Propagation- and Mobility-Aware D2D Social Content Replication

Zhi Wang, Lifeng Sun, *Member, IEEE*, Miao Zhang, Haitian Pang, Erfang Tian, and Wenwu Zhu

Abstract—Mobile online social network services have seen rapid expansion; thus, the corresponding huge amounts of user-generated social media contents propagating between users via social connections have significantly challenged the traditional content delivery paradigm. First, replicating all the contents generated by users to edge servers that well "fit" the receivers becomes difficult due to limited bandwidth and storage capacities. Motivated by device-to-device (D2D) communication, which allows users with smart devices to transfer content directly, we propose replicating bandwidth-intensive social contents in a device-to-device manner. Based on large-scale measurement studies on social content propagation and user mobility patterns in edge-network regions, we observe the following: (1) Device-to-device replication can significantly help users download social contents from neighboring peers. (2) Both social propagation and mobility patterns affect how contents should be replicated. (3) The replication strategies depend on regional characteristics (e.g., how users move across regions). Using these measurement insights, we propose a *propagation- and mobility-aware* content replication strategy for edge-network regions, in which social contents are assigned to users in edge-network regions according to a joint consideration of social graphs, content propagation, and user mobility. We formulate the replication scheduling as an optimization problem and design a distributed algorithm using only historical, local, and partial information to solve it. Trace-driven experiments further verify the superiority of our proposal: compared with conventional pure-movement-based and popularity-based approaches, our design can significantly improve ($2-4$-fold improvement) the amount of social content successfully delivered via device-to-device replication.

Index Terms—Device-to-device communication, social propagation, mobility behavior, measurement, algorithm

1 Introduction

Mobile social network services based on the convergence of wireless networks, smart devices, and online social networks have witnessed rapid expansion in recent years [1]. According to YouTube, over 100 hours worth of videos have been produced by individuals and shared among themselves, and the traffic resulting from delivering these content items to mobile devices has exceeded 50 percent [2], significantly challenging the traditional content delivery paradigm, in which content is replicated by a hierarchical infrastructure using the same scheme [3]. It is usually expansive and inefficient to replicate the massive number of social content items to traditional CDN servers [4].

With the development of *device-to-device* (D2D) communication [5], it has become promising to offload the bandwidth-intensive social content delivery to users' mobile devices and let them serve each other. Previous studies have demonstrated that such device-to-device content sharing is possible when users are close to each other and when the content to be delivered is delay tolerant [6]. In this paper, we use *mobile edge networks* (or edge networks for short) to define the local area whereby users move across regions and can directly communicate with each other. It is intriguing to investigate content delivery strategies in the context of edge networks because both users' behaviors and network properties have to be studied.

In traditional device-to-device content sharing, a user typically sends the generated content to a set of users that are close to the user in a broadcast-like manner, therein causing the following problems: (1) Due to the broadcasting mechanism, users' devices have to expend high amounts of power to cache and forward many content items in the edge network. As the number of user-generated social content items increases, such a mechanism becomes inherently inscalable. (2) Social content—due to the dynamical social propagation—has heterogeneous popularity, whereas conventional approaches treat all such content the same, resulting in wasted resources to replicate unpopular content items. (3) Due to the dynamic mobility patterns, it is difficult to guarantee any quality of experience.

To address these problems, we propose a joint propagation- and mobility-aware replication strategy based on social propagation characteristics and user mobility patterns in the edge-network *regions*, e.g., 100×100 m^2 areas that users can move across and deliver content to other users.

The idea behind our proposal is as follows. (1) Instead of letting content flood between users that are merely close to one another, we propose to replicate social content according to the social influence of users and the social propagation of content. (2) We develop a regional social popularity prediction model that captures the popularity of content items based on both regional and social information. (3) We

- *Z. Wang and M. Zhang are with the Graduate School, Tsinghua University, Shenzhen 518055, China. E-mail: wangzhi@sz.tsinghua.edu.cn,.*
- *L. Sun, H. Pang, E. Tian, and W. Zhu are with the Department of Computer Science and Technology, Tsinghua University, Tsinghua National Laboratory for Information Science and Technology, Beijing Key Laboratory of Networked Multimedia, Beijing 100084, China. E-mail: {sunlf, wwzhu}@tsinghua.edu.cn.*

Manuscript received 19 Aug. 2015; revised 31 May 2016; accepted 3 June 2016. Date of publication 20 June 2016; date of current version 2 Mar. 2017. 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.2016.2582159

propose to replicate social content items according to not only regional social popularity but also user mobility patterns, which capture how users move across and remain in these edge-network regions.

