of each time slot is $1\,\mathrm{sec}$. Moreover, ξ is set to 2% in the simulations.

The data flows are created by the following scheme. We have a set of digital objects $\mathcal{O}$. In our simulations, each user $i \in \mathcal{U}$ randomly chooses 50 objects from set $\mathcal{O}$ as his interested objects, *i.e.*, $|\mathcal{O}_i| = 50$. For each object $k \in \bigcup_{i\in\mathcal{U}} \mathcal{O}_i$, the value of x_k (sec) is uniformly distributed on $[0, 3600]$. That is, the diffusion start time x_k of object k is randomly selected within the first $1\,\mathrm{hr}$ (*i.e.*, 3600 sec) in a simulation run. We first set $y_{i,k} = x_k$ for a random user i in set $\{u \,|\, k \in \mathcal{O}_u\}$. Then, the time when user $j \in \{u \,|\, k \in \mathcal{O}_u\} \backslash \{i\}$ reveals his interest in object k (*i.e.*, $y_{j,k}$) is set to x_k plus a random number generated by log-normal distribution $\ln \mathcal{N}(\mu = 3.91, \sigma^2 = 6.86)$ [44], unless otherwise stated.

4.2 System Model Validation

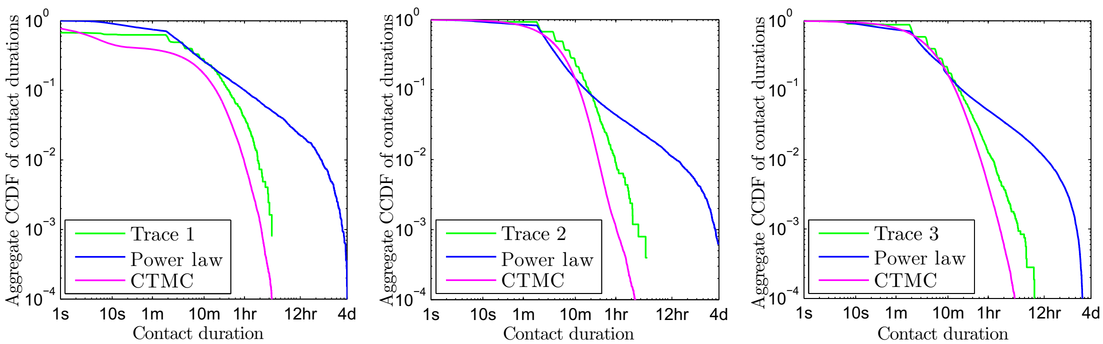
4.2.1 Validation for the Pairwise Connectivity Model

Let $\mathcal{U}_1$, $\mathcal{U}_2$, and $\mathcal{U}_3$ denote the sets of users in Traces 1–3, respectively. We have $|\mathcal{U}_1| = 8$, $|\mathcal{U}_2| = 12$, and $|\mathcal{U}_3| = 41$. For each record in Traces 1–3 with the zero contact duration, its actual contact duration is randomly selected from $[0\,\mathrm{sec}, 120\,\mathrm{sec}]$. The values of $\widetilde{\lambda}_{i,j}$ and $\widetilde{\mu}_{i,j}$ estimated for each pair of users $i, j \in \mathcal{U}$ by MLE are presented in Fig. 3. We find that the value of $\widetilde{\lambda}_{i,j}$ is almost with the order of magnitude of 10^{-4} and the value of $\widetilde{\mu}_{i,j}$ is mainly with the order of magnitude of 10^{-3}. We further predict the D2D connectivity for each pair of mobile users with both power law and CTMC models. Specifically, we simulate the pairwise connectivity with CTMC model by using parameters $\widetilde{\lambda}_{i,j}$ and $\widetilde{\mu}_{i,j}$ for each pair of users i and j in a trace. For the power law model, we consider that both pairwise contact and intercontact durations for a given user pair follow the power law distributions and estimate the parameters for both distributions, respectively. We then simulate the pairwise connectivity of each pair of users by the obtained power law distributions. We gather the simulation results of all user pairs in each trace for both connectivity models. We compare the aggregate CCDFs of contact and intercontact durations given by simulations



(a) Trace 1 (b) Trace 2 (c) Trace 3

Fig. 4. The aggregate CCDFs of the intercontact durations of all user pairs when power law and CTMC models are used to predict the connectivity between each user pair.



(a) Trace 1 (b) Trace 2 (c) Trace 3

Fig. 5. The aggregate CCDFs of the contact durations of all user pairs when power law and CTMC models are used to predict the connectivity between each user pair.

with the corresponding CCDFs given by empirical results in real-world traces, respectively. The comparison is shown in Figs. 4 and 5 for the aggregate CCDFs of intercontact and contact durations. From the aggregate CCDFs in Fig. 4, we find that the power law model can better predict the pairwise intercontact duration than the CTMC model for the connectivity between users in Trace 1. However, the CTMC model achieves similar performance as the power law model to predict the pairwise intercontact duration for users in Traces 2 and 3. On the other hand, it is observed from Fig. 5 that the CTMC and power law models can also obtain the similar performance in predicting the pairwise contact durations in all traces.

To better show the accuracy of the CTMC and power law models in predicting the pairwise contact and intercontact durations, we conduct the Kolmogorov-Smirnov test to show the maximum gap from the aggregate CCDFs of contact and intercontact durations given by simulation results to the corresponding aggregate CCDFs given by empirical data sets, respectively. The results of Kolmogorov-Smirnov test are given in Fig. 6. We observe that the power law model is more accurate than the CTMC model to predict the pairwise intercontact time in Traces 1 and 3. However, the adopted CTMC model obtains a better prediction for both pairwise contact and intercontact durations in Trace 2. Meanwhile, the CTMC model also achieves a better performance to predict the pairwise contact time in Trace 1.

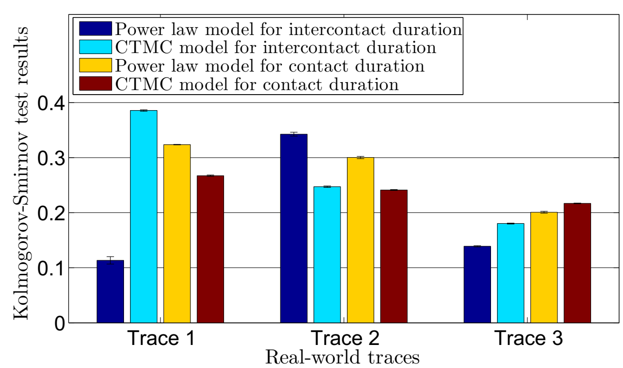


Fig. 6. Kolmogorov-Smirnov test results obtained by comparing aggregate CCDFs of contact and intercontact durations given by simulations with power law and CTMC models with aggregate CCDFs given by empirical results.

4.2.2 *Effect of Bayesian Inference*

We now present the benefit of using Bayesian inference in interest estimation. Specifically, we take the estimate $p^t_{i,j,k}$ made by user i at time t for the interest of user j in object k as the baseline case. Note that $p^t_{i,j,k}$ defined in Section 2.2.2 considers only the social influence aspect. We compare the estimate $p^t_{i,j,k}$ with the estimate $q^t_{i,j,k}$ defined in Section 2.2.3 by noting that $q^t_{i,j,k}$ takes Bayesian inference into account to estimate the interest of user j in object k. For our simulations, we apply the scheme introduced in

