



Research Article

Social-Aware Device-to-Device Offloading Based on Experimental Mobility and Content Similarity Models

Lynn Aoude,¹ Zaher Dawy^{ID},¹ Sanaa Sharafeddine,² Karim Frenn,¹ and Karim Jaled²

¹Department of Electrical and Computer Engineering, American University of Beirut, Lebanon

²Department of Computer Science and Mathematics, Lebanese American University, Beirut, Lebanon

Correspondence should be addressed to Zaher Dawy; zd03@aub.edu.lb

Received 23 December 2017; Revised 27 March 2018; Accepted 18 April 2018; Published 16 May 2018

Academic Editor: Maode Ma

Copyright © 2018 Lynn Aoude et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Device-to-device (D2D) offloading has been shown to be a highly effective technique to enhance the performance of wireless networks. Yet, for any two mobile users to share data efficiently and reliably via D2D links, they should be in close proximity for long enough period of time, share similar content interests, and have some level of incentive and trust to cooperate. In this work, we focus on the practical implementation aspects of D2D data sharing taking into account realistic operational conditions. To this end, we design and conduct an experimental study to collect location and neighbor discovery data from 38 mobile users in a university campus over several weeks using our own customized crowdsourcing Android mobile application. The collected data is then processed and utilized to empirically model mobility-related parameters that include contact frequency, contact duration, and inter-contact duration. The participating users did also fill a user interest survey in order to correlate mobility and connectivity patterns with content interests and social network relations. The obtained insights are then used to develop a practical implementation framework for designing effective D2D data sharing strategies. To test the proposed ideas under realistic operational constraints, we design and implement a social-aware D2D data sharing Android mobile application and demonstrate its functionality and effectiveness using an example case study scenario.

1. Introduction

The computing and communications capabilities and the user-friendly interfaces of smartphones are facilitating the creation and sharing of content between devices without relying on existing wireless infrastructure, such as WiFi access points or cellular base stations [1–3]. Data sharing over device-to-device (D2D) links has been shown to provide a multitude of performance gains in terms of throughput and spectral efficiency enhancement [4–6], energy consumption reduction [7–10], and end user cost savings and incentives [11–13]. However, reliable and efficient D2D data sharing in practice hinges on several factors that include content availability, users' proximity and ability to connect to each other, users' mobility patterns, and contact intervals [14], in addition to other aspects such as cooperation incentives and trust establishment [15, 16].

Supporting D2D communications in wireless networks can be achieved with different levels of network operator

assistance ranging from full to partial to none; for example, see [17, 18]. We focus on a fully device-centric approach which requires direct coordination and communication between the devices over short-range wireless links (e.g., Bluetooth or WiFi-Direct) via mobile application solutions. This makes the proposed ideas and techniques independent from existing long-range wireless technologies such as WiFi or cellular networks and, thus, does not require any modification to existing wireless standards.

The main objective of this work is to address practical aspects of D2D data sharing following an experimental approach using our own developed crowdsourcing Android mobile application in order to capture realistic system and performance constraints. We model empirically key metrics for effective D2D data sharing including the number of contacts, contact duration, and inter-contact duration among a set of mobile users in a bounded geographical environment. We complement these empirical models with data collected via user surveys on content similarity and

social networking relations. We then overlay the contact models, content similarity, and social relationship links to generate novel insights that can help in designing efficient and reliable D2D data sharing strategies. Finally, we demonstrate the practical feasibility of the various presented ideas by performing experiments using our own developed social-aware D2D data sharing solution based on an Android mobile application. The key novelty of our work is twofold: (i) holistic approach by jointly capturing within one framework the users' mobility and inter-contact statistics, content interests, and social relations; (ii) experimental methodology with empirical models in order to obtain practical results that can bridge a gap between theoretical studies and real wireless devices, systems, and technologies.

In the literature, there have been several experimental studies focused on mobility and contact modeling for D2D data sharing. In [19], location data was collected by using LifeMap application on mobile devices carried by four graduate students for eight weeks in a $15 \times 20 \text{ km}^2$ area in Seoul, Korea; mobility information was used to devise an efficient routing algorithm that learns the users' mobility patterns and predicts future interactions. In [20], the authors gathered traces on a university campus using HYCCUPS tracer [21], a mobile application running in the background that collects information about a device's encounters with other nodes or access points. Information was gathered via Bluetooth and AllJoyn. This experiment spanned 64 days and involved 66 participants. A node's past encounters were analyzed and their time series was approximated as a Poisson distribution, representing the probability that a node will make a given number of distinct contacts within a specific time frame. The authors show that the Poisson distribution prediction is accurate if computed as an average per hour per day of the week, since academic schedules are repetitive and have an hour as a unit of time. In [22], six different traces were analyzed to study smartphone mobility. Data was gathered via WiFi, GPS, GSM, and Bluetooth. Analysis of these traces showed a power-law distribution of inter-contact time between smartphones up to a characteristic time, then an exponential decay. In [23], the number of connected pairs over time was evaluated, taking into consideration multihop contacts; it was observed that number of nodes in k -contact is much larger than the number of nodes in direct contact. All these references do relate to the scope of our work, yet they are mostly focused on mobility and contact duration modeling and analysis without capturing the other dimensions of content similarity and social relations. Moreover, they demonstrate the importance of relatively small-scale experimental studies in a bounded geographical areas as they can provide insights into use cases with hot spot locations and homogenous communities (e.g., university campuses, schools, factory facilities, and enterprises).

Another line of related work in the literature has focused on content similarity among mobile users. For example, in [24], data was collected from various services to obtain the distribution characteristics of requests across videos, the evolution of viewer's focus, and the shifts in content popularity; data was collected by crawling the websites of four

different online video providers. User generated content was modeled using a power law with a truncated tail probability distribution due to the fetch-at-most-once user behavior in peer-to-peer environments. The effect of this behavior is amplified when the number of videos is small and/or the average number of requests per user is large. The authors in [25] collected a list of shared iTunes songs from 239 users in 2006 at University of Missouri School of Journalism. The popularity distribution of song names proved to follow a Zipf-like distribution in a log-log scale, indicating that few objects are extremely popular while many objects are rare. User availability information was also collected and the number of unique users seen daily was plotted. The authors in [26] measured a file's popularity by its replication degree. They observed that few files are highly replicated, while most are not replicated at all. Another observation was the correlation between geographical clustering and video files: peers requesting a certain video file are more liable to download it from peers in their own country, thus reducing the delay and download time. In our work, we consider content interests within a given community in a university campus environment which will lead to new perspectives that do complement insights generated from the large-scale studies discussed above.

Recently, there has been increased interest in utilizing social networking relations among users to leverage content downloading and sharing for improved quality of experience. In [27], the authors presented interesting observations on the benefits of exploring social relations to enhance the performance of D2D communication systems and used simulation studies based on real traces. In [28], the authors proposed a novel framework to extract useful information from social networks and utilize it to improve the performance of D2D communications; they applied their framework to the problem of spectrum sharing for in-band D2D communications and presented numerical results to show significant improvements in spectrum utilization efficiency. The authors in [29] defined social spreading impact and mobility patterns of users to evaluate one's significance in propagating posts and sharing content via encounters; several factors were integrated in a social-aware D2D sharing framework to achieve major offloading of mobile traffic while satisfying users' delay requirements. From a different perspective, the authors in [30] used social relationship among users to address the problem of optimal peer selection in cooperative D2D communications to enhance data privacy in dynamic scenarios where users move within a given area. Users that suffer from poor channel quality opt to use nearby users with close social relation as relays to protect their privacy. Simulation results demonstrated the effectiveness of the proposed physical-social model to provide better data rates while protecting privacy. In [31], the authors considered social relations and user mobility attributes to select core users that may reliably and effectively assist base stations in distributing video content through D2D multicast. In [32], the authors utilized user mobility, opportunistic encounters, and social relations to build D2D data dissemination model based on a three-phase approach with analytical and simulation based performance analysis. In [11], the authors proposed

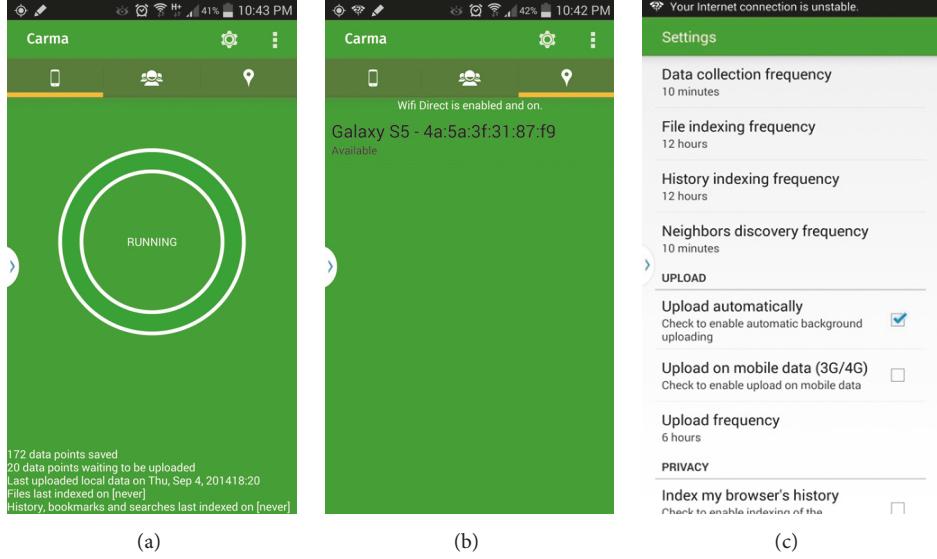


FIGURE 1: CARMA graphical user interface snapshots: (a) home screen; (b) neighbor discovery screen; (c) settings screen.

the integration of a social networking layer to enhance the efficiency of physical D2D communications using an analytical approach based on Indian Buffet Process with evaluation using simulation results based on real traces. This summary of selected recent works from the literature demonstrates the benefits of utilizing metrics based on social networks' relationships in the design of D2D data sharing mechanisms; this is one of the key components of our proposed framework and analysis.