In our proposal, we face the following challenges: How does one capture the joint propagation and mobility behaviors? How does one identify the parameters that affect the performance of mobile social content replication? How does one design efficient strategies/algorithms for our proposal that work in real world? Our contributions represent a set of answers to address these challenges.

⊳ Based on large-scale measurement studies, including $450,000$ content items shared by $240,000$ users on an online social network, and on $300,000$ users moving across hundreds of edge-network regions, we reveal the possibility of device-to-device replication for social content and the design principles for utilizing both social propagation and user mobility patterns. We present a number of measurement insights, including how social propagation and mobility patterns affect D2D content replication.

⊳ Based on our measurement studies, we build social propagation and user mobility predictive models to capture the popularity distribution of content in different regions. Using the predictive models, we then formulate the D2D content replication as an optimization problem, which is inherently centralized. We then design a heuristic algorithm to practically solve this problem in a distributed manner, therein only requiring historical, local and partial information.

⊳ We use both model-driven and trace-driven experiments to verify the effectiveness of our design: Compared to traditional approaches, our design can significantly enhance the chance for users to download content from nearby edge-network devices. In particular, our design improves the D2D delivery fraction by 4 times compared to a pure movement-based approach and by 2 times compared to a pure popularity-based approach. Based on our model-driven experiments, we also present the limitations of such a D2D delivery approach.

The remainder of the paper is organized as follows. We survey related works in Section 2. We give the motivation of our design in Section 3. Using large-scale measurement studies, we present the principles for our design in Section 4. In Section 5, we present the details of our design based on mobility and propagation predictive models. We evaluate our design in Section 6. Finally, we conclude the paper in Section 7.

2 Related Works

We survey the literature on social propagation, social content distribution, D2D content delivery, as well as user mobility characteristics.

2.1 Social Propagation in Online Social Networks

Online social network has greatly changed content delivery, e.g., the distribution of social contents has shifted from a "central-edge" manner to an "edge-edge" manner. Bakshy et al. [7] studied the social influence of individuals in an online social network and observed that some users can be very influential in social propagation. Li et al. [8] studied content sharing in an online social network and observed the skewed popularity distribution of contents and the *power-law* activity of users. Comarela et al. [9] investigated the response time of social contents using collected traces and confirmed the nature of the *delay tolerance* of social media, which motivated our study.

Because online social networks are facilitating the *dissemination* of all types of online contents, conventional content delivery paradigms requirement improvement based on social information. Pujol et al. [10] designed a social partition and replication middleware wherein users' friends' data can be co-located in data center servers. Scellato et al. [11] investigated using social *cascading* information for content delivery over edge networks. Wang et al. [12], [13] investigated the possibility to infer social propagation according to users' social profiles and behaviors and allocate network resources at edge-cloud servers based on propagation predictions. Wen et al. [14] further proposed the cloud mobile media concept to utilize cloud-based resources for mobile media content processing and distribution. Wang et al. [15] proposed a novel peer-assisted paradigm using social relationships to achieve improved social media distribution.

The limitation of previous studies is that they were focused on social content delivery using server-based hierarchical infrastructure. Our study will investigate how bandwidth-intensive social contents can be distributed via D2D resources by jointly inferring social propagation and user mobility.

2.2 Mobility Characteristics

Understanding the mobility of users is key to designing effective delivery strategies for mobile social contents. Kim et al. [16] proposed to use traces of users' associations with Wi-Fi access points to investigate how users move among popular locations. Based on user mobility models, Yoon et al. [17] found that it is possible to generate movement patterns that are statistically similar to real movements. Karamshuk et al. [18] surveyed the usage of spatial, temporal and social properties to capture the mobility behaviors.

Rhee et al. [19] then noted that human movements are not random walks and that the patterns characterizing human walks and *Levy* walks contain some statistical similarity; in particular, features including heavy-tail flight and the super diffusive nature of mobility are observed. In [20], the power-law and exponential decay relations characterizing inter-contact times between mobile devices are observed. Zhuang et al. [21] studied the mobility and encountering patterns for users in regions during particular events, e.g., conferences.

By jointly studying the mobility and social network of users, Cho et al. [22] observed that, although human movement and mobility patterns have a high degree of variation, they exhibit structural patterns due to geographic and social constraints. In particular, short-ranged travel is periodic both spatially and temporally and is not affected by the social network structure. Recently, such mobility studies have improved the edge network and content delivery design. Wang et al. [23] studied the mobility characteristics of people to guide wireless network deployment. Wang et al. [24] found a similarity between individuals' mobility patterns and their social proximities.