Section 4.1 to create the data flows with $|\mathcal{O}| = 400$ digital objects. We let all users reveal their interest within the first 1 hr (*i.e.*, $y_{i,k} \leq 3600\,\text{sec}, \forall\, i \in \mathcal{U}, k \in \mathcal{O}_i$). We compare the aforementioned two estimates in terms of the *estimation error per object* by running trace-driven simulations in Traces 1–3. Let $E_{\text{bse}}(t)$ denote the estimation error per object of the baseline case at $t > 3600$ sec, which is defined as follows

$$E_{\text{bse}}\left(t\right) \triangleq \frac{\sum_{i\in\mathcal{U}}\sum_{k\in\mathcal{O}_i}\sum_{j\in\mathcal{U}\setminus\{i\}}\left|p_{i,j,k}^t - I_{j,k}\right|}{\left(|\mathcal{U}|-1\right)\sum_{i\in\mathcal{U}}|\mathcal{O}_i|}, \qquad (13)$$

where $I_{j,k} = 1$ if user j is interested in object k according to the data flows created in simulations, and is equal to zero otherwise. The estimation error per object of our proposed model at time $t > 3600$ sec is denoted by $E_{\text{Bay}}(t)$, which is determined by replacing $p_{i,j,k}^t$ in (13) by $q_{i,j,k}^t$. In Fig. 7, we compare $E_{\text{bse}}\left(t\right)$ and $E_{\text{Bay}}\left(t\right)$ at various sample time t. We find that $E_{\text{Bay}}\left(t\right)$ is smaller than $E_{\text{bse}}\left(t\right)$ at all sample time t in each trace-driven simulation, which means that the accuracy of the interest estimation model can be improved by Bayesian inference. Besides, it is observed from Fig. 7 that using only social influence model to estimate the interest of users may not always provide a better result at a longer simulation time t (*e.g.*, $E_{\text{bse}}(t)$ of Trace 3 in Fig. 7). This is because a social influence model cannot be perfect. Specifically, the information exchanged between users during contacts can let them be certain of the interest of each other on the objects that they are waiting to download via D2D data offloading. However, this may also introduce the interest estimation error, since users may overestimate the interest of other users for the popular objects. With the same reason, we have also observed in Fig. 7 that even though the estimation error per object may decrease when the simulation evolves over time, it does not decrease significantly (*e.g.*, $E_{\text{bse}}(t)$ of Traces 1 and 2 in Fig. 7). On the other hand, the estimation error per object with Bayesian inference can be significantly reduced when the simulation evolves over time (*i.e.*, $E_{\text{Bay}}(t)$ of Traces 1–3 in Fig. 7). We would like to show that the proposed interest estimation model with Bayesian inference still works when the digital objects are diffused with a different pattern from the model that is used for the proposed interest estimation. To this end, we create data flows with another information diffusion model given by a log-normal distribution $\ln\mathcal{N}(\mu = 5.547, \sigma^2 = 4.519)$ [45]. From Fig. 8, we observe that similar simulation results as shown in Fig. 7 can be obtained. Thus, the Bayesian inference can increase the accuracy of interest estimation when digital objects have different log-normal diffusion patterns.

4.2.3 Validation of Interest Estimation

We now show that a larger value of $q_{i,j,k}^t$ obtained by user i at time t can better convince him that user j is interested in object k. For such a purpose, we introduce the *hit ratio* for our interest estimation model. Let $g(t,h)$ denote the hit ratio at time t with probability threshold h, where $0 \leq h \leq 1$. We define

$$g\left(t,h\right) \triangleq \frac{\sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{U}\setminus\{i\}}\left|\left\{k \,|\, k\in\mathcal{O}_i^t, q_{i,j,k}^t \geq h\right\}\cap\mathcal{O}_j\right|}{\sum_{i\in\mathcal{U}}\sum_{j\in\mathcal{U}\setminus\{i\}}\left|\left\{k \,|\, k\in\mathcal{O}_i^t, q_{i,j,k}^t \geq h\right\}\right|}. \qquad (14)$$

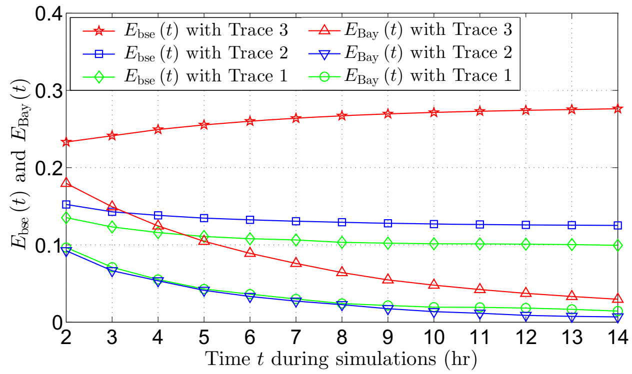


Fig. 7. The estimation error per object obtained by applying Bayesian inference in interest estimation with the information diffusion model $\ln\mathcal{N}(\mu = 3.91, \sigma^2 = 6.86)$ [44].

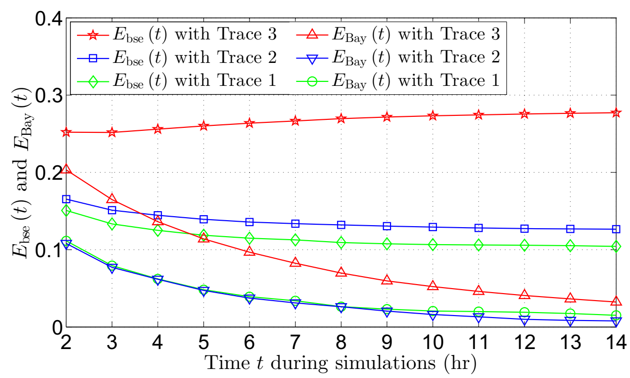


Fig. 8. The estimation error per object obtained by applying Bayesian inference in interest estimation with the information diffusion model $\ln\mathcal{N}(\mu = 5.547, \sigma^2 = 4.519)$ in [45].

Specifically, $\left|\left\{k \,|\, k\in\mathcal{O}_i^t, q_{i,j,k}^t \geq h\right\}\right|$ in the denominator of (14) is the number of objects that (a) user i is interested in at time t and (b) user j is also interested in with probability h or higher. Each element of the summation in the nominator of (14) is the cardinality of a subset of the aforementioned objects that user j is indeed interested in. We create data flows with $|\mathcal{O}| = 200$ to run simulations for each trace introduced above. At time $t = 8\times3600$ sec (*i.e.*, 8 hr) in each simulation run, we evaluate (14) for $h = 0, 0.02, 0.04, \ldots, 1$. The denominator in (14) may decrease to 0 as h increases. When this happens, we stop increasing h and plot the results that have been obtained in Fig. 9. The positive relation between h and $q\left(t,h\right)$ means that a larger value of $q_{i,j,k}^t$ can increase the confidence of user i that user j is interested in object k. This shows the correctness of our interest estimation model.

4.3 Performance of Proposed Algorithm

We now present the performance of Algorithm 2. We use 400 digital objects in our trace-driven simulations (*i.e.*, $|\mathcal{O}| = 400$). We first show that using our EAD metric with the proposed model can help a user choose the object that has the shortest available duration for D2D data offloading. Specifically, when both of the data flow (given by our data flow creation scheme in Section 4.1) and the D2D topologies (given by trace files) are known *a priori*, the available duration that a mobile user can download an object from his

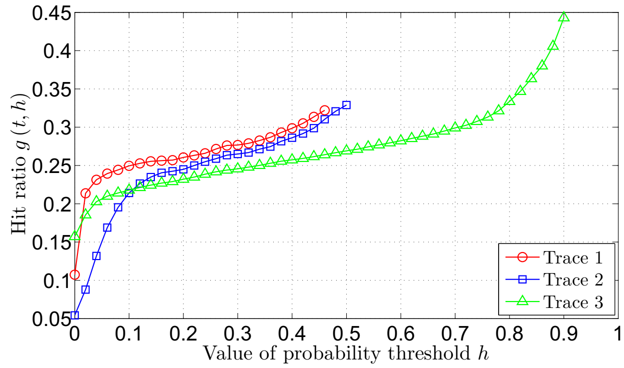


Fig. 9. Hit ratio $g(t,h)$ at $t = 8\times 3600$ sec (*i.e.*, 8 hr) vs. the threshold value h. The positive relation between $g(t,h)$ and h shows the correctness of our interest estimation model.