The rest of the paper is organized as follows. We describe in Section 2 the experimental study design and tools including the data collection process. We summarize in Section 3 the obtained mobility and content similarity related results, in addition to the derivation of empirical models with discussion and analysis. We then present in Section 4 a practical framework for designing effective D2D data sharing strategies, in addition to the architecture, implementation, and evaluation of a social-aware mobile solution for dynamic D2D data sharing using Android devices. Finally, conclusions are drawn in Section 5.

2. Experimental Study: Design

Our experimental study design consists of two independent and complementary phases: a mobility-related data collection phase and a content similarity data collection phase.

2.1. Mobility-Related Data Collection. Mobility-related data is collected using our own crowdsourcing mobile application called *Context Aware Resource Management App*, or CARMA for short (see Figure 1 for snapshots of the graphical user interface). CARMA runs in the background on Android smartphones and collects seamlessly device location and neighbor discovery information and then uploads the data points periodically to a database at a remote server.

This database contains two tables: Data and Neighbors. The information stored in the Data table is as follows: time stamp (msec), device ID, number of bytes sent and received, battery percentage, latitude, longitude, accuracy, and location provider. As for the Neighbors table, the data stored includes the following: time stamp (msec), device ID, discovered device name and MAC address, latitude, longitude, and accuracy.

The aim of this experimental study is to collect data in a confined and controlled environment, with a pool of participants having different pairwise contact frequencies; the scale of the study is similar to [19], yet the dynamics of the environment are different. Data collection was done in two stages on the 60.9 acre campus of the American University of Beirut (AUB): the first stage during the 2015 spring semester and the second stage during the 2015 summer semester. The first stage started on March 9, 2015, and ended on May 20, 2015, for a duration of 11 weeks with 37 mobile user participants. Data collection frequency for both neighbor discovery and location information were set to 10 minutes, resulting in 21,289 Neighbors table entries and 354,528 Data table entries stored in our database. The second stage started on June 11, 2015, and ended on July 6, 2015, for a duration of approximately 4 weeks with 13 mobile user participants. Neighbor discovery interval was set to one minute, while location information was logged every two minutes. Consequently, 34,920 Neighbors table entries and 326,539 Data table entries were stored in the database. In Section 3, we will use this collected data to derive empirically the cumulative distribution function (CDF) of various mobility-related metrics that include number of pairwise contacts with respect to day and time interval aggregated over all device pairs, in addition to contact and inter-contact durations.

2.2. Content Similarity Data Collection. Due to privacy limitations, it was not possible to activate the tracking of

TABLE 1: CARMA data collection settings during spring and summer stages.

Settings	Spring	Summer
Area	AUB campus	AUB campus
Trace duration (days)	77	25
Location data collection frequency	10 minutes	2 minutes
Neighbor discovery frequency	10 minutes	1 minute
Data points upload frequency	1 hour	1 hour
Number of participants recruited	37	13
Number of points in data table	21,289	34,920
Number of points in neighbors table	354,528	326,539
Number of participants correlated with survey	8	4

users' stored content and download activity on their devices using the CARMA mobile application. To deal with this constraint, we designed and deployed an online survey to collect relevant content similarity data from mobile users including a subset who participated in the mobility-related data collection experiment. The survey has 28 questions that focus on users' interests and content stored on their smartphones, downloaded or shared, as well as on social network activity.

The survey questions were divided into five groups, each centered around a certain aspect that relates to D2D data sharing. The first group of questions focused on the type of files stored on the user devices, as well as on the user interests in the five categories of files mentioned. The second group inquired about the users' sources of media files. The third group concentrated on the number of files stored and streamed on the users' devices. The fourth group served to assess the users' willingness to download and share their files via direct D2D data sharing. The final group captured the users' social activity regarding four social networks: Facebook, Instagram, Twitter, and LinkedIn. Using the data collected, content similarity level between any two users can be calculated. To correlate the survey's results with the mobility and contact data analysis, participants were asked to include their names along with their responses. We received 31 responses in total for the survey, including 12 who included their names and participated in the mobility-related data collection experiments.

Table 1 summarizes key parameters related to the experimental study for both stages. In the data collection stage during the summer, we decreased the location data collection frequency and the neighbor discovery frequency since the duration of the summer term is shorter than the spring term and the number of participants is relatively lower. This facilitated the collection of enough data to generate reliable models and helped capture user mobility at a faster rate.

As in [19, 20], even though the data collected is of relatively limited size and focused on a hot spot geographical area, it provides important insights as dynamic and effective D2D data sharing is expected to be more feasible in closed communities with low mobility and high level of trust; in addition, it can lead to useful observations on how to scale up to wider geographical regions with larger density of mobile users. Having existing similar published experimental

studies does not diminish the value of our experimental work as every study focuses on different goals and has its own specificities due to the variations in the environment and users' mobility patterns. The outcomes of these various studies provide real traces that can be valuable to other researchers for evaluation purposes.

3. Experimental Study: Results and Analysis

In this section, we present a summary of the obtained results from the two stages (spring semester experiment and summer semester experiment) of data collection using CARMA; we use this data to derive several empirical models for mobility-related metrics that would be useful for future studies and enhancements related to D2D communications. Moreover, we analyze the user survey results and social network relations to correlate content similarity and social links with mobility and contact data in the context of D2D data sharing between mobile users in a given geographical area.

3.1. Mobility-Related Parameters

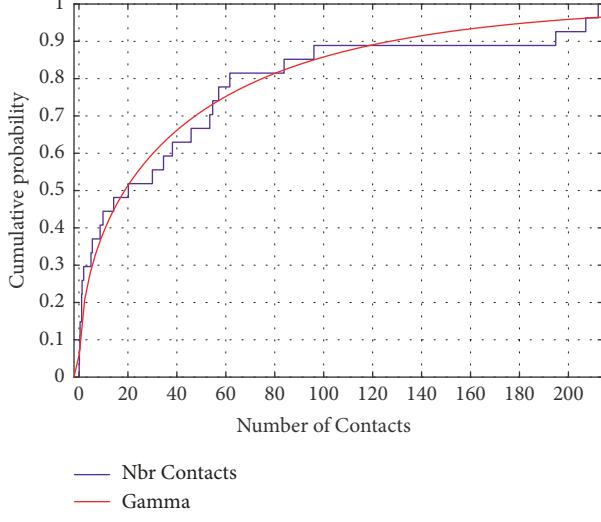
3.1.1. Number of Contacts. The number of contacts between two devices is a key measure to determine whether D2D data sharing is likely between them. The CDF of the normalized number of contacts per day and that of the normalized number of contacts per time interval are derived empirically using the collected data. The empirical aggregated CDF for the normalized number of contacts per day, as well as per time interval, are shown in Figures 2 and 3 to follow a Gamma distribution with an acceptable accuracy level. The Gamma distribution is given by the following:

$$F(x) = \frac{\Gamma(\beta, x/\alpha)}{\Gamma(\beta)} \quad \text{for } x > 0, \quad (1)$$

where α is the shape parameter, β the scale parameter, and $\Gamma(\beta) = \int_0^\infty t^{\beta-1} e^{-t} dt$ the incomplete Gamma function. The Gamma distribution parameters for the data collected from both semesters are presented in Table 2.

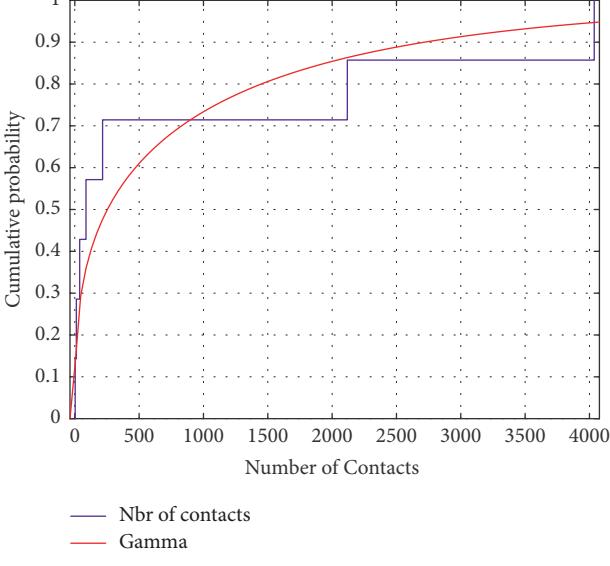
Referring to Figure 2, 80% of the number of contacts recorded are less than 61 contacts per day during the sample results from the spring data collection stage, and less than

Monday



(a)

Tuesday

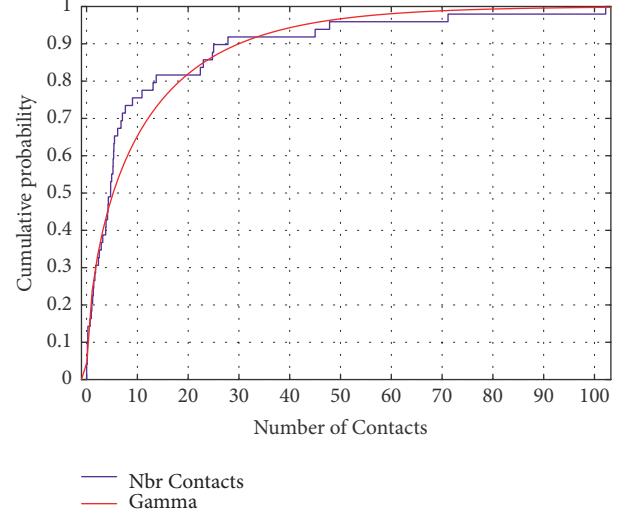


(b)

FIGURE 2: Selected empirical aggregated CDF for the normalized number of contacts per day for (a) Monday in spring (b) Tuesday in summer.