These studies have focused on human mobility characteristics from a general perspective, i.e., how people move in their daily lives. Our study will particularly focus on how users move in edge-network regions, e.g., regions associated with Wi-Fi access points, where users can serve as peers for D2D content delivery.

2.3 Message Forwarding in Delay-Tolerant Networks

The delay-tolerant network architecture and application interface were proposed to structure around optionally reliable asynchronous message forwarding [25], with limited expectations of end-to-end connectivity and node resources. Since then, many efforts have been devoted to efficient message routing and forwarding in this paradigm. Jain et al. investigated the routing problem in such delay-tolerant networks and demonstrated that algorithms using the least knowledge tend to perform poorly, whereas with limited additional knowledge, namely far less than complete global knowledge, efficient algorithms can be constructed for routing in such environments [26]. Daly et al. studied the small-world dynamics for characterizing information propagation in wireless networks and also confirmed that using local information is promising for message routing in DTNs [27]. Helgason et al. [28] developed an opportunistic framework in which content is divided into different topics for lookup and forwarding. Haillot et al. [29] proposed a content-based communication scheme wherein users can subscribe to content categories according to their preferences and content is disseminated based on the subscription. Hui et al. [30] studied the patterns of contact in pocket-switched networks and exploited two social and structural metrics, centrality and community, using real human mobility traces. Karamshuk et al. [18] surveyed the usage of spatial, temporal and social properties to capture the mobility behaviors and utilized mobility models for opportunistic networks.

In these studies, the concept of a social network is used as the real-world contact graph, i.e., the contact of people forms a social network. However, today's online social networks and the user contact graphs can be highly independent on each other because users are allowed to interact with any users, who are not necessarily nearby, thanks to the wireless network and mobile devices. Our study attempts to investigate how social propagation and mobility patterns can be jointly utilized to improve content replication in edge networks and the limitations of such D2D replication for social media content.

2.4 Mobile Social Content Delivery

As soon as Mobile Ad hoc Networks (MANETs) were proposed, researchers started to envision using device-to-device communications for content delivery. Pelusi et al. [31] studied opportunistic networking and discussed possible scenarios of its use. In the context of using smart devices with limited energy for content delivery, Li et al. [32] formulated the optimization problem of opportunistic forwarding, with a constraint concerning the energy consumed by message delivery for both two-hop and epidemic forwarding. In Han et al.'s study [33], they investigated the selection of a target set of users for content deployment for minimizing the mobile data traffic over cellular networks.

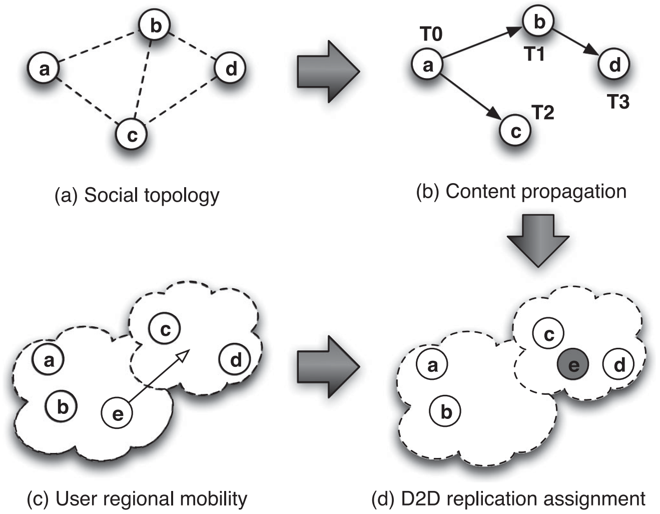


Fig. 1. D2D replication affected by social topology, content propagation, and user regional mobility.

Recently, Bao et al. [34] explored the possibility of serving user requests from other mobile devices located geographically close to the user, where the cellular operator learns users' content requests and guides smart devices to other devices that have the requested content. Wang et al. [6] investigated the problem of offloading social media delivery based on a joint online social network and mobile social network framework. Recently, a number of contextual factors have also been investigated for new content delivery paradigms [35].

In previous studies, it was assumed that users directly help each other in their *trajectories*, which significantly limits the ability of D2D delivery possibilities. In our study, we propose to replicate social contents for regional social propagation, i.e., users are supposed to serve edge-network regions where they will move to and stay, based on joint social propagation and user mobility characteristics such that edge-network regions will be covered by peering users according to their regional social popularities.

3 Motivation and Assumption

3.1 Motivation

In an online social network, users share contents with their friends through the social connections. The propagation of contents is determined by both social graphs and user behaviors. Because such social graphs and user behaviors are inside the online social network, they can be independent of users' mobility patterns in the physical world. For example, a user can intensively interact with their friends online without having to be at the same location thanks to the online social network.