neighborhood within the preferred MWT can be accurately determined. In this ideal case with the *priori knowledge*, we can find the object that the user should download from his neighborhood in each time slot. The ideal case is not applicable in practice. However, it can be taken as the benchmark to show the accuracy of Algorithm 2 in choosing object $k^\star$ (Line 23). We compare our work with two existing scheduling policies, namely, the *earliest deadline first* (EDF) policy and the *shortest remaining processing time first* (SRPTF) policy. The EDF policy has been used to improve the quality of service in wired networks [48]. The SRPTF policy has been adopted for server-side task scheduling in peer-to-peer systems [49]. We compare our work with EDF and SRPTF policies because both of them are also used for task scheduling in D2D data offloading [50]. To simulate the EDF policy, in each time slot of our simulation run, we let each mobile user download the object (from his neighbors) that has the shortest remaining time before the end of its MWT preferred by the user. To simulate the SRPTF policy, we refer to the study in [51] and consider that the D2D communication data rate at time t_s is uniformly distributed between $[1\,\text{Mbps}, 4\,\text{Mbps}]$. Let $\gamma_{i,j}^{t_s}$ denote the D2D communication data rate between users $i, j \in \mathcal{U}$ at time t_s. For simplicity, we assume that $\gamma_{i,j}^{t_s}$ remains unchanged during time slot s. In particular, we have $\gamma_{i,j}^{t_s} > 0$ if users i and j are neighbors (*i.e.*, $C_{i,j}^{t_s} = 1$). Otherwise, we have $\gamma_{i,j}^{t_s} = 0$. Let $F_{i,k}^{t_s}$ denote the residual file size of object k that has not been completely downloaded by user i at time t_s. We run simulations for SRPTF policy by letting user i first download object k from neighbor j, where the tuple (j,k) is given by the solution of the following problem:

$$\underset{(j,k)\in(\mathcal{U}\backslash\{i\})\times\mathcal{Q}_i^{t_s}}{\text{argmin}} \quad \frac{F_{i,k}^{t_s}}{\gamma_{i,j}^{t_s}}, \ \text{ subject to } \ C_{i,j}^{t_s} = 1, \ \xi_{j,k}^{t_s} \geq \xi, \tag{15}$$

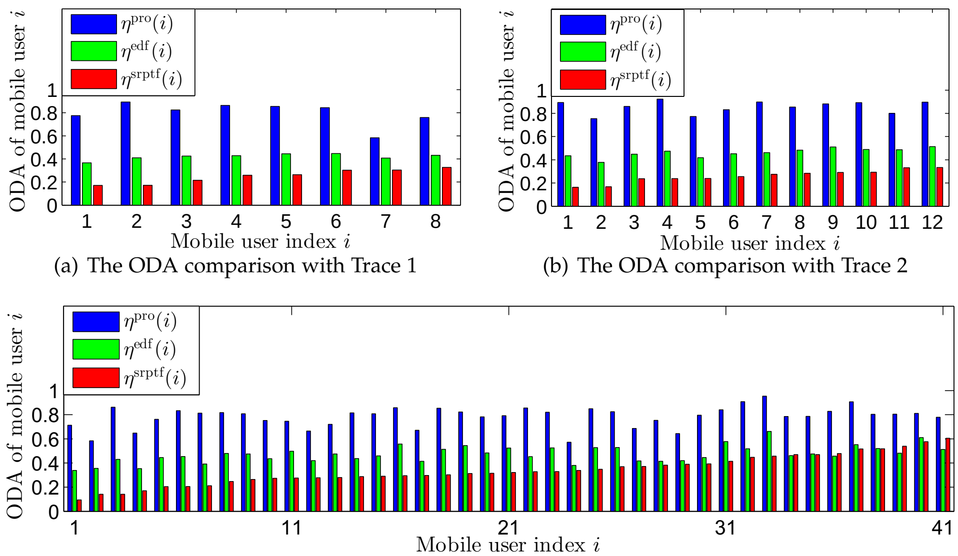
where the objective function is based on the definition of SRPTF, and the constraints restrict that only the neighbors who have already downloaded ξ or a higher percentage of object k can transmit data to user i.

We refer to the object selection made by a user to download in a time slot as an *offloading decision*. For user i in time slot s, let $d_{i,s}^{\text{ideal}}$, $d_{i,s}^{\text{pro}}$, $d_{i,s}^{\text{edf}}$, and $d_{i,s}^{\text{srptf}}$ denote the offloading decisions made in the ideal case, in Algorithm 2, with the EDF policy, and with the SRPTF policy, respectively. User i may not always find an interested object available on his neighbors. We set $d_{i,s}^{\text{ideal}} = d_{i,s}^{\text{pro}} = d_{i,s}^{\text{edf}} = d_{i,s}^{\text{srptf}} = 0$ if no object can be transferred from neighbors to user i in time slot s. After running a simulation with $T = 80000$ time slots, we evaluate the *offloading decision accuracy* (ODA) for user $i \in \mathcal{U}$ in our proposed Algorithm 2, which is defined as

$$\eta^{\text{pro}}(i) \triangleq \frac{\sum_{s=1}^{T} \mathbf{1}_{\{d_{i,s}^{\text{ideal}}\}\backslash\{0\}}(d_{i,s}^{\text{pro}})}{\sum_{s=1}^{T} \mathbf{1}_{\mathcal{O}}(d_{i,s}^{\text{ideal}})}, \tag{16}$$

where $\mathbf{1}_{\Omega}(\omega)$ is the indicator function which returns the value $\mathbf{1}_{\Omega}(\omega) = 1$ if $\omega \in \Omega$, and $\mathbf{1}_{\Omega}(\omega) = 0$ otherwise. We denote $\eta^{\text{edf}}(i)$ and $\eta^{\text{srptf}}(i)$ as the ODA when user i downloads the interested object by using the EDF and SRPTF policies, respectively. Specifically, $\eta^{\text{edf}}(i)$ and $\eta^{\text{srptf}}(i)$ are determined by replacing $d_{i,s}^{\text{pro}}$ in (16) by $d_{i,s}^{\text{edf}}$ and $d_{i,s}^{\text{srptf}}$, respectively. The values of ODA at user i (*i.e.*, $\eta^{\text{pro}}(i)$, $\eta^{\text{edf}}(i)$, and $\eta^{\text{srptf}}(i)$) are the proportions that the objects selected by user i in each time slot are the same objects selected by the ideal case in corresponding time slots. In this set of simulations, the MWT for each object is set to be uniformly distributed on $[1\,\text{hr}, 7\,\text{hr}]$. The values of $\eta^{\text{pro}}(i)$, $\eta^{\text{edf}}(i)$, and $\eta^{\text{srptf}}(i)$ for each user i in each real-world trace are given in Fig. 10. In this figure, we have sorted the users in ascending order according to their ODA when the SRPTF policy is used (*i.e.*, $\eta^{\text{srptf}}(i), \forall\, i \in \mathcal{U}$). Results in Fig. 10 show that the ODA of each user obtained by Algorithm 2 is almost between 0.7 and 0.9. However, the ODA of each user by using EDF and SRPTF is almost lower than 0.6. Thus, using the EAD metric can effectively determine the object that should be first downloaded from neighbors.

We now show that users can download more data from their neighbors by using Algorithm 2 compared with downloading an available object by the EDF or SRPTF policy. We consider that the size of each digital object is 100 MB. Since we have $|\mathcal{O}_i| = 50, \forall\, i \in \mathcal{U}$, the average mobile data traffic demand on each user is $100\,\text{MB} \times 50 = 5\,\text{GB}$. Note that $\gamma_{i,j}^{t_s}$ has been introduced above for the SRPTF policy to denote the D2D communication data rate between users i and j. Moreover, $\gamma_{i,j}^{t_s}$ has been assumed to remain unchanged in time slot s. To compare our proposed algorithm with the SRPTF and EDF policies in a fair manner, when we run simulations for the proposed algorithm and the EDF policy, the D2D communication data rate between users i and j in time slot s is also set to $\gamma_{i,j}^{t_s}$. We vary the average MWT for digital objects from $0.5\,\text{hr}$ to $3.5\,\text{hr}$ with a step size of $0.5\,\text{hr}$ in simulations. To do so, the MWT preferred by each user for each object that the user is interested in is uniformly distributed on $[0\,\text{hr}, T\,\text{hr}]$, where $T = 1, \ldots, 7$ for different simulation trials. We obtain the performance for four kinds of offloading decisions introduced above. Specifically, we refer to the ideal case as Case 1, where the offloading decision $d_{i,s}^{\text{ideal}}$ is used by user $i \in \mathcal{U}$ in time slot s. Cases 2–4 refer to the situations that the offloading decisions $d_{i,s}^{\text{pro}}$, $d_{i,s}^{\text{edf}}$, and $d_{i,s}^{\text{srptf}}$ are made by user i in time slot s, respectively. We evaluate the average size of data that can be downloaded from neighbors for these cases. For Cases 2–4, we denote $R_{2,1}$, $R_{3,1}$, and $R_{4,1}$ as the *relative performances*, which are defined as the ratios between the average data traffic offloaded in Cases 2–4 and the average data traffic