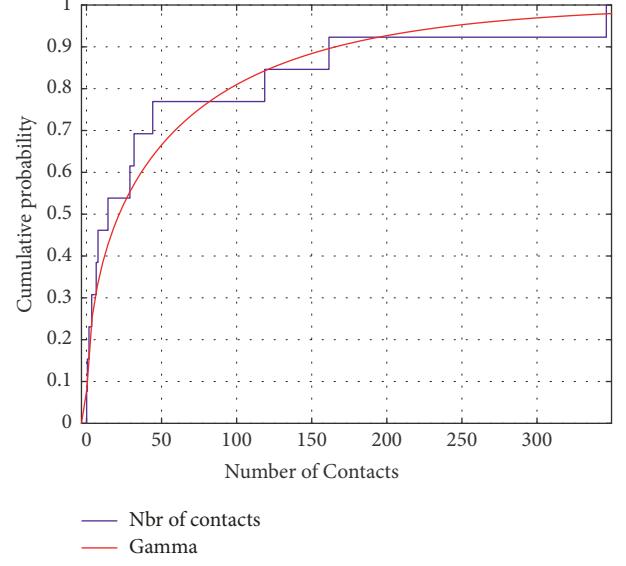
2,118 contacts per day during the summer data collection stage. The disparity in scale between these two numbers is due to the different neighbor discovery data collection frequencies between spring and summer stages (set to 10 minutes in the spring compared to 1 minute in the summer). Referring to Figure 3, 80% of the number of contacts recorded are less than 15 every four hours during the spring, and less than 125 every four hours during the summer, provided that these four-hour time intervals fall during workdays. We notice that the distribution of the number of contacts is steeper in the summer in comparison with the spring. This means that, during the spring, the number of contacts is well distributed within the ranges 1 to 61 per day or 1 to 15

10:00–14:00 hours



(a)

6:00–10:00 hours



(b)

FIGURE 3: Selected empirical aggregated CDF for the normalized number of contacts per time interval for (a) 10:00 to 14:00 in spring and (b) 6:00 to 10:00 in summer.

per time interval. On the other hand, during the summer, the number of contacts is distributed unevenly with values having mostly small or large values and with few values being in the middle of the ranges (1 to 2118 contacts per day and 1 to 125 contacts) per time interval. These results demonstrate the notable variability in users' group behavior during different time periods (spring versus summer) even with a homogenous group of users in the same geographical area.

Additionally, we investigate how the number of users in a study affects the statistics of the total number of contacts. Thus, we arbitrarily select different sets of users of size ranging from 5 to 25. We start with a set of 5 users and then

TABLE 2: Gamma distribution parameters for the number of contacts CDF.

	Spring		Summer	
	per day	per time interval	per day	per time interval
Shape (α)	0.45	0.52	0.32	0.41
Scale (β)	103.12	21.57	2889.52	144.68
% fit	90%	90%	76%	88%

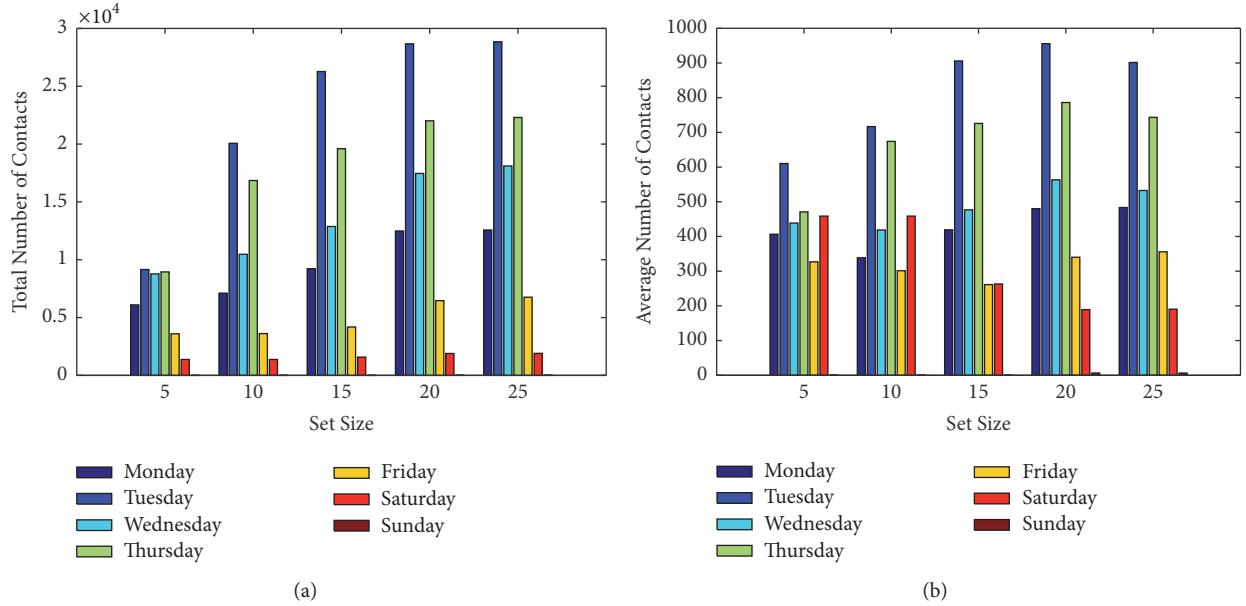


FIGURE 4: (a) Total and (b) average number of contacts in function of different sets of users with respect to day.

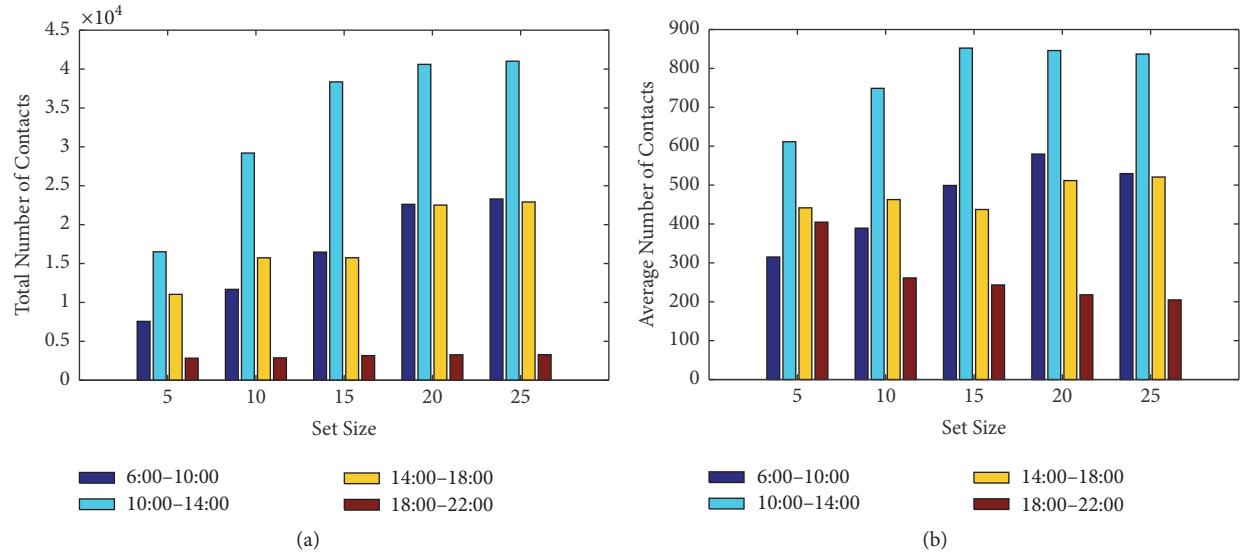


FIGURE 5: (a) Total and (b) average number of contacts in function of different sets of users with respect to time interval.

incrementally increase by 5 additional users up to 25. The total and average number of contacts with respect to day and time interval are plotted in Figures 4 and 5. As expected, the larger the set of users, the higher the number of contacts obtained, yet the relation is not necessarily linear. In addition,

the average number of contacts per user is dependent on the selected set of users.