As illustrated in Fig. 1, based on the social graph and propagation patterns, we estimate how contents will be received by users; based on the regional mobility, we will then predict which regions users will be moving to and how long they will stay. Simultaneously, we decide which users will replicate which social contents on the move. In this example, user e – while not a friend of any other user—

is moving to the region where user c and d are located. Thus, e will be selected to replicate the content generated by user a, and both user c and user d will receive the content shared by user a in the social propagation at times $T2$ and $T3$, respectively.

In our study, we propose to jointly utilize the social graph, user behaviors and user mobility patterns for D2D replication to serve edge-network regions. To this end, we first study the propagation of mobile social contents, user mobility patterns, and characteristics of edge-network regions. Based on the measurement insights, we then propose a propagation- and mobility-aware D2D replication for edge-network regions.

3.2 Assumption

In our proposal, we decouple the content replication for the social propagation on the online social network from users' mobility in the physical world, i.e., a user may cache content and deliver the content to other users who are not socially connected to the user. Note that, although our replication is decoupled from the mobility, the social propagation and mobility patterns do not have to be independent, e.g., users may share more content at some locations compared to other location. In our design, we present an algorithm to adaptively select content for users to cache. We will also verify the impact of the correlation between the social propagation and user mobility in our experiments.

4 Measurement Studies

We conduct large-scale measurement studies on social propagation, user mobility and edge-network regions.

4.1 Measurement Setup

We use a data-driven approach for the measurement study using the following valuable datasets provided by our partners.

4.1.1 Traces of Mobile Social Content Propagation

We have obtained content upload and request traces from our partnership with Tencent Weishi, a mobile social network service. In Weishi, short videos are generated by individuals and shared with their friends. Our dataset records $240,000$ users sharing over $450,000$ videos in 2014 with the following information: (1) Content generation, which records when a video is generated and shared by a particular user; (2) Content download, which records which videos are downloaded by which users; (3) Social graph, which records how users are socially connected to each other; and (4) Sharing information, which records when content is shared by users, including the ID, name, IP address of the users, time stamp corresponding to when content is shared, and the IDs of the parent and root users if it is a reshare.

4.1.2 Traces of Edge-Network Region and User Mobility

To investigate how users move across edge-network regions, we use traces provided by NextWiFi, a local Wi-Fi provider. NextWiFi sampled over $300,000$ users associating to hundreds of Wi-Fi access points in a shopping mall in 2014. The information NextWiFi recorded includes (1) The timestamp representing when users connect to Wi-Fi access points along with the duration of the AP *association* and (2) The BSSID, Service Set Identification (SSID) and locations of these access points.

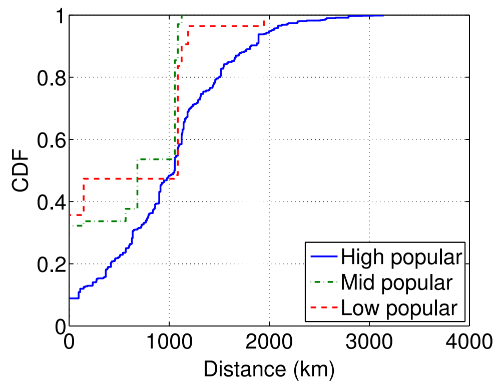


Fig. 2. CDF of distance between users that are sharing in the same propagation of social content items.

Based on these datasets,[1] we first study how social content items are shared by users moving across edge-network regions, and the measurement studies will reveal the possibility of D2D social content replication. Then, we study users' mobility patterns, which reveal the principles used to design the replication strategies.

4.2 Propagation of Social Content

4.2.1 Propagation in Local Regions

We first investigate the possibility of D2D content delivery for social content based on propagation traces. In Fig. 2, we plot the CDF of distances between users who join the social propagation of the same content in the online social network. Different curves are for content items with different popularities, i.e., the number of users who receive the content. We observe that, in contrast to popular content that is shared by users randomly located in different places, unpopular content is more likely to be shared in *local* regions, where users are closer to each other. For example, approximately 40 percent of the distances between users sharing unpopular content (low popular) are close to 0 km, indicating that users may be in the same edge-network regions.

This observation indicates that, in the propagation of such unpopular content, which represents the majority of social content [36], users are likely to be able to receive content through a D2D scheme. Notice that these distances are between users sharing the same social content; in our design, we also allow users who are in different content propagations to deliver content for each other, which can further reduce the distance between users for D2D delivery.