(a) The ODA comparison with Trace 1

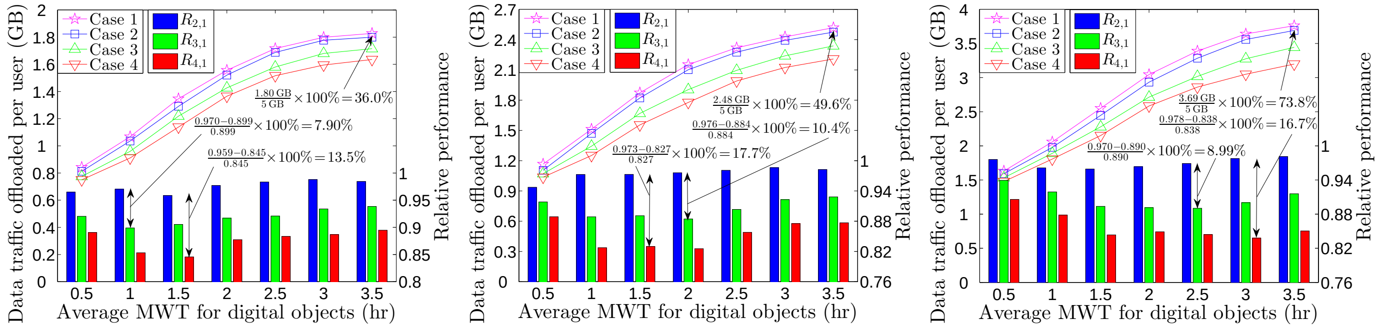
(b) The ODA comparison with Trace 2

(c) The ODA comparison with Trace 3

Fig. 10. Comparing Algorithm 2 with the EDF and SRPTF policies in terms of ODA.

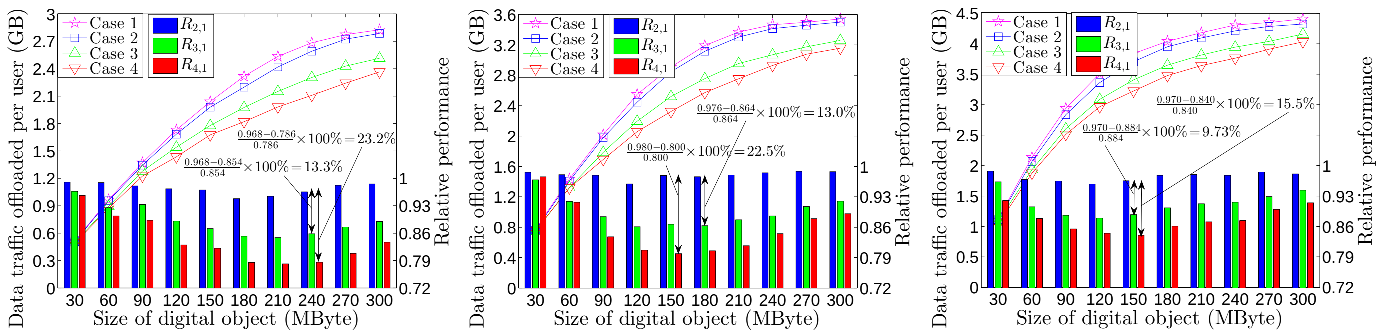
offloaded in Case 1, respectively. We observe from Fig. 11 that $36.0\% - 73.8\%$ of the $5\,\text{GB}$ data can be downloaded from neighbors when the users prefer to wait for $4\,\text{hr}$ on average for their interested objects in D2D data offloading. Fig. 11 also shows that the increment of the offloaded data traffic gradually decreases when the average MWT increases. This indicates that the size of data downloaded by a user from his neighbors cannot increase linearly with the average MWT for digital objects, as some of his interested objects may no longer be of interest to others. From the bar charts in Fig. 11, we find that Case 2 can achieve a performance close to Case 1 and download more data from neighbors compared with Cases 3 and 4. This is because first downloading the object that has the smallest EAD can better utilize the transient D2D data offloading opportunity which may not occur again. First downloading the object with the EDF policy (*i.e.*, Case 3) also achieves a higher performance than that with the SRPTF policy (*i.e.*, Case 4). With our simulation settings, up to $13.5\% - 17.7\%$ more data can be downloaded by the proposed algorithm (*i.e.*, Case 2) when compared with Case 4. Meanwhile, $7.9\% - 10.4\%$ more data can be downloaded when we compare Case 2 with Case 3. It is worth mentioning that our proposed algorithm as well as the EDF and SRPTF polices all benefit from the assumption that users who have partially downloaded the same object have non-overlapped portions to transfer to each other. However, the SRPTF policy can benefit the most from the above assumption. To explain this, let us consider object $k \in \mathcal{O}$ which has almost been downloaded by user $i \in \mathcal{U}$ (*i.e.*, only a few packets are missing). Since the residual file size of object k on user i is small, object k has a short remaining processing time on user i. That is, using SRPTF policy may prioritize the D2D data offloading for the object which has the smallest residual file size (*i.e.*, the least missing packets). However, it has been shown in [52] that a user who has received more packets for an object has smaller probability to find the missing packets of that object on his neighbors. Since we have assumed that the users who have partially downloaded the same object always have non-overlapped portions to transfer to each other, the SRPTF policy which chooses an object that has the smallest probability for a user to obtain missing packets in the neighborhood will benefit the most from the aforementioned assumption.

We study the performance of Algorithm 2 where the size of each object varies from $30\,\text{MB}$ to $300\,\text{MB}$ and the MWT of each object is uniformly distributed on $[0\,\text{hr}, 5\,\text{hr}]$. The simulation results are presented in Fig. 12. We find that the size of data downloaded via D2D data offloading first increases linearly when the size of each objects increases from $30\,\text{MB}$ to $120\,\text{MB}$. Then, the increment decreases when the size of each object keeps increasing. The amount of data traffic obtained by D2D data offloading eventually starts to saturate when the size of each object reaches $300\,\text{MB}$. From the bar charts in Fig. 12, we find that Algorithm 2 in Case 2 obtains almost the same performance as the ideal Case 1 when the size of each object is either smaller than $90\,\text{MB}$ or greater than $270\,\text{MB}$. Moreover, we find that the relative performance of Algorithm 2 comparing with EDF and SRPTF (*i.e.*, Cases 3 and 4, respectively) increases first and then decreases. In particular, when the size of each object is small, an object can be completely downloaded once it appears in the neighborhood. In this case, offloading decisions slightly affect the amount of data traffic obtained by D2D data offloading. When the size of each object increases, the proposed offloading algorithm also achieves a better relative performance. The relative performance compared with the baseline policies starts to decrease if the size of each object keeps increasing. This is because large-size objects can provide users the persistent D2D data offloading opportunities before they have been completely downloaded. We observe from Fig. 12 that using Algorithm 2 can help users download up to $15.5\% - 23.2\%$ more data from neighbors than using the SRPTF policy. Using Algorithm 2 can also help users offload up to $9.73\% - 13.3\%$ more data from the cellular network than using the



(a) Data traffic offloaded in Trace 1 (b) Data traffic offloaded in Trace 2 (c) Data traffic offloaded in Trace 3

Fig. 11. Mobile data traffic offloaded per user and relative performance vs. the average MWT (size of each object is 100 MB).



(a) Data traffic offloaded in Trace 1 (b) Data traffic offloaded in Trace 2 (c) Data traffic offloaded in Trace 3

Fig. 12. Mobile data traffic offloaded per user and relative performance vs. the size of each object (average MWT is 2.5 hr).

EDF policy.