We also calculate the pairwise number of contacts, defined as the number of contacts between a specific user pair, for both semesters (see Figure 6). We obtain an average of 155

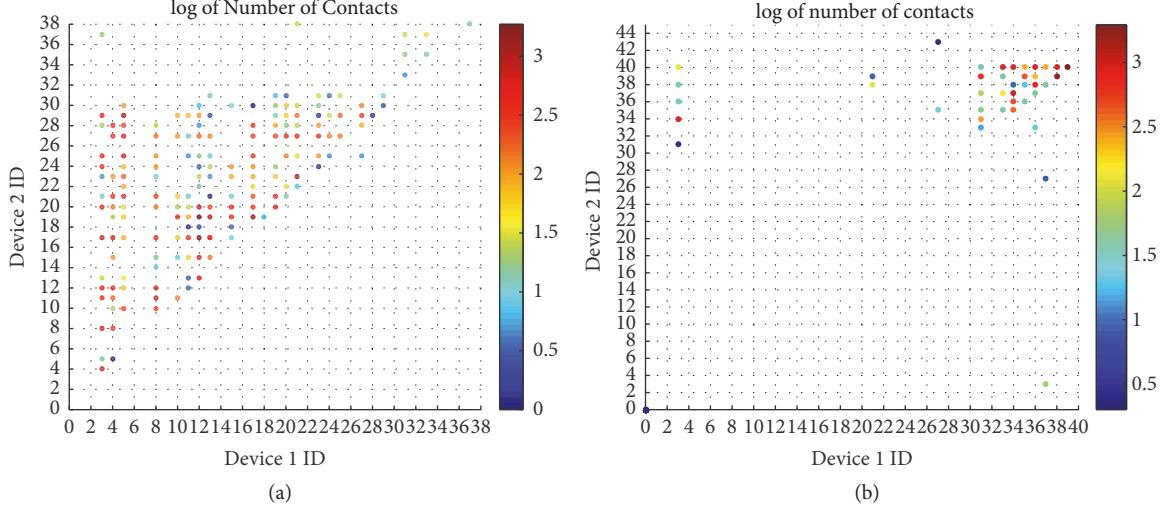


FIGURE 6: Pairwise number of contacts in log-scale for (a) spring and (b) summer.

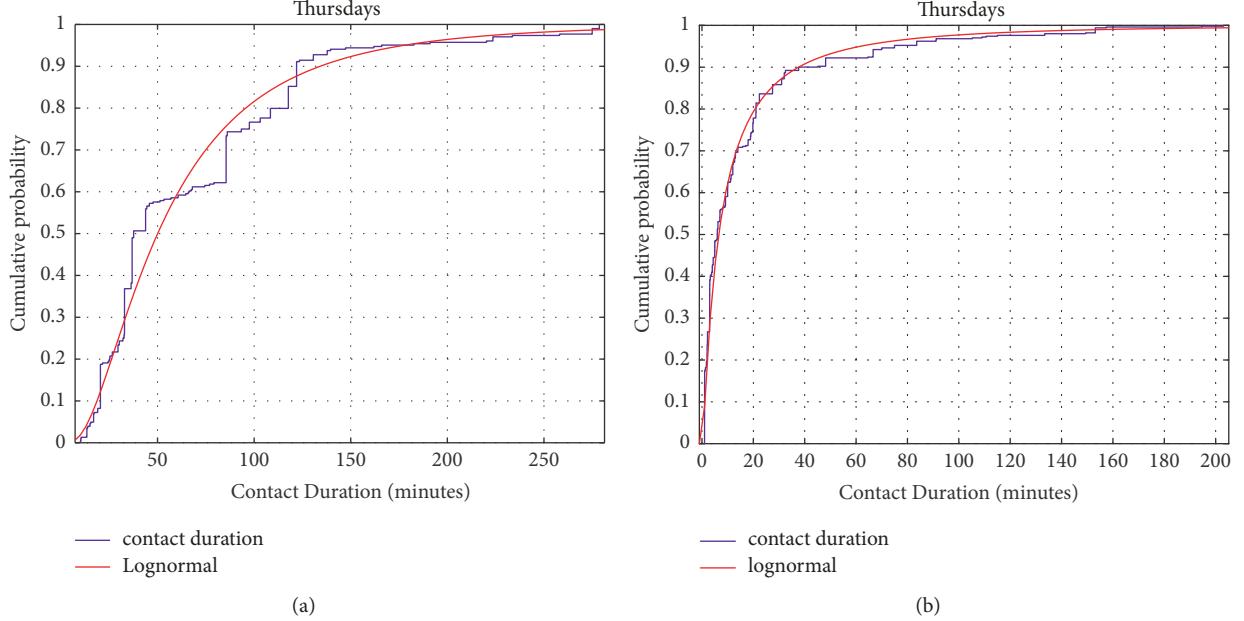


FIGURE 7: Pairwise aggregate CDF of contact duration with respect to day for (a) spring and (b) summer.

contacts per pair, with a minimum of 1 contact per pair and a maximum of 1867 contacts per pair during the spring. During this semester, the following five pairs show an exceptionally high number of contacts: (5; 29), (12; 17), (12; 19), (17; 19), and (21; 23). As for the summer, the maximum pairwise number of contacts is 1965, the minimum is 2, and the average is 265. Three pairs show an exceptionally high number of contacts: (3; 34), (38; 39), and (39; 40). For example, user device 3, which participated in both data collection phases, is shown to have consistently a high number of contacts with its neighbors. Thus, device 3 can be considered as a reliable source for D2D data sharing use cases and, thus, can be labeled as reliable collaborator and provided with incentives

to help offload network traffic; it can be also targeted for mobile caching services.

3.1.2. Contact Duration. The pairwise contact duration CDF aggregated over all users is studied with respect to day and time (see Figures 7 and 8). In both cases, the CDF is shown to follow accurately a lognormal distribution:

$$F(x | \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} \int_0^x \frac{e^{-(\ln(t)-\mu)^2/2\sigma^2}}{t} dt, \quad (2)$$

where μ and σ are the location and scale parameters, respectively. The obtained empirical parameter values for

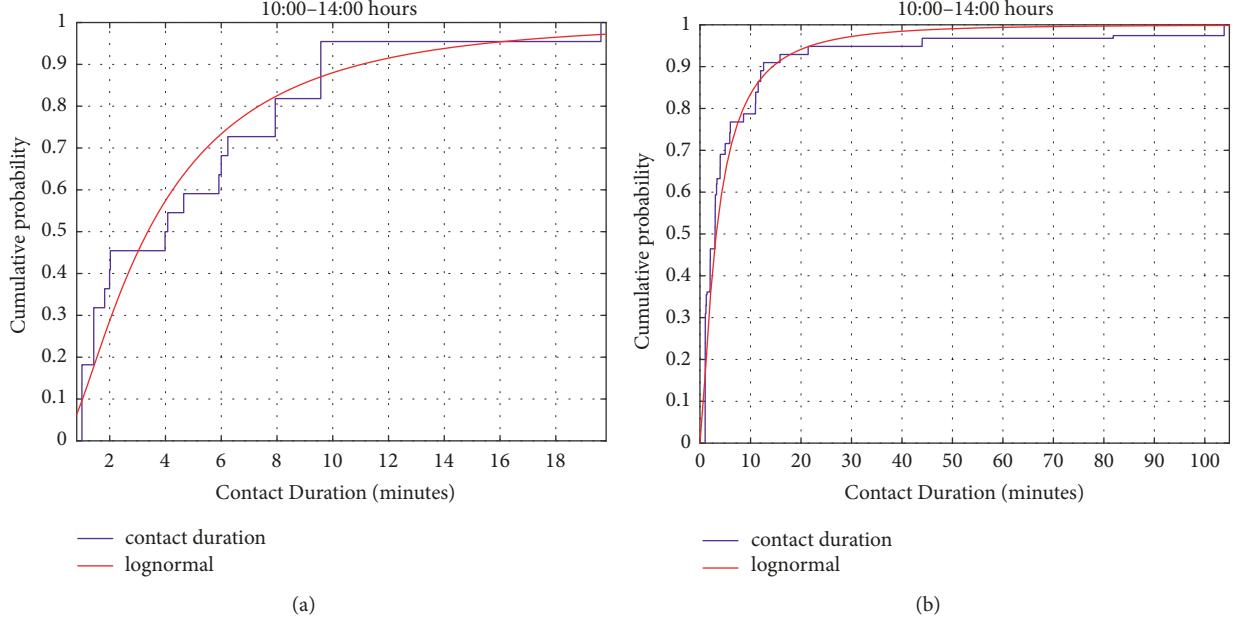


FIGURE 8: Pairwise aggregate CDF of contact duration with respect to time interval for (a) spring and (b) summer.

TABLE 3: Lognormal distribution parameters for pair-wise contact duration CDF.

	Spring		Summer	
	per day	per time interval	per day	per time interval
μ	3.91	1.88	1.21	1.16
σ	0.77	1.37	0.93	1.17
% fit	95%	95%	80%	90%

selected time windows during spring and summer stages are summarized in Table 3.

For example, results show that 80% of the pairwise contact duration values recorded are less than 125 minutes per the selected day during the spring and less than 20 minutes during the summer. On the other hand, 80% of the pairwise contact duration values recorded are less than eight minutes per the selected four-hour interval during the spring and less than 12 minutes during the summer. The contact duration is a critical factor in D2D data sharing applications as it determines the time needed to download a given content (or parts of it) as a function of the communications bit rate quality between the cooperating devices.