4.2.2 Delay Tolerance of Receiving Social Content

D2D content delivery depends on users who are moving across edge-network regions, e.g., a user can carry a social content from one region and then serve it in another region

1. The datasets used in our study are available at http://mmlab.top:8080/social-mobility-data/

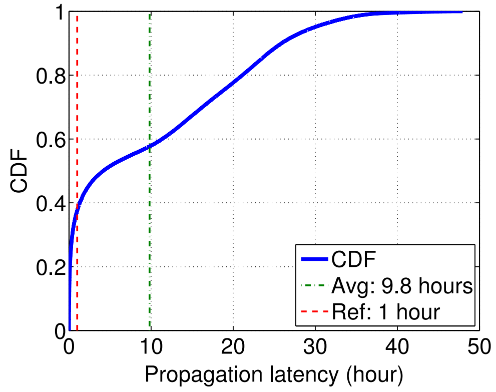


Fig. 3. CDF of propagation latency.

when the user moves there. As a result, D2D content delivery itself may not be able to guarantee a latency to successfully pass a content due to the dynamics of user mobility because the content has to wait for users to encounter each other to be delivered.

Thus, we have to investigate whether this technique is feasible for today's social content delivery from the perspective of the latency tolerance of users. We study *propagation latency*, which defines the time elapsed between the time a content is shared by a person and the time that the content is reshared or viewed by another user. D2D delivery is supposed to be performed within this propagation latency.

As illustrated in Fig. 3, the curve represents the CDF of the propagation latencies of social content shared in our traces. In particular, the average propagation latency is approximately 9.8 hours, and over 62 percent (*resp.*, 69 percent) of the propagation latencies are longer than 1 hour (*resp.*, 30 minutes). These observations indicate that, in social propagation, social content tends to be delay tolerant, allowing us to design D2D replication for content delivery.

4.3 Characteristics of Edge-Network Regions

Next, we study the characteristics of edge-network regions, where users are supposed to replicate social content and serve others in our design.

4.3.1 Popularity of Edge-Network Regions

We first study the popularity of edge-network regions in terms of the number of users in different regions. Several previous works assumed user movement as random walks [19]; however, this is not true for users' daily mobility patterns. In our study, we investigate the popularity of different edge regions, which captures how many users, who can be either requesting users or D2D peers, are currently in these regions.

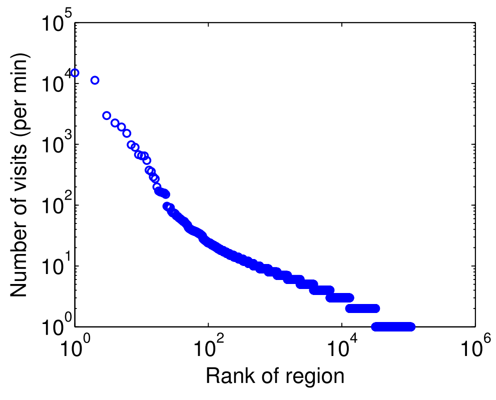


Fig. 4. Number of user visits to edge-network regions.

As illustrated in Fig. 4, each sample is the number of visits per minute versus the rank of regions (in descending order of the number of visits). We observe that the popularity roughly follows a *zipf*-like distribution, indicating that there are a few regions that can attract substantially more users than others. In our design, this observation is considered so that a region with a higher "load" (i.e., the number of requests that can potentially be issued from a region) will be assigned more D2D peers to replicate the social content.

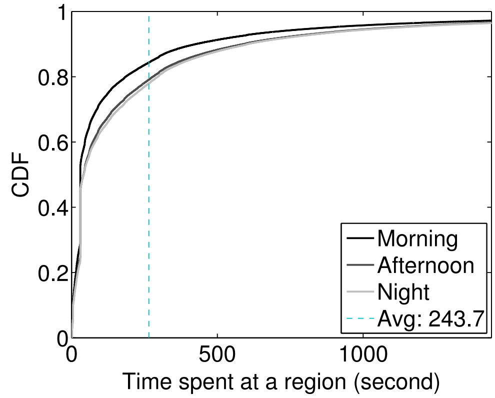


Fig. 5. CDF of duration of user association in edge-network regions.

4.3.2 Association Duration in Different Regions

After leaving an edge-network region, a user will no longer be able to serve users in that region. Thus, it is also important to investigate how long users can stay in a region. From the Wi-Fi traces, we have sampled 5 million Wi-Fi association records spanning 10 days to study the association duration of users in the regions they have visited. In Fig. 5, we plot the CDF of the time that users stay in regions they visited at different hours of a day. The average duration for users to stay in a region is approximately 4 minutes, and we also observe that users tend to spend less time in the same region in the morning than during other hours of the day. These observations indicate that there remain many users that only instantly pass edge-network regions and that are most likely unable to serve as stable replication peers. In our design, we take how long users stay in different regions into account.