5 Conclusion

In this paper, we have proposed the EAD metric to evaluate the D2D data offloading opportunity. The EAD metric takes into account the pairwise connectivity between mobile devices, the social influence between users, diffusion of digital objects, and the MWT preferred by users for their interested objects. We have extended the idea of the rarest first strategy to wireless domain for D2D data offloading by letting a mobile device first download an available object with the smallest EAD. An efficient D2D data offloading algorithm was also proposed, which can be applied by mobile devices in a distributed manner. The correctness of the pairwise connectivity model and the interest estimation model were validated by trace-driven simulations. Simulation results showed that our D2D data offloading algorithm can effectively find the object that should be downloaded from the neighboring devices in each time slot. Comparing with downloading an available object with the EDF and SRPTF policies, trace-driven simulation results showed that more data can be downloaded via D2D data offloading by our work. For mobile users having similar activity (*e.g.*, colleagues working together or classmates taking the same course), the pairwise connectivity may be known in advance. Thus, the value of EAD can be determined more accurately. Moreover, D2D communication data rate may vary with the number of nearby users due to limited resources in wireless channels. For future work, we will consider users' activity and D2D communication link quality to further improve the performance of D2D data offloading.

Appendix A
Sketch of Proof of Lemma 1

We have $p^t_{i,j,k} = \mathbb{P}(I_{j,k} = 1 \,|\, \boldsymbol{r}^t_{i,k})$ by definition. We now evaluate the conditional probability $\mathbb{P}(I_{j,k} = 1 \,|\, \boldsymbol{r}^t_{i,k})$ obtained by user i at time t. Given the interest record $\boldsymbol{r}^t_{i,k} = (R^t_{i,1,k}, \ldots, R^t_{i,U,k})$ for object $k \in \mathcal{Q}^t_i$ on user $i \in \mathcal{U}$, user i is certain of a set of users that may influence user $j \in \mathcal{B}^t_{i,k}$ to be interested in object k. This set is $\{u \,|\, R^t_{i,u,k} = 1\}$, which is equivalent to $\mathcal{A}^t_{i,k} \cup \{i\}$. With our adopted social influence model for the distributed interest estimation, given the interest record $\boldsymbol{r}^t_{i,k}$ on user i at time t, the conditional probability determined by user i that user j is not influenced by any user in set $\mathcal{A}^t_{i,k} \cup \{i\}$ is $\prod_{u \in \mathcal{A}^t_{i,k} \cup \{i\}} (1 - \tilde{\theta}_{u,j})$. Thus, we have $\mathbb{P}(I_{j,k} = 1 \,|\, \boldsymbol{r}^t_{i,k}) = 1 - \prod_{u \in \mathcal{A}^t_{i,k} \cup \{i\}} (1 - \tilde{\theta}_{u,j})$. ■

Appendix B
Proof of Theorem 1

Since we have $q^t_{i,j,k} = \mathbb{P}(I_{j,k} = 1 \,|\, \boldsymbol{r}^t_{i,k}, \tau^{t,c}_{i,j})$, we need to evaluate the conditional probability $\mathbb{P}(I_{j,k} = 1 \,|\, \boldsymbol{r}^t_{i,k}, \tau^{t,c}_{i,j})$ obtained by user i at time t when $\boldsymbol{r}^t_{i,k}$ and $\tau^{t,c}_{i,j}$ are given. We take both the social influence and Bayesian inference into account to calculate the conditional probability. Note that x_k is the diffusion start time of object k. We first consider the case when $\tau^{t,c}_{i,j} \leq x_k$, which means that user i has not yet connected with user j after object k starts to be diffused among users. Therefore, we cannot use Bayesian inference in this case and only the social influence model is applicable. Thus, for $\tau^{t,c}_{i,j} \leq x_k$, we have

$$\mathbb{P}(I_{j,k} = 1 \,|\, \boldsymbol{r}^t_{i,k}, \tau^{t,c}_{i,j}) = \mathbb{P}(I_{j,k} = 1 \,|\, \boldsymbol{r}^t_{i,k}) = p^t_{i,j,k}. \quad (17)$$

We now consider the case when $\tau_{i,j}^{t,c} > x_k$. For user $j \in \mathcal{B}_{i,k}^t$, we denote $\boldsymbol{r}_{i,-j,k}^t$ to represent the vector after removing the component $R_{i,j,k}^t = 0$ from the interest record $\boldsymbol{r}_{i,k}^t$. We have

$$\begin{aligned}
&\mathbb{P}(I_{j,k} = 1 \mid \boldsymbol{r}_{i,k}^t, \tau_{i,j}^{t,c}) \\
&= \mathbb{P}(I_{j,k} = 1 \mid R_{i,j,k}^t = 0, \boldsymbol{r}_{i,-j,k}^t, \tau_{i,j}^{t,c}) \\
&\overset{(a)}{=} \mathbb{P}(I_{j,k} = 1 \mid Y_{j,k} > \tau_{i,j}^{t,c}, \boldsymbol{r}_{i,-j,k}^t, \tau_{i,j}^{t,c}) \\
&= \frac{\mathbb{P}(I_{j,k} = 1, Y_{j,k} > \tau_{i,j}^{t,c} \mid \boldsymbol{r}_{i,-j,k}^t, \tau_{i,j}^{t,c})}{\mathbb{P}(Y_{j,k} > \tau_{i,j}^{t,c} \mid \boldsymbol{r}_{i,-j,k}^t, \tau_{i,j}^{t,c})} \\
&\overset{(b)}{=} \frac{\mathbb{P}(Y_{j,k} > \tau_{i,j}^{t,c}, I_{j,k} = 1 \mid \boldsymbol{r}_{i,-j,k}^t, \tau_{i,j}^{t,c})}{\sum_{m=0}^{1} \mathbb{P}(Y_{j,k} > \tau_{i,j}^{t,c}, I_{j,k} = m \mid \boldsymbol{r}_{i,-j,k}^t, \tau_{i,j}^{t,c})} \\
&= \frac{\mathbb{P}(Y_{j,k} > \tau_{i,j}^{t,c} \mid I_{j,k} = 1, \boldsymbol{r}_{i,-j,k}^t, \tau_{i,j}^{t,c}) \mathbb{P}(I_{j,k} = 1 \mid \boldsymbol{r}_{i,-j,k}^t, \tau_{i,j}^{t,c})}{\sum_{m=0}^{1} \mathbb{P}(Y_{j,k} > \tau_{i,j}^{t,c} \mid I_{j,k} = m, \boldsymbol{r}_{i,-j,k}^t, \tau_{i,j}^{t,c}) \mathbb{P}(I_{j,k} = m \mid \boldsymbol{r}_{i,-j,k}^t, \tau_{i,j}^{t,c})}.
\end{aligned} \tag{18}$$

Equality (a) follows because $R_{i,j,k}^t = 0$ implies that user j is not interested in object k before his most recent connection with user i (*i.e.*, $Y_{j,k} > \tau_{i,j}^{t,c}$). Equality (b) is obtained by using the law of total probability. It has been shown that the time when a user shares his interested information with his friends in the OSN follows the log-normal distribution $\ln \mathcal{N}(\mu = 3.91, \sigma^2 = 6.86)$ after the information is initially posted [44]. Let random variable $S_{i,k} > 0$ denote the duration from $Y_{i,k}$ to the time when user i downloads ξ percentage of object k. Recall that constant parameter ξ is the percentage of an object that a user has to obtain before sharing the object with others. Thus, given $I_{j,k} = 1$, the time when user j shares object $k \in \mathcal{O}_j$ (*i.e.*, $Y_{j,k} + S_{j,k}$ cf. Section II-B3) is independent from both $\boldsymbol{r}_{i,-j,k}^t$ and $\tau_{i,j}^{t,c}$, and follows the above log-normal distribution after diffusion start time x_k. The cumulative distribution function of $Y_{j,k} + S_{j,k}$ at time t' is

$$\begin{aligned}
&\mathbb{P}(Y_{j,k} + S_{j,k} \le t' \mid I_{j,k} = 1, \boldsymbol{r}_{i,-j,k}^t, \tau_{i,j}^{t,c}) \\
&= \mathbb{P}(Y_{j,k} + S_{j,k} \le t' \mid I_{j,k} = 1) = \frac{1 + f(t' - x_k)}{2},
\end{aligned} \tag{19}$$