3.1.3. Inter-Contact Duration. We have also analyzed the inter-contact duration statistics per day and time interval. The inter-contact duration is defined as the time period between two contact events for given a pair of users; that is, it captures the time during which the users are not in proximity with respect to each other. For example, results show that, during the spring, 479 inter-contacts occurred during the whole trace, with a minimum duration of 10 minutes, a maximum duration of approximately 58 days, and an average duration of approximately 26 hours. However, during the summer, 1111 inter-contacts occurred during the whole trace, with a minimum duration of 1 minute, a maximum duration of

approximately 5 days, and an average duration of around 46 minutes. Note that the minimum durations are determined by the neighbor discovery frequency parameter which was set to 10 minutes and 1 minute during the spring and summer data collection stages, respectively (see Table 1).

3.2. Content-Related Parameters. The content similarity analysis is solely based on the survey responses. Based on the 31 responses received, statistics show that the majority of the responders store about four to six media files per day on their hand-held devices. The files they store are distributed to include all seven categories considered (personal documents, public documents, personal images, public images, personal videos, public videos, and music) with high percentages. Even though personal files occupy the majority of files stored by the participants, public media files also occupy a high percentage of user storage with the most popular file type being images, followed by music. Table 4 presents a summary of the survey results relating to users' interests. When asked about their source of music, responders mostly use online sharing (download) websites such as Rapidgator, rather than streaming websites. However, the majority of responders prefer streaming videos rather than downloading and storing them on their devices. In general, most responders stream less than 10 music and/or video files per week. As for the total number of music files stored on their devices, the majority

TABLE 4: User interest survey results.

Category	Popular Genres	Percentage
Documents	Lectures and course material	58.06%
	Articles of interest	51.61%
	Technical documents	45.16%
Images	Photos from the internet (e.g., 9 gag, reddit)	67.74%
Videos	Short humor clips/stand-up comedy	58.06%
	Music Videos	25.81%
Music	Blues and Jazz/Classical	48.39%
	Rock/R&B/Rap/Hip-Hop	45.16%
	Arabic music	41.94%
Movies	Comedy	80.65%
	Action	54.84%
	Science Fiction/Fantasy/Paranormal	48.39%

of our responders have less than 50 files; the same applies to document and video files. As for image files, the majority of the responders (83.9%) have more than 300 files stored on their devices.

Furthermore, when asked about D2D data sharing, responders indicated that they are willing to share and/or download all file categories over direct D2D links, that is, documents, images, videos, and music, with music (90.32%) and videos (87.1%) being the most popular choices. As for incentives to share files with neighboring devices using D2D links, 61.3% were favorable without any incentives and 96.8% were favorable when incentives are offered. The most popular incentives were extra 3G/4G cellular data quota (90.3%), followed by free call minutes for each file exchanged (58.1%) and credits for online shopping (54.9%).

For the content similarity analysis, two parts of the survey are relevant: the one related to user interest and the one related to the device content. A weight is calculated for each part by taking the number of common answers between any two participants divided by the corresponding total number of questions as follows:

$$\omega_1(i, j) = \frac{\# \text{ common answers of users } i \& j}{\text{total } \# \text{ user interest related questions}}, \quad (3)$$

$$\omega_2(i, j) = \frac{\# \text{ common answers of users } i \& j}{\text{total } \# \text{ device content related questions}}.$$

The final content similarity weight is then calculated as the average between these two weights as follows:

$$S(i, j) = \frac{\omega_1(i, j) + \omega_2(i, j)}{2}. \quad (4)$$

These have been only calculated for the responders who participated in the CARMA data collection phase during the spring semester. Responses are available for eight CARMA participants, since the rest did not include their names in their survey responses. The ID of these participants are {4, 10, 12, 17, 19, 27, 29, 30}.

The similarity matrix \mathbf{S} of the content similarity between the spring participants' devices is calculated according to (4) as follows:

$$\mathbf{S} = \begin{bmatrix} 1 & 0.558 & 0.678 & 0.602 & 0.703 & 0.609 & 0.435 & 0.533 \\ 0.558 & 1 & 0.644 & 0.526 & 0.661 & 0.686 & 0.484 & 0.637 \\ 0.678 & 0.644 & 1 & 0.57 & 0.713 & 0.772 & 0.57 & 0.585 \\ 0.602 & 0.526 & 0.57 & 1 & 0.484 & 0.543 & 0.529 & 0.501 \\ 0.703 & 0.661 & 0.713 & 0.484 & 1 & 0.713 & 0.484 & 0.568 \\ 0.609 & 0.686 & 0.772 & 0.543 & 0.713 & 1 & 0.509 & 0.703 \\ 0.435 & 0.484 & 0.57 & 0.529 & 0.484 & 0.509 & 1 & 0.425 \\ 0.533 & 0.637 & 0.585 & 0.501 & 0.568 & 0.703 & 0.425 & 1 \end{bmatrix}. \quad (5)$$

3.3. Social Network Relationships. The social relationship strength is analyzed using the survey participants' Facebook friends lists. In order to correlate mobility and content similarity results with participants' social network relationships, the number of mutual friends is considered to assess the strength level between any two participants. The number of mutual Facebook friends between participants is presented in the matrix \mathbf{M}_{spr} for the eight spring semester participants who also filled the user survey. The number of mutual friends compared to oneself is set to zero to simplify the graph construction.

$$\mathbf{M}_{\text{spr}} = \begin{bmatrix} 0 & 0 & 0 & 11 & 11 & 0 & 0 & 0 \\ 0 & 0 & 0 & 9 & 2 & 14 & 7 & 9 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 11 & 9 & 0 & 0 & 16 & 1 & 0 & 0 \\ 11 & 2 & 0 & 16 & 0 & 1 & 0 & 0 \\ 0 & 14 & 0 & 1 & 1 & 0 & 7 & 19 \\ 0 & 7 & 0 & 0 & 0 & 7 & 0 & 7 \\ 0 & 9 & 0 & 0 & 0 & 19 & 7 & 0 \end{bmatrix}. \quad (6)$$

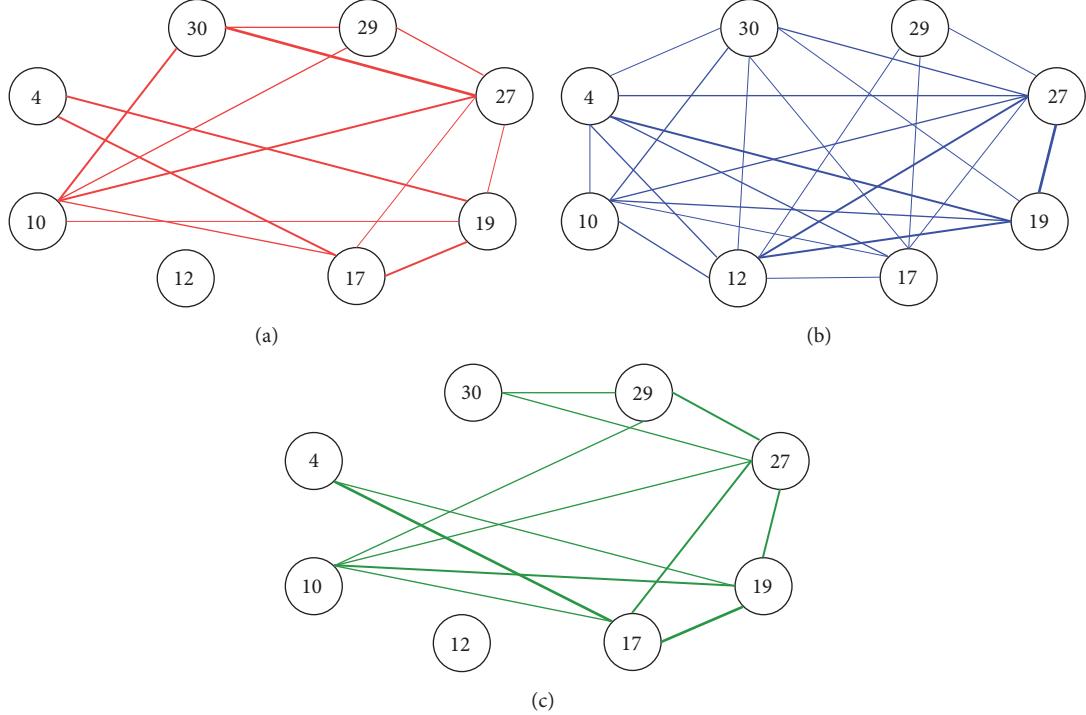


FIGURE 9: Multilayered graph representing the pairwise social network relationship strength (a), content similarity (b), and number of contacts (c) during the spring semester.

3.4. Correlation between Mobility Parameters, Content Similarity, and Social Network Links. In this section, the social relationship strength, the content similarity, and the number of pairwise contacts are correlated to identify favorable quantifiable conditions for successful D2D file sharing. The correlation between mobility, content, and social networks is done only for the spring semester, with results being summarized using graphs in Figure 9. Note that the thicker the edges connecting any two vertices, the higher the layer-respective measure. Looking at this figure, for each user pair one can roughly determine whether they have the potential to successfully share files via D2D links. When the edges connecting a pair of users are thickest on the three levels simultaneously, a D2D link has the highest chance of being successful while keeping in mind the constraints pertaining to the contact and inter-contact durations.