4.4 Regional Mobility Patterns

Finally, we study users' mobility patterns in edge-network regions to guide our D2D replication design.

4.4.1 Revisits to the Same Regions

We investigate the possibility for users to revisit the same region many times because this helps us to design social content replication strategies according to users' periodical appearance patterns. We study how users revisit the same region using the Wi-Fi traces that we collected. In Fig. 6, the curves are the CDFs of the number of user revisits to the same region within a time span of 12 hours, 24 hours and 48 hours. We observe that over 30 percent of users' visits occur to the same region over 2 times. This observation indicates

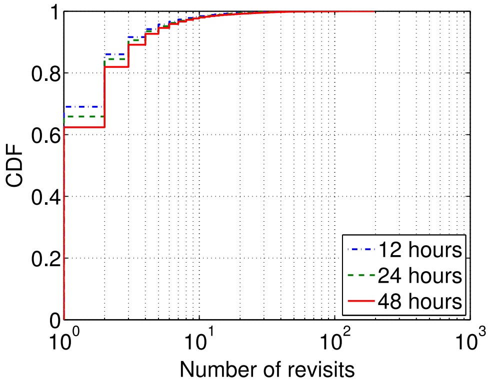


Fig. 6. CDF of the number of user visits to the same region.

that users are actually not randomly walking across edge-network regions; instead, they have inherent preferences for certain regions. As a result, replicating social content should take this region preference into consideration.

4.4.2 *Migration Patterns between Regions*

Furthermore, we investigate how users migrate between different regions. For each region pair $(r1, r2)$, we calculate a migration number of users who connect to Wi-Fi hotspots in $r1$ and $r2$, respectively, in two consecutive associations. Without loss of generality, a user associates with APs in the same region if $r1 = r2$. We then rank the region pairs in descending order of the migration number. In Fig. 7, we plot the migration number versus the rank of sorted region pairs. Notice that users can associate with the same Wi-Fi access point many times; thus, there are "migrations" staying in the same region. We observe that this number roughly follows a power-law distribution, indicating that there are a few pairs of regions that users tend to move across.

This observation indicates that edge-network regions are actually (e.g., geographically) correlated with each other. Based on this observation, we propose to predict which region a user may move to in the future according to the migration patterns in our design.

5 Detailed Design: Propagation- and Mobility-Aware D2D Regional Replication

Motivated by the measurement insights, we design a joint propagation- and mobility-aware replication strategy for

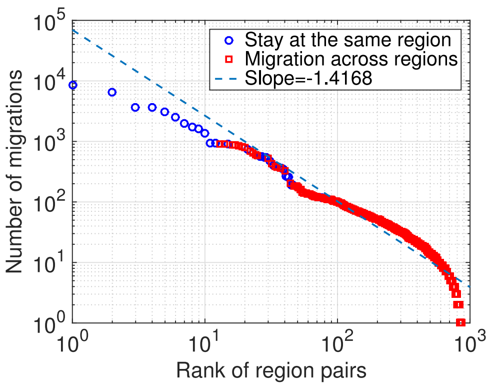


Fig. 7. Number of user migrations across regions.

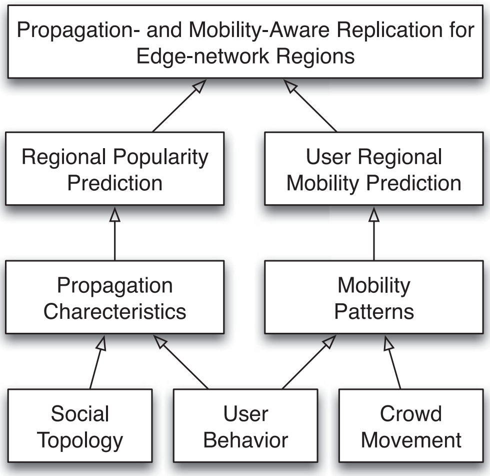


Fig. 8. Framework of propagation- and mobility-aware D2D replication for social content.

D2D social content delivery. The framework of our design is illustrated in Fig. 8. We design models to capture the propagation patterns of social content and the regional popularity and user mobility across regions. Based on these models, we predict if a content will be highly requested due to social propagation in a particular region, and we predict where a user may visit in the future. Using these predictions, we assign users to replicate social content for D2D delivery.

Before we present the details, we give some important notations in Table 1.