where $f(\varphi) = \operatorname{erf}(\frac{\ln \varphi - 3.91}{\sqrt{13.72}})$ is obtained by applying the above log-normal distribution and $\operatorname{erf}(\cdot)$ is the error function. The probability $\mathbb{P}(Y_{j,k} > \tau_{i,j}^{t,c} \mid I_{j,k} = 1, \boldsymbol{r}_{i,-j,k}^t, \tau_{i,j}^{t,c})$ in (18) satisfies

$$\begin{aligned}
&\mathbb{P}(Y_{j,k} > \tau_{i,j}^{t,c} \mid I_{j,k} = 1, \boldsymbol{r}_{i,-j,k}^t, \tau_{i,j}^{t,c}) \\
&= 1 - \mathbb{P}(Y_{j,k} \le \tau_{i,j}^{t,c} \mid I_{j,k} = 1, \boldsymbol{r}_{i,-j,k}^t, \tau_{i,j}^{t,c}) \\
&\overset{(c)}{<} 1 - \mathbb{P}(Y_{j,k} + S_{j,k} \le \tau_{i,j}^{t,c} \mid I_{j,k} = 1, \boldsymbol{r}_{i,-j,k}^t, \tau_{i,j}^{t,c}) \\
&\overset{(d)}{=} \frac{1 - f(\tau_{i,j}^{t,c} - x_k)}{2}.
\end{aligned} \tag{20}$$

Inequality (c) is due to $S_{j,k} > 0$. Equality (d) is obtained by substituting $\tau_{i,j}^{t,c}$ for t' in (19).

We now consider the conditional probability $\mathbb{P}(I_{j,k} = 1 \mid \boldsymbol{r}_{i,-j,k}^t, \tau_{i,j}^{t,c})$ in (18). In the considered social influence model, a user is interested in an object if he is influenced by one of the users who is interested in the object. Thus, random variable $I_{j,k}$ is independent from $\tau_{i,j}^{t,c}$. Moreover, given vector $\boldsymbol{r}_{i,-j,k}^t$, user i is aware of the set of users that may influence user j on his interest in object k, which is $\{u \mid R_{i,u,k}^t = 1\} = \mathcal{A}_{i,k}^t \cup \{i\}$. Noting that $R_{i,j,k}^t = 0$, we thus have

$$\begin{aligned}
&\mathbb{P}(I_{j,k} = 1 \mid \boldsymbol{r}_{i,-j,k}^t, \tau_{i,j}^{t,c}) \\
&= \mathbb{P}(I_{j,k} = 1 \mid \boldsymbol{r}_{i,-j,k}^t) \\
&= \mathbb{P}(I_{j,k} = 1 \mid \boldsymbol{r}_{i,k}^t) \\
&\overset{(e)}{=} 1 - \textstyle\prod_{u \in \mathcal{A}_{i,k}^t \cup \{i\}} (1 - \tilde{\theta}_{u,j}).
\end{aligned} \tag{21}$$

Equality (e) holds by the proof of Lemma 1 given in Appendix A. Furthermore, we have $\mathbb{P}(Y_{j,k} > \tau_{i,j}^{t,c} \mid I_{j,k} = 0, \boldsymbol{r}_{i,-j,k}^t, \tau_{i,j}^{t,c}) = 1$ because $y_{j,k} = \infty$ when $I_{j,k} = 0$. By substituting (20) and (21) into (18) and considering $\mathbb{P}(I_{j,k} = 0 \mid \boldsymbol{r}_{i,-j,k}^t, \tau_{i,j}^{t,c}) = 1 - \mathbb{P}(I_{j,k} = 1 \mid \boldsymbol{r}_{i,-j,k}^t, \tau_{i,j}^{t,c}) = \prod_{u \in \mathcal{A}_{i,k}^t \cup \{i\}} (1 - \tilde{\theta}_{u,j})$, we have

$$\begin{aligned}
&\mathbb{P}(I_{j,k} = 1 \mid \boldsymbol{r}_{i,k}^t, \tau_{i,j}^{t,c}) \\
&< \frac{(1 - f(\tau_{i,j}^{t,c} - x_k))(1 - \prod_{u \in \mathcal{A}_{i,k}^t \cup \{i\}} (1 - \tilde{\theta}_{u,j}))}{(1 - f(\tau_{i,j}^{t,c} - x_k))(1 - \prod\limits_{u \in \mathcal{A}_{i,k}^t \cup \{i\}} (1 - \tilde{\theta}_{u,j})) + 2 \prod\limits_{u \in \mathcal{A}_{i,k}^t \cup \{i\}} (1 - \tilde{\theta}_{u,j})},
\end{aligned} \tag{22}$$

which completes the proof. ■

Appendix C
Proof of Theorem 2

We have $z_{i,j,k}^t(t') = \mathbb{P}(Z_{i,j,k}^t(t') = 1 \mid \boldsymbol{r}_{i,k}^t, \tau_{i,j}^{t,c}, C_{i,j}^t)$ by definition. Thus, we evaluate the conditional probability $\mathbb{P}(Z_{i,j,k}^t(t') = 1 \mid \boldsymbol{r}_{i,k}^t, \tau_{i,j}^{t,c}, C_{i,j}^t)$ determined by user i at time $t' \ge t$ when $\boldsymbol{r}_{i,k}^t$, $\tau_{i,j}^{t,c}$, and $C_{i,j}^t$ are given. Specifically, we have $Z_{i,j,k}^t(t') = 1$ if users i and j are connected at time t' (*i.e.*, $C_{i,j}^{t'} = 1$) and user j has obtained ξ or a higher percentage of object k at time t' (*i.e.*, $Y_{j,k} + S_{j,k} < t'$). We have assumed that the stochastic D2D connections between mobile users are independent from both the interest of users and the diffusion of digital objects. Thus, we have

$$\begin{aligned}
&\mathbb{P}(Z_{i,j,k}^t(t') = 1 \mid \boldsymbol{r}_{i,k}^t, \tau_{i,j}^{t,c}, C_{i,j}^t) \\
&= \mathbb{P}(Y_{j,k} + S_{j,k} < t', C_{i,j}^{t'} = 1 \mid \boldsymbol{r}_{i,k}^t, \tau_{i,j}^{t,c}, C_{i,j}^t) \\
&= \mathbb{P}(Y_{j,k} + S_{j,k} < t' \mid \boldsymbol{r}_{i,k}^t, \tau_{i,j}^{t,c}) \mathbb{P}(C_{i,j}^{t'} = 1 \mid C_{i,j}^t).
\end{aligned} \tag{23}$$

Moreover, we have

$$\begin{aligned}
&\mathbb{P}(Y_{j,k} + S_{j,k} < t' \mid \boldsymbol{r}_{i,k}^t, \tau_{i,j}^{t,c}) \\
&\overset{(a)}{=} \sum_{m=0}^{1} \mathbb{P}(Y_{j,k} + S_{j,k} < t', I_{j,k} = m \mid \boldsymbol{r}_{i,k}^t, \tau_{i,j}^{t,c}) \\
&\overset{(b)}{=} \mathbb{P}(Y_{j,k} + S_{j,k} < t', I_{j,k} = 1 \mid \boldsymbol{r}_{i,k}^t, \tau_{i,j}^{t,c}) \\
&= \mathbb{P}(Y_{j,k} + S_{j,k} < t' \mid I_{j,k} = 1, \boldsymbol{r}_{i,k}^t, \tau_{i,j}^{t,c}) \mathbb{P}(I_{j,k} = 1 \mid \boldsymbol{r}_{i,k}^t, \tau_{i,j}^{t,c}) \\
&\overset{(c)}{=} \frac{(1 + f(t' - x_k)) \mathbb{P}(I_{j,k} = 1 \mid \boldsymbol{r}_{i,k}^t, \tau_{i,j}^{t,c})}{2}.
\end{aligned} \tag{24}$$

Equality (a) follows the law of total probability. Equality (b) is due to $\mathbb{P}(Y_{j,k} + S_{j,k} < t', I_{j,k} = 0 \mid \boldsymbol{r}_{i,k}^t, \tau_{i,j}^{t,c}) = 0$ as $y_{j,k} = \infty$ for $I_{j,k} = 0$. Equality (c) holds due to (19).