The aggregation of the three layers is visualized in Figure 10; the thicker the edges connecting any two vertices, the higher the correlation between the pairwise social relationship strength, content similarity, and number of contacts. Dotted edges indicate a correlation that allows only opportunistic or limited D2D data sharing, due to the low pairwise number of contacts. For example, according to Figure 10, pairs (4; 17), (4; 19), (27; 29), and (27; 30) have the highest chance of success in D2D data sharing as they meet the various conditions needed for successful cooperation. It is interesting to note that several edges are correlated across the three layers which reflects the potential for direct D2D data sharing mobile applications in practice.

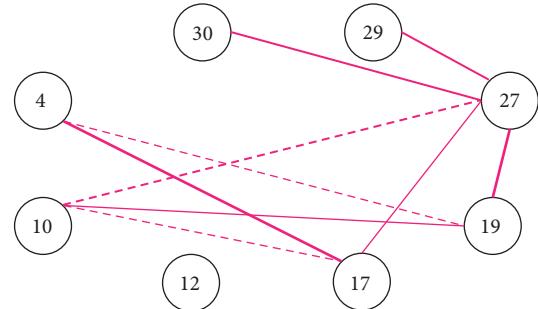


FIGURE 10: Aggregation of social network relationship strength, content similarity, and number of contacts during the spring semester based on the level of correlation.

4. Social-Aware D2D Data Sharing Mobile Solution

In this section, we use the insights derived from the mobility and content similarity experimental measurements and models from Section 3 in order to shed light on the design of practical D2D data sharing strategies. Even though our experimental study was confined to a limited geographical area, it does provide important information for similar hot spot scenarios and can serve as a best case for larger scale scenarios where mobile users are more spread geographically with high mobility and belong to diverse communities.

TABLE 5: Conditions for designing efficient and reliable D2D data sharing mobile solutions.

Case	Result
Low N_{ij} AND Low D_{ij}	Users are unreliable pair for D2D cooperation
High N_{ij} AND High D_{ij}	D2D cooperation likely to be successful, with large content size and low delay
Low N_{ij} AND High D_{ij}	D2D cooperation likely to be successful, with medium content size and low delay
High N_{ij} AND Low D_{ij} AND High ID_{ij}	D2D cooperation can be successful, but with limited content size and high delay
High N_{ij} AND Low D_{ij} AND Low ID_{ij}	D2D cooperation can be successful, but with limited content size and low delay

4.1. Practical Design Considerations. The following is a summary of key factors to design practical D2D data sharing strategies:

- (1) Social network relationship between users is a prerequisite enabler for D2D data sharing, as its strength is an indicator of the users' level of trustworthiness with respect to each other; for example, see [14, 15, 27]. The importance of using social network relationship as a trust level indicator is that it can be implemented seamlessly within D2D data sharing mobile solutions without the need for direct user intervention or dedicated server authentication. Mobile applications can extract and aggregate this information using existing social media APIs.
- (2) For a user to download a data file over a direct D2D connection, the content of interest should be available in one of its neighboring devices. This requires running service and peer discovery procedures to identify devices that are willing to cooperate and within radio range from the requesting device; this will be followed by content discovery phase to check if any of the trusted devices has the content locally cached. The user survey results presented in Section 3.2 demonstrated that users in hub locations (such as university campus and factory facility) do indeed have common interests when it comes to types of content stored on their devices; yet, the survey also confirmed the need to provide attractive incentives for users to accept to participate.
- (3) In addition to trust, incentives, and content availability, D2D data sharing success depends on the number and duration of contacts between a pair of users over time and space. Let N_{ij} represent the pairwise number of contacts, D_{ij} the pairwise contact duration, and ID_{ij} the pairwise inter-contact duration over a given period of time where i and j are device indices. The five cases in Table 5 summarize a set of conditions for designing decision making strategies in D2D data sharing mobile solutions.

The experimental measurements and derived empirical models in Section 3 can be utilized for multiple purposes. First, they can be used to derive theoretical performance results quantifying the probability of successful content sharing over D2D links in a given hub location as a function of parameters such as content size, content availability probability, and content download delay target; the models are generic and parameterized which allows for customization to various

network scenarios. Second, they can be used to develop effective mobile application solutions for direct D2D data sharing based on historical mobility and content similarity data. It is important to note that initiating service and device discovery functions and setting up connections with nearby devices to request content are time and resource consuming functions which can deplete energy and overload devices. Therefore, these activities need to be initiated only when the probability of successful D2D data sharing is high enough; this success probability is a function of various parameters that can be monitored through a sliding window over time, such as N_{ij} , D_{ij} , and ID_{ij} . Finally, D2D data sharing can be highly effective even if different parts of the same data content are downloaded from different peer devices as long as they are aggregated properly and the total cumulative download delay meets any end user target constraints. In the following section, we demonstrate how such ideas can be realized in practice using our own developed mobile application solution.

4.2. Mobile Solution: Design and Implementation. We have developed and implemented a social-aware D2D data sharing mobile solution for Android smartphones based on the architecture in Figure 11. The application is divided into the following five inter-connected modules that include back-end components and an interactive user interface.

4.2.1. Authentication Service. This module is responsible for authenticating the device by connecting the mobile application to the user's social networking profile which includes list of friends or contacts; we used Facebook in our implementation, without loss of generality. This process starts by the user providing her/his Facebook credentials which are then authenticated with a Facebook server. If valid, the Facebook server will return a unique identifier for the user's profile along with the user's friends list. The user's friends list is then saved in a local database while the user's ID and D2D wireless link MAC address are saved in a remote global database; the wireless link can be either WiFi-Direct or Bluetooth as both provide the needed discovery and connectivity features to facilitate stable direct D2D connections. In our implementation, we used Bluetooth.

4.2.2. Database Service. This module is responsible for a background task whose job is to synchronize a global database of files and devices whenever feasible; for instance, whenever the device connects to the Internet. The database contains only the meta-data about the files that are being shared by the devices. More precisely, for each file that is

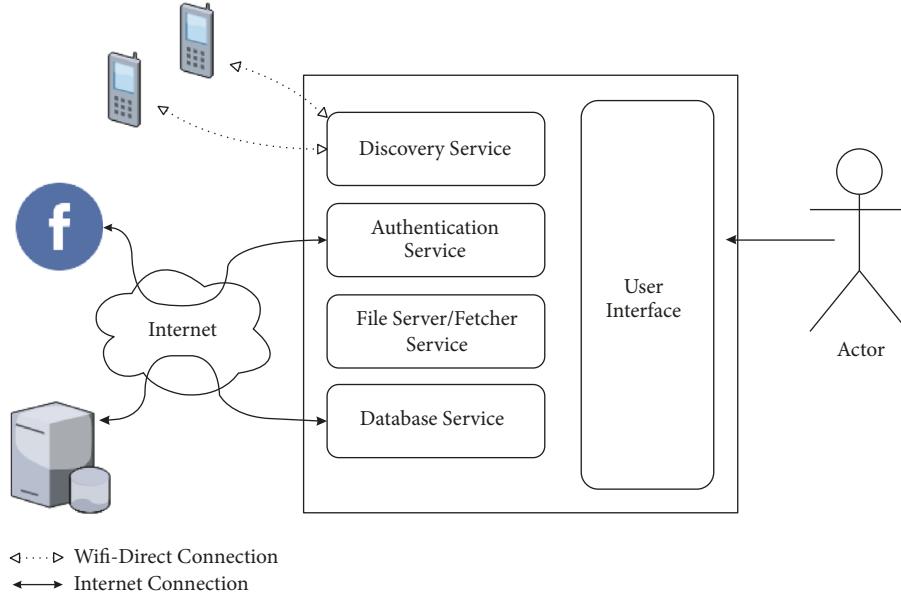


FIGURE 11: D2D data sharing mobile solution architecture.

shared by some mobile device, we maintain the following information: the file's unique signature (md5 hash), file size, user-friendly file name, and a list of device identifiers that have the file available for sharing; we assume the device identifier is the MAC address of its WiFi-Direct or Bluetooth interface. In addition, the database is used to map between Facebook user identifiers and device MAC addresses. The database contains a tuple (MAC_ADD, FB_UID) for each device with a WiFi-Direct/Bluetooth interface having MAC_ADD as a MAC address and authenticated with a Facebook profile having FB_UID as identifier.

4.2.3. Discovery Service. This module is intermittently launched in the background, at predetermined intervals, to discover nearby devices. The device discovery mechanism is implemented as a standard Bluetooth discovery procedure. A list containing the MAC addresses of discovered devices is sent to the File Server Service module.

4.2.4. File Server Service. This module runs periodically in the background and can operate in two modes: either as a file server or as a fetcher. By default, this module is in the file server mode waiting for incoming requests from other devices. This is implemented using a standard TCP file transfer procedure; the device listens on a specific TCP port for incoming file transfer requests. A request is composed of two parameters: the unique signature of the requested file along with a byte offset; the byte offset is needed to allow a device to download different parts of the content from multiple devices. Whenever such request is received, the device file server will begin sending the requested file after skipping to the specified offset.

In the file fetcher mode, this module keeps a lookout for files that are requested by the user. Whenever a new list of nearby devices is received from the device discovery module

and the device is not actively transmitting a file, the file fetcher module checks whether it can obtain one of the files requested by the user by the following:

- (i) Filtering the list of discovered devices and keeping only those that meet a set of conditions, for example, as discussed in Section 4.1 and Table 5
- (ii) Fetching the list of requested files and cross-matching it to the list of filtered discovered devices to check whether any of the selected devices has a file of interest.