5.1 Propagation-Based Regional Popularity

We attempt to characterize the popularity of social content. In particular, we seek to answer the following question: How many users in total are expected to download a social content in a particular edge-network region in the next time slot? Notice that we propose to schedule D2D replication strategies in a discrete time slot-based manner, and there are numerous studies on converting such strategies to real implementations [37]. The length of a time slot depends on the observed average propagation latency and the duration that users are staying in the regions.

5.1.1 *Social Influence of Users*

We propose to build an *influence* estimation between users for regional propagation prediction. Let $I_{u,v}^{(T)}$ represent a social influence index from user u to v, which captures the ability for user u to attract user v to receive content shared by user u in time slot T. $I_{u,v}^{(T)}$ is affected by many factors [38], including the content itself, the preference of users, and the context under which the content is shared.

In our design, we use a data-driven approach to estimate the influence index by considering the historical statistics as follows:

$$I_{u,v}^{(T)} = \frac{\sum_{t=T-W}^{T-1} H_{u,v}^{(t)}}{\sum_{t=T-W}^{T-1} G_{u}^{(t)}}, \tag{1}$$

where $H_{u,v}^{(t)}$ is the number of social content items that are shared by user u and received (accepted) by user v in time slot t, and $G_u^{(t)}$ is the number of content items generated and reshared by user u in time slot t. W is a time window that we are referring to for the statistical data and is chosen to be sufficiently large to contain the content shared by user u. The rationale is that we use users' social influence recorded in previous time slots to estimate to what extent they attract others in the future. A larger $I_{u,v}^{(T)}$ indicates that user v is more likely to accept content shared by user u. If user u has a large influence index to many friends, the popularity affected by social propagation then cannot be ignored.

5.1.2 User Regional Preference

According to our measurement studies in Section 4.4, users have different preferences for different edge-network regions, thereby reflecting possibilities for where they stay. We investigate users' inherent preferences for different regions. Suppose that a user has a preference $\mathcal{P}_{u,r}^{(T)}$ for a region r. We estimate $\mathcal{P}_{u,r}^{(T)}$ using the records of how long users are staying in different regions in a previous time window W' as follows:

$$\mathcal{P}_{u,r}^{(T)} = \frac{d_{u,r}^{(T-W',T-1)}}{\sum_{r' \in \mathcal{R}_u^{(T)}} d_{u,r'}^{(T-W',T-1)}}, \tag{2}$$

where $d_{u,r}^{(T-W',T-1)}$ is the duration user u has stayed in region r in the previous time window $[T-W', T-1]$, $\mathcal{R}_u^{(T)}$ is the set of regions that user u has visited previously, and W' is the time window used to study users' regional preferences. In our design, W' is chosen as a weekly basis, i.e., the previous week is referred to for estimation.

5.1.3 Regional Social Popularity

We study the regional social popularity based on both the user influence and users' region preferences.

▻ *Inherent content popularity*. In our design, the inherent content popularity is determined by the content itself. Some content items tend to be more interesting and can attract greater attention than other content items. We use $p_{c,r}^{(T)}$ to denote the inherent popularity of content c in region r in time slot T, which can be inferred using traditional popularity prediction approaches [39].

▻ *Influential popularity*. After a content is distributed over the social connections, its popularity is highly affected by the social networks [13]. We thus incorporate the social influence of users into our popularity inference.

Based on the inherent popularity and influential popularity, we design a *social popularity index* $A_{c,r}^{(T)}$, which reflects the popularity of content c in region r in time slot T, as follows:

$$A_{c,r}^{(T)} = p_{c,r}^{(T)} + \alpha \sum_{u \in \mathcal{S}(c)} \sum_{v \in \mathcal{F}_u} I_{u,v}^{(T)} \mathcal{P}_{v,r}^{(T)}, \tag{3}$$

where $\mathcal{S}(c)$ is the set of users who have shared content c and $\mathcal{F}_u$ is the set of users who are socially connected to user u, which may become the next resharers. According to our previous definitions, $\sum_{u \in \mathcal{S}(c)} \sum_{v \in \mathcal{F}_u} I_{u,v}^{(T)} \mathcal{P}_{v,r}^{(T)}$ reflects the popularity of the social influence in region r. $\alpha \in [0,1]$ is a running parameter used to determine how much the social influential popularity can contribute to the content requests. For example, some content items are only shared between friends, and the social influence thus dominates the popularity. On the other hand, some other content items are published by the central content provider, and their popularity is affected by the content itself [38].

α can be learnt using historical data: α is assigned a larger value for content whose popularity is more strongly influenced by the social propagation, i.e., the correlation between content popularity and social influence is strong. In our experiments, we use the fraction of the influenced users (i.e., users whose friends are already resharers of a content item) over all viewers to represent such an inference level of a particular content.