For probability $\mathbb{P}(I_{j,k} = 1 \mid \boldsymbol{r}_{i,k}^t, \tau_{i,j}^{t,c})$ in (24), we have $\mathbb{P}(I_{j,k} = 1 \mid \boldsymbol{r}_{i,k}^t, \tau_{i,j}^{t,c}) = 1$ given $R_{i,j,k}^t = 1$ in vector $\boldsymbol{r}_{i,k}^t$ (*i.e.*, $j \in \mathcal{A}_{i,k}^t$). This is because $j \in \mathcal{A}_{i,k}^t$ means that user j has informed user i that he is interested in object k. On the other hand, we have $\mathbb{P}\big(I_{j,k} = 1 \,|\, \boldsymbol{r}^t_{i,k}, \tau^{t,c}_{i,j}\big) = q^t_{i,j,k}$ by definition given $R^t_{i,j,k}=0$ (*i.e.*, $j \in \mathcal{B}^t_{i,k}$). This completes the proof. ■

REFERENCES

[1] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2015-2020," Feb. 2016. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html

[2] M.-R. Ra, J. Paek, A. B. Sharma, R. Govindan, M. H. Krieger, and M. J. Neely, "Energy-delay tradeoffs in smartphone applications," in *Proc. of ACM MobiSys*, San Francisco, CA, Jun. 2010.

[3] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, "Mobile data offloading: How much can WiFi deliver?" *IEEE/ACM Trans. on Networking*, vol. 21, no. 2, pp. 536–550, Apr. 2013.

[4] A. Ding, B. Han, Y. Xiao, P. Hui, A. Srinivasan, M. Kojo, and S. Tarkoma, "Enabling energy-aware collaborative mobile data offloading for smartphones," in *Proc. of IEEE SECON*, New Orleans, LA, Jun. 2013.

[5] F. Mehmeti and T. Spyropoulos, "Is it worth to be patient? Analysis and optimization of delayed mobile data offloading," in *Proc. of IEEE INFOCOM*, Toronto, Canada, Apr. 2014.

[6] WiFi Alliance, "Wi-Fi peer-to-peer (P2P) technical specification v1.5," [Online]. Available: http://www.wi-fi.org/discover-wi-fi/specifications

[7] A. Beach, M. Gartrell, S. Akkala, J. Elston, J. Kelley, K. Nishimoto, B. Ray, S. Razgulin, K. Sundaresan, B. Surendar, M. Terada, and R. Han, "Whozthat? Evolving an ecosystem for context-aware mobile social networks," *IEEE Network*, vol. 22, no. 4, pp. 50–55, Jul. 2008.

[8] B. Han, P. Hui, V. Kumar, M. Marathe, J. Shao, and A. Srinivasan, "Mobile data offloading through opportunistic communications and social participation," *IEEE Trans. on Mobile Computing*, vol. 11, no. 5, pp. 821–834, May 2012.

[9] X. Wang, M. Chen, Z. Han, D. O. Wu, and T. T. Kwon, "TOSS: Traffic offloading by social network service-based opportunistic sharing in mobile social networks," in *Proc. of IEEE INFOCOM*, Toronto, Canada, Apr. 2014.

[10] K.-J. Lin, C.-W. Chen, and C.-F. Chou, "Preference-aware content dissemination in opportunistic mobile social networks," in *Proc. of IEEE INFOCOM*, Orlando, FL, Mar. 2012.

[11] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on opportunistic forwarding algorithms," *IEEE Trans. on Mobile Computing*, vol. 6, no. 6, pp. 606–620, Jun. 2007.

[12] A. Chaintreau, P. Hui, J. Cowcroft, C. Diot, R. Gass, and J. Scott, "Pocket switched networks: Real-world mobility and its consequences for opportunistic forwarding," Technical Report, UCAM-CL-TR-617, Feb. 2005.

[13] T. Karagiannis, J.-Y. Le Boudec, and M. Vojnović, "Power law and exponential decay of intercontact times between mobile devices," *IEEE Trans. on Mobile Computing*, vol. 9, no. 10, pp. 1377–1390, Oct. 2010.

[14] Y. Wang, B. Krishnamachari, and T. W. Valente, "Findings from an empirical study of fine-grained human social contacts," in *Proc. of Int'l Conf. on Wireless On-demand Network Systems and Services (WONS)*, Snowbird, UT, Feb. 2009.

[15] V. Conan, J. Leguay, and T. Friedman, "Characterizing pairwise inter-contact patterns in delay tolerant networks," in *Proc. of ACM Int'l Conf. on Autonomic Computing and Communication Systems*, Rome, Italy, Oct. 2007.

[16] H. Cai and D. Y. Eun, "Crossing over the bounded domain: From exponential to power-law intermeeting time in mobile ad hoc networks," *IEEE/ACM Trans. on Networking*, vol. 17, no. 5, pp. 1578–1591, Oct. 2009.

[17] W. Gao, Q. Li, B. Zhao, and G. Cao, "Multicasting in delay tolerant networks: A social network perspective," in *Proc. of ACM MobiHoc*, New Orleans, LA, May 2009.

[18] P. E. Greenwood and M. S. Nikulin, *A Guide to Chi-Squared Testing*. Wiley, 1996.

[19] B. Zhang, Y. Li, D. Jin, P. Hui, and Z. Han, "Social-aware peer discovery for D2D communications underlaying cellular networks," *IEEE Trans. on Wireless Communications*, vol. 14, no. 5, pp. 2426–2439, May 2015.

[20] X. Zhuo, W. Gao, G. Cao, and S. Hua, "An incentive framework for cellular traffic offloading," *IEEE Trans. on Mobile Computing*, vol. 13, no. 3, pp. 541–555, Mar. 2014.

[21] B. Cohen, "Incentives build robustness in BitTorrent," in *Proc. of the Workshop on Economics of Peer-to-Peer Systems*, Berkeley, CA, Jun. 2003.

[22] C.-H. Lee and D. Y. Eun, "On the forwarding performance under heterogeneous contact dynamics in mobile opportunistic networks," *IEEE Trans. on Mobile Computing*, vol. 12, no. 6, pp. 1107–1119, Jun. 2013.

[23] Y. Li, M. Qian, D. Jin, P. Hui, Z. Wang, and S. Chen, "Multiple mobile data offloading through disruption tolerant networks," *IEEE Trans. on Mobile Computing*, vol. 13, no. 7, pp. 1579–1596, Jul. 2014.

[24] P. Sermpezis and T. Spyropoulos, "Modelling and analysis of communication traffic heterogeneity in opportunistic networks," *IEEE Trans. on Mobile Computing*, vol. 14, no. 11, pp. 2316–2331, Nov. 2015.

[25] S. M. Ross, *Introduction to Probability Models, Eleventh Edition*. Academic Press, 2014.

[26] V. Sciancalepore, D. Giustiniano, A. Banchs, and A. Hossmann-Picu, "Offloading cellular traffic through opportunistic communications: Analysis and optimization," *IEEE J. Select. Areas Commun.*, vol. 34, no. 1, pp. 122–137, Jan. 2016.

[27] M. Zhao, Y. Li, and W. Wang, "Modeling and analytical study of link properties in multihop wireless networks," *IEEE Trans. on Communications*, vol. 60, no. 2, pp. 445–455, Feb. 2012.

[28] Y. Li and W. Wang, "Can mobile cloudlets support mobile applications?" in *Proc. of IEEE INFOCOM*, Toronto, Canada, Apr. 2014.

[29] A. S. Das, M. Datar, A. Garg, and S. Rajaram, "Google news personalization: Scalable online collaborative filtering," in *Proc. of Int'l World Wide Web Conf.*, Banff, Canada, May 2007.

[30] L. Li, W. Chu, J. Langford, and R. E. Schapire, "A contextual-bandit approach to personalized news article recommendation," in *Proc. of Int'l World Wide Web Conf.*, Raleigh, NC, Apr. 2010.

[31] D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, and S. Suri, "Feedback effects between similarity and social influence in online communities," in *Proc. of ACM SIGKDD*, Las Vegas, NV, Aug. 2008.