If such a device is found, a Bluetooth connection request is sent to the device. If the connection is successfully established, the file fetcher service connects to the file server running on the other device and the file transfer process begins (see Figure 12 for interaction flowchart). When a file is fully fetched, it will be removed from the list of requested files and the WiFi-Direct connection will be terminated. Otherwise, if the transfer was interrupted for any reason, the current download progress (number of bytes downloaded so far) is saved in a local database so that download may resume when the opportunity presents itself again.

4.2.5. Front-End Graphical User Interface. Figures 13 and 14 present snapshots from the user interface of the mobile application demonstrating different features. Figure 13(a) shows the list of files that the user would like to download with a symbol indicating the size and location of each file; Figure 13(b) shows the list of Facebook friends; Figure 14(a) shows the list of nearby discovered WiFi-Direct devices; and Figure 14(b) demonstrates initiation of a download process for the file “Wonders_of_Nature.mp4” starting at position 91.5 MB from a user nearby who belongs to the friends list. We ran several test cases using our developed application and were able to demonstrate the validity of

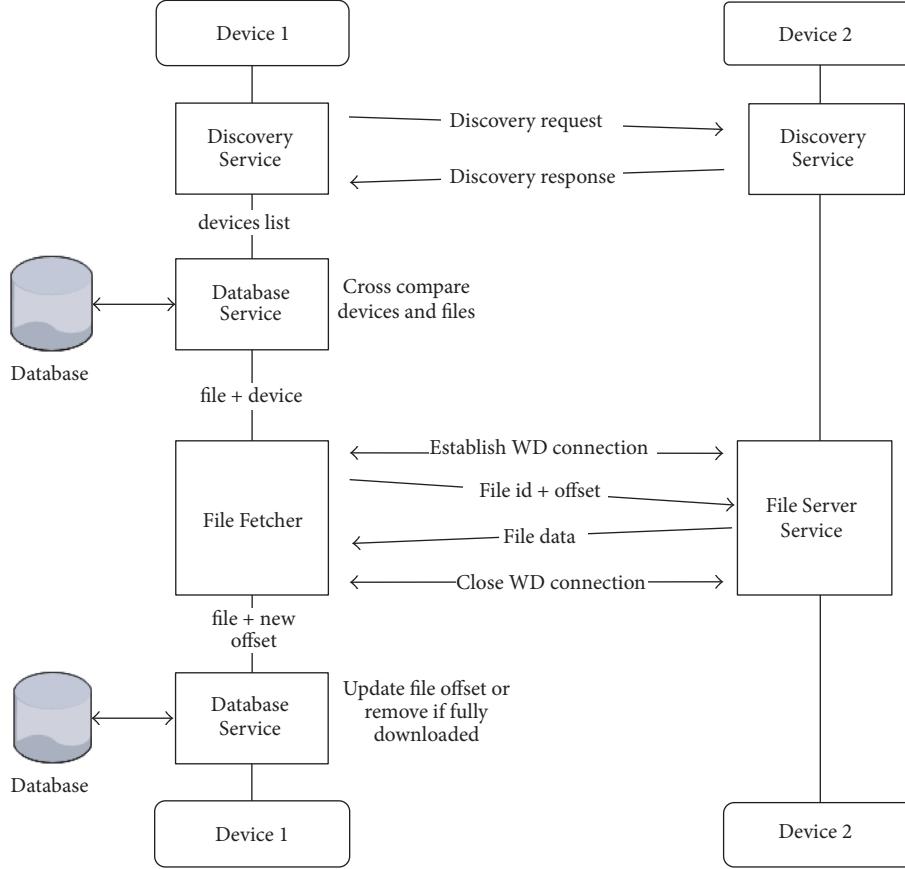


FIGURE 12: Device-to-device interaction flowchart including both discovery and file fetching services.

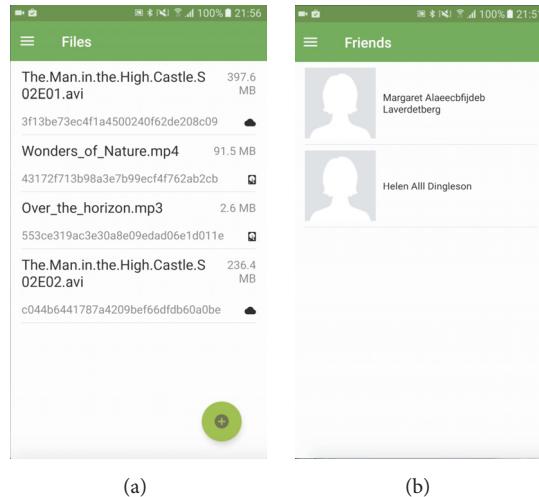


FIGURE 13: (a) List of files; (b) list of Facebook friends.

social-aware D2D data sharing with a given file downloaded to completeness from multiple devices over a window of time. Our implementation fits under the scope of the general area of delay tolerant networks whereby content download allows for a certain level of delay to benefit from D2D connectivity and reduce overhead on existing wireless infrastructure.

4.3. Mobile Solution: Case Study Validation. In order to validate the functionality of the developed mobile solution, we present in this section an example scenario using six Android smartphones to download a 5000 KB file; all devices are running the mobile application based on the architecture explained in Section 4.2. Five devices serve as potential

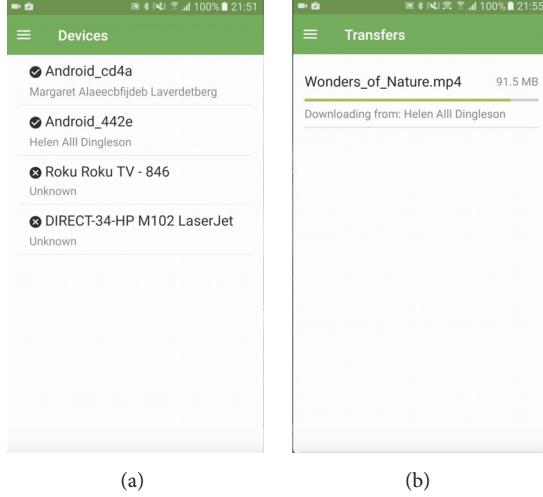


FIGURE 14: (a) List of discovered devices; (b) file transfer in progress.

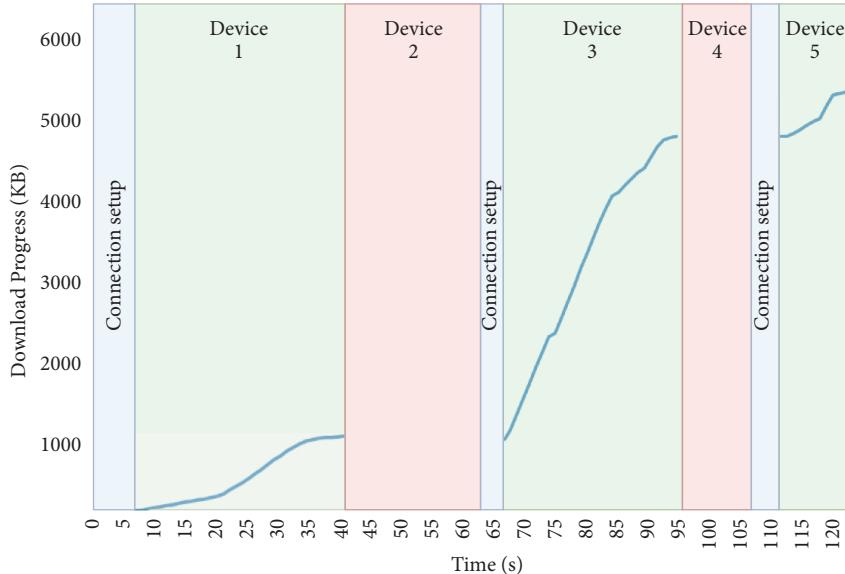


FIGURE 15: Example social-aware D2D data sharing case study using Android smartphones.

collaborators for D2D data sharing and are positioned equidistantly along a hallway while in file server mode (see description of File Server Service in Figure 11). The sixth device is serving as a data requester (file fetch mode) and is moving along the hallway and, thus, crossing the radio range boundaries of the five devices sequentially. Figure 15 presents real measurement results showing the download progress versus time. In this scenario, it is assumed that all five devices have the content available but only devices 1, 3, and 5 have a trusted social network relation with the requesting device. As the mobile solution is social-aware, devices 2 and 4 were excluded from the cooperation; that is, they will be filtered out after being discovered by the requesting device.

Figure 15 demonstrates how the download process progresses as the requesting device moves along the hallway; around 1000 KB were downloaded from device 1, next

3500 KB downloaded from device 3, and last 500 KB downloaded from device 5. The figure also quantifies the time spent by the device to perform device discovery and connection setup before it can proceed with D2D data transfer; this time is shown to be around five seconds on average and it adds to the total delay overhead of D2D mobile solutions in practice. The total time needed to download the content ended up to be around 125 seconds. We conducted similar experiments for other configurations and parameters, and the obtained results provided coherent insights and validated the robustness and effectiveness of our implementation under practical operational constraints.

5. Conclusion

In this work, we addressed several aspects related to the practical design and implementation of direct D2D data

sharing mobile solutions without relying on intelligence or coordination from any existing radio access network. We performed an experimental study using our own developed crowdsourcing Android application CARMA in order to collect data and derive empirical models for several user mobility and contact parameters. To this end, we managed two data collection stages, the first including 37 mobile user participants over a period of 11 weeks and the second including 13 mobile user participants over a period of 4 weeks. Our focus was on use cases with bounded geographical location (university campus) and homogenous community (university students). We also designed and managed a user survey after obtaining the needed IRB (Institutional Research Board) approval in order to collect data related to content stored on users' mobile devices and their willingness to participate in D2D data sharing applications. Then, we collected data on the social network relationships among the users and presented a set of conditions necessary for successful D2D file sharing based on joint consideration and correlation of mobility parameters (e.g., number of contacts and contact duration per time period), content similarity level, and social network relationship strength. Finally, we presented an architecture, implementation, and validation of a mobile solution for social-aware D2D data sharing using Android mobile application with several back-end module and a friendly user interface.

Even though the conclusions derived in the course of our work are specific to a university environment, they provide valuable insights that can guide the development of D2D data sharing solutions. The research outcomes illustrate clearly the dynamic nature of inter-user interactions whether in terms of contact and mobility, social connectivity, or mobile content. They also shed light on the temporal variability of all these factors (e.g., by comparing summer versus spring or comparing different days within the week) even by a homogenous group of users in a bounded geographical area, which somehow mimics a best case scenario when compared to dynamic variations over the scale of a city or country. Regarding the number of participants in our crowdsourcing experiments, it is similar to other interesting studies reported in the literature, for example, [19, 20]. Finally, all the collected data traces will be added as supplementary material with this paper to be openly accessible by the research community, as they can be utilized as inputs for evaluating the performance of new proposed D2D related enhancement techniques. In addition, the data traces can be used by researchers interested in predictive analytics using big data techniques, whereby devices or some centralized intelligence unit can predict future performance based on historical data in order to optimize the D2D offloading experience.

Disclosure

The statements made herein are solely the responsibility of the authors.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was made possible by Qatar National Research Fund (a member of the Qatar Foundation), NPRP Grant 7-1529-2-555.

References

- [1] P. Gandomi and R. K. Jha, "Device-to-Device Communication in Cellular Networks: A Survey," *Journal of Network and Computer Applications*, vol. 71, pp. 99–117, 2016.
- [2] F. Rebecchi, M. Dias de Amorim, V. Conan, A. Passarella, R. Bruno, and M. Conti, "Data offloading techniques in cellular networks: a survey," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 580–603, 2015.
- [3] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 4, pp. 1801–1819, 2014.
- [4] A. Detti, B. Ricci, and N. Blefari-Melazzi, "Mobile peer-to-peer video streaming over information-centric networks," *Computer Networks*, vol. 81, pp. 272–288, 2015.
- [5] N. Golrezaei, P. Mansourifard, A. F. Molisch, and A. G. Dimakis, "Base-station assisted device-to-device communications for high-throughput wireless video networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 7, pp. 3665–3676, 2014.
- [6] K. Jahed, S. Sharafeddine, A. Moussawi et al., "Scalable multimedia streaming in wireless networks with device-to-device cooperation," in *Proceedings of the 24th ACM Multimedia Conference*, (MM '16), pp. 762–764, October 2016.
- [7] S. Mumtaz, H. Lundqvist, K. M. S. Huq, J. Rodriguez, and A. Radwan, "Smart Direct-LTE communication: An energy saving perspective," *Ad Hoc Networks*, vol. 13, pp. 296–311, 2014.
- [8] E. Yaacoub, Z. Dawy, S. Sharafeddine, and A. Abu-Dayya, "Joint energy-distortion aware algorithms for cooperative video streaming over LTE networks," *Signal Processing: Image Communication*, vol. 28, no. 9, pp. 1114–1131, 2013.
- [9] L. Al-Kanj, Z. Dawy, and E. Yaacoub, "Energy-aware cooperative content distribution over wireless networks: design alternatives and implementation aspects," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 1736–1760, 2013.
- [10] K. Jahed, M. Younes, and S. Sharafeddine, "Energy measurements for mobile cooperative video streaming," in *Proceedings of the 2012 IFIP Wireless Days*, (WD '12), November 2012.
- [11] Y. Zhang, E. Pan, L. Song, W. Saad, Z. Dawy, and Z. Han, "Social network aware device-to-device communication in wireless networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 1, pp. 177–190, 2015.
- [12] Y. Zhang, L. Song, W. Saad, Z. Dawy, and Z. Han, "Contract-Based Incentive Mechanisms for Device-to-Device Communications in Cellular Networks," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 10, pp. 2144–2155, 2015.
- [13] M.-F. Leung and S.-H. G. Chan, "Broadcast-based peer-to-peer collaborative video streaming among mobiles," *IEEE Transactions on Broadcasting*, vol. 53, no. 1, pp. 350–361, 2007.
- [14] S. Andreev, J. Hosek, T. Olsson et al., "A unifying perspective on proximity-based cellular-assisted mobile social networking," *IEEE Communications Magazine*, vol. 54, no. 4, pp. 108–116, 2016.
- [15] Y. Meng, C. Jiang, H.-H. Chen, and Y. Ren, "Cooperative Device-to-Device Communications: Social Networking Perspectives," *IEEE Network*, vol. 31, no. 3, pp. 38–44, 2017.

- [16] Y. Zhang, M. Pan, L. Song, Z. Dawy, and Z. Han, "A Survey of Contract Theory-Based Incentive Mechanism Design in Wireless Networks," *IEEE Wireless Communications Magazine*, vol. 24, no. 3, pp. 80–85, 2017.
- [17] M. Tehrani, M. Uysal, and H. Yanikomeroglu, "Device-to-device communication in 5G cellular networks: challenges, solutions, and future directions," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 86–92, 2014.
- [18] D. Camps-Mur, A. Garcia-Saavedra, and P. Serrano, "Device-to-device communications with WiFi direct: overview and experimentation," *IEEE Wireless Communications Magazine*, vol. 20, no. 3, pp. 96–104, 2013.
- [19] E. Talipov, Y. Chon, and H. Cha, "Content sharing over smartphone-based delay-tolerant networks," *IEEE Transactions on Mobile Computing*, vol. 12, no. 3, pp. 581–595, 2013.
- [20] R.-I. Ciobanu, R.-C. Marin, and C. Dobre, "Interaction predictability of opportunistic networks in academic environments," *Transactions on Emerging Telecommunications Technologies*, vol. 25, no. 8, pp. 852–864, 2014.
- [21] R.-C. Marin, C. Dobre, and F. Xhafa, "A methodology for assessing the predictable behaviour of mobile users in wireless networks," *Concurrency and Computation: Practice and Experience*, vol. 26, no. 5, pp. 1215–1230, 2014.
- [22] T. Karagiannis, J.-Y. Le Boudec, and M. Vojnović, "Power law and exponential decay of intercontact times between mobile devices," *IEEE Transactions on Mobile Computing*, vol. 9, no. 10, pp. 1377–1390, 2010.
- [23] A. Tatar, T. Phe-Neau, M. Dias De Amorim, V. Conan, and S. Fdida, "Beyond contact predictions in mobile opportunistic networks," in *Proceedings of the 11th Annual Conference on Wireless On-Demand Network Systems and Services*, (IEEE/IFIP WONS '14), pp. 65–72, April 2014.
- [24] M. Cha, H. Kwak, P. Rodriguez, Y. Y. Ahnt, and S. Moon, "I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system," in *Proceedings of the ACM SIGCOMM Internet Measurement Conference*, pp. 1–14, October 2007.
- [25] S. Chandra and X. Yu, "An empirical analysis of serendipitous media sharing among campus-wide wireless users," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 7, no. 1, article no. 6, 2011.
- [26] F. Le Fessant, S. Handurukande, A. M. Kermarrec, and L. Massoulié, "Clustering in Peer-to-Peer File Sharing Workloads," in *Peer-to-Peer Systems III*, vol. 3279 of *Lecture Notes in Computer Science*, pp. 217–226, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
- [27] Y. Li, T. Wu, P. Hui, D. Jin, and S. Chen, "Social-aware D2D communications: qualitative insights and quantitative analysis," *IEEE Communications Magazine*, vol. 52, no. 6, pp. 150–158, 2014.
- [28] Y. Xiao, D. Niyato, K.-C. Chen, and Z. Han, "Enhance device-to-device communication with social awareness: A belief-based stable marriage game framework," *IEEE Wireless Communications Magazine*, vol. 23, no. 4, pp. 36–44, 2016.
- [29] X. Wang, Z. Sheng, S. Yang, and V. C. M. Leung, "Tag-assisted social-aware opportunistic device-to-device sharing for traffic offloading in mobile social networks," *IEEE Wireless Communications Magazine*, vol. 23, no. 4, pp. 60–67, 2016.
- [30] Y. Meng, C. Jiang, Q. D. La, T. Q. S. Quek, and Y. Ren, "Dynamic social-aware peer selection scheme for cooperative device-to-device communications," in *Proceedings of the 2017 IEEE Wireless Communications and Networking Conference (WCNC '17)*, March 2017.
- [31] D. Wu, J. Yan, H. Wang, D. Wu, and R. Wang, "Social Attribute Aware Incentive Mechanism for Device-to-Device Video Distribution," *IEEE Transactions on Multimedia*, vol. 19, no. 8, pp. 1908–1920, 2017.
- [32] Y. Zhao, W. Song, and Z. Han, "Social-Aware data dissemination via device-to-device communications: Fusing social and mobile networks with incentive constraints," *IEEE Transactions on Services Computing*, vol. PP, no. 99, 2017.