[32] A. Goyal, F. Bonchi, and L. V. S. Lakshmanan, "Learning influence probabilities in social networks," in *Proc. of ACM Int'l Conf. on Web Search and Data Mining (WSDM)*, New York, NY, Feb. 2010.

[33] J. Goldenberg, B. Libai, and E. Muller, "Talk of the network: A complex systems look at the underlying process of word-of-mouth," *Marketing Letters*, vol. 12, no. 3, pp. 211–223, Aug. 2001.

[34] U. Shardanand and P. Maes, "Social information filtering: Algorithms for automating 'word of mouth'," in *Proc. of ACM SIGCHI Conf. on Human Factors in Computing Systems*, Denver, CO, May 1995.

[35] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *Journal of the American Society for Information Science and Technology*, vol. 58, no. 7, pp. 1019–1031, May 2007.

[36] B. Sigurbjörnsson and R. van Zwol, "Flickr tag recommendation based on collective knowledge," in *Proc. of Int'l World Wide Web Conf.*, Beijing, China, Apr. 2008.

[37] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[38] F. Qiu and J. Cho, "Automatic identification of user interest for personalized search," in *Proc. of Int'l World Wide Web Conf.*, Edinburgh, Scotland, May 2006.

[39] J. Davidson, B. Liebald, J. Liu, P. Nandy, and T. Van Vleet, "The YouTube video recommendation system," in *Proc. of ACM Conf. on Recommender Systems*, Barcelona, Spain, Sep. 2010.

[40] Z. Weinberg, E. Y. Chen, P. R. Jayaraman, and C. Jackson, "I still know what you visited last summer: Leaking browsing history via user interaction and side channel attacks," in *Proc. of IEEE Symposium on Security and Privacy*, Oakland, CA, May 2011.

[41] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *Proc. of ACM SIGKDD*, Washington, D.C., Aug. 2003.

[42] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins, "Information diffusion through blogspace," in *Proc. of Int'l World Wide Web Conf.*, New York, NY, May 2004.

[43] D. Kempe, J. Kleinberg, and E. Tardos, "Influential nodes in a diffusion model for social networks," in *Proc. of Int'l Conf. on Automata, Languages and Programming (ICALP)*, Lisbon, Portugal, Jul. 2005.

[44] W. Galuba, K. Aberer, D. Chakraborty, Z. Despotovic, and W. Kellerer, "Outtweeting the Twitterers — Predicting information

cascades in microblogs," in *Proc. of ACM Workshop on Online Social Networks*, Boston, MA, Jun. 2010.

[45] J. L. Iribarren and E. Moro, "Impact of human activity patterns on the dynamics of information diffusion," *Physical Review Letters*, vol. 103, no. 3, p. 038702, Jul. 2009.

[46] A. Pyattaev, O. Galinina, S. Andreev, M. Katz, and Y. Koucheryavy, "Understanding practical limitations of network coding for assisted proximate communication," *IEEE J. Select. Areas Commun.*, vol. 33, no. 2, pp. 156–170, Feb. 2015.

[47] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, and A. Chaintreau, "CRAWDAD dataset cambridge/haggle (v. 2006-09-15)," Sep. 2006. [Online]. Available: http://crawdad.org/cambridge/haggle/20060915/imote

[48] F. M. Chiussi and V. Sivaraman, "Achieving high utilization in guaranteed services networks using early-deadline-first scheduling," in *Proc. of Int'l Workshop on Quality of Service*, Napa, CA, May 1998.

[49] Y. Qiao, D. Lu, F. E. Bustamante, and P. A. Dinda, "Looking at the server side of peer-to-peer systems," in *Proc. of Workshop on Languages, Compilers, and Run-time Support for Scalable Systems*, Houston, TX, Oct. 2004.

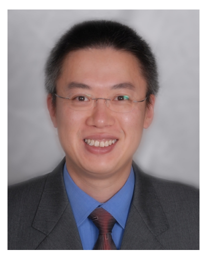
[50] Z. Lu and G. De Veciana, "Application-aware opportunistic D2D link schedulers: Traffic offloading and user perceived QoS," Technical Report, Department of Electrical and Computer Engineering, The University of Texas at Austin, Jan. 2015.

[51] A. Pyattaev, K. Johnsson, S. Andreev, and Y. Koucheryavy, "3GPP LTE traffic offloading onto WiFi direct," in *Proc. of IEEE Int'l Workshop on Mobile Internet: Traffic Modeling, Subscriber Perception Analysis and Traffic-aware Network Design*, Shanghai, China, Apr. 2013.

[52] Y. Lin, B. Li, and B. Liang, "Stochastic analysis of network coding in epidemic routing," *IEEE J. Select. Areas Commun.*, vol. 26, no. 5, pp. 794–808, Jun. 2008.

Zehua Wang (S'11) received the B.Eng. degree in Software Engineering from Wuhan University, Wuhan, China, in 2009, and the M.Eng. degree in Electrical and Computer Engineering from Memorial University of Newfoundland, St John's, NL, Canada, in 2011. He is currently a Ph.D. candidate at the University of British Columbia (UBC), Vancouver, BC, Canada. His research interests include device-to-device communications, social networks, machine-type communications, beamforming design, and routing and forwarding in mobile ad hoc networks. He is the recipient of the *Chinese Government Award for Outstanding Self-Financed Students Abroad* in 2015. He has been the recipient of the *Four Year Doctoral Fellowship (FYF)* at UBC since 2012. He was also awarded the *Graduate Support Initiative (GSI) Award* at UBC in 2014 and 2015. Mr. Wang served as the technical program committee (TPC) member for several conferences including *IEEE International Conference on Communications (ICC) 2012, 2014–2017*, *IEEE Global Communications Conference (Globecom) 2014–2016*, and *IEEE Vehicular Technology Conference (VTC) 2012-Fall, 2016-Fall*.

Hamed Shah-Mansouri (S'06, M'14) received the B.Sc., M.Sc., and Ph.D. degrees from Sharif University of Technology, Tehran, Iran, in 2005, 2007, and 2012, respectively all in electrical engineering. He ranked first among the graduate students. From 2012 to 2013, he was with Parman Co., Tehran, Iran. Currently, Dr. Shah-Mansouri is a post-doctoral research and teaching fellow at the University of British Columbia, Vancouver, Canada. His research interests are in the area of stochastic analysis, optimization and game theory and their applications in economics of cellular networks and mobile cloud computing systems. He has served as the publication co-chair for the IEEE Canadian Conference on Electrical and Computer Engineering 2016 and as the technical program committee (TPC) member for several conferences including *IEEE Global Communications Conference (Globecom) 2015* and *IEEE Vehicular Technology Conference (VTC) 2016-Fall*.



Vincent W.S. Wong (S'94, M'00, SM'07, F'16) received the B.Sc. degree from the University of Manitoba, Winnipeg, MB, Canada, in 1994, the M.A.Sc. degree from the University of Waterloo, Waterloo, ON, Canada, in 1996, and the Ph.D. degree from the University of British Columbia (UBC), Vancouver, BC, Canada, in 2000. From 2000 to 2001, he worked as a systems engineer at PMC-Sierra Inc. (now Microsemi). He joined the Department of Electrical and Computer Engineering at UBC in 2002 and is currently a Professor. His research areas include protocol design, optimization, and resource management of communication networks, with applications to wireless networks, smart grid, mobile cloud computing, and Internet of Things. Dr. Wong is an Editor of the *IEEE Transactions on Communications*. He is a Guest Editor of *IEEE Journal on Selected Areas in Communications*, special issue on "Emerging Technologies" in 2016. He has served on the editorial boards of *IEEE Transactions on Vehicular Technology* and *Journal of Communications and Networks*. He has served as a Technical Program Co-chair of *IEEE SmartGridComm'14*, as well as a Symposium Co-chair of *IEEE SmartGridComm'13* and *IEEE Globecom'13*. He is the Chair of the IEEE Communications Society Emerging Technical Sub-Committee on Smart Grid Communications and the IEEE Vancouver Joint Communications Chapter. He received the 2014 UBC Killam Faculty Research Fellowship.