A content with a large $A_{c,r}^{(T)}$ is likely to attract more requests from region r. The rationale is that a content with a large social popularity index is either very popular in a region in the previous time window or has been shared by many influential people whose resharers are located in the region.

5.2 Regional Mobility Prediction

5.2.1 Regional Migration Model

We propose to let users cache content and serve as edge peers in local regions. To this end, we have to understand how users move across different regions. Based on our previous measurement studies, we observe that the edge regions not only have different popularities but also present different correlation levels between each other—more users tend to migrate between certain pairs of regions more so than other pairs of regions.

Let $\mathbf{E}_{r,s}^{(T)}$ denote the number of migrations of users who have moved from region r directly to region s in the previous time slot $T-1$ (i.e., only the recent previous time slot is referred to). We normalize $\mathbf{E}_{r,s}^{(T)}$ to $\bar{\mathbf{E}}_{r,s}^{(T)}$ as follows: $\bar{\mathbf{E}}_{r,s}^{(T)} = \frac{\mathbf{E}_{r,s}^{(T)}}{\sum_{r' | r' \in \mathcal{R}} \mathbf{E}_{r's}^{(T)}}$, where $\mathcal{R}$ is the set of all edge-network regions. In our later formulation of the replication strategy, we use this *migration index* $\bar{\mathbf{E}}_{r,s}^{(T)}$ as an optimization coefficient.

5.2.2 User Regional Mobility Prediction

Based on the user regional migration index and the regional preference of users, we then study a user's *mobility index* $\mathcal{Q}_{u,r}^{(T)}$ to capture the possibility for user u to visit region r in time slot T from its current region. The calculation of user mobility index is defined as follows:

$$\mathcal{Q}_{u,r}^{(T)} = \frac{\bar{\mathbf{E}}_{R_u^{(T-1)},r}^{(T)} \mathcal{P}_{u,r}^{(T)}}{\sum_{s \in \mathcal{R}} \bar{\mathbf{E}}_{R_u^{(T-1)},s}^{(T)} \mathcal{P}_{u,s}^{(T)}}, \tag{4}$$

where $R_u^{(T-1)}$ is the region where user u is in the previous time slot (i.e., $T-1$) and $\mathcal{P}_{u,r}^{(T)}$ is a user's region preference defined previously. The rationale of the user mobility index is that it integrates the user mobility statistics, which capture the correlations between regions as well as individuals' preferences of regions. A larger $\mathcal{Q}_{u,r}^{(T)}$ indicates that user u will be more likely to visit region r in the next time slot.

TABLE 1
Important Notations

Symbol	Definition
u, v	User index
r, s	Region index
$A_{c,r}^{(T)}$	Regional social popularity of content c in region r
$K_{u,c}^{(T)}$	Strategy variable indicating whether user u will carry c for D2D replication
$\mathcal{Q}_{u,r}^{(T)}$	Probability for user u to visit region r
$I_{u,v}^{(T)}$	A social influence index from user u to v
$H_{u,v}^{(t)}$	The number of content items that are shared by user u and received (accepted) by user v in time slot t
$G_{u}^{(t)}$	The number of contents posted by user u in time slot t
$\mathcal{P}_{u}^{(T)}$	A user's preference vector for different regions.
$\mathbf{E}_{r,s}^{(T)}$	Number of user migrations from region r to s.

5.3 Formulation and Analysis

We design a time-slot-based strategy for the D2D replication of social content. Let $\mathbf{K}^{(T)}$ denote the strategies users are going to apply for content replication in D2D social content delivery in time slot T. In $\mathbf{K}^{(T)}$, each entry $K_{u,c}^{(T)}$ is a strategy variable: $K_{u,c}^{(T)} = 1$ indicates that user u will *carry/cache* content c for D2D delivery in time slot T, and $K_{u,c}^{(T)} = 0$ otherwise.

Our objective is then to find an assignment for users to best match the regional popularities of the social content. To capture how much a strategy matches the current social popularity of content in the edge regions, we define $Y_c^{(T)}$ as a replication *gain* for social content c, under a given strategy $\mathbf{K}^{(T)}$, as follows:

$$Y_c^{(T)} = \sum_{u \in \mathcal{U}} \beta_u K_{u,c}^{(T)} \sum_{r \in \mathcal{R}} \mathcal{Q}_{u,r}^{(T)} A_{c,r}^{(T)}, \tag{5}$$

where $\mathcal{U}$ is the set of users who can perform social content replication and β_u is the upload capacity that user u can contribute in the D2D delivery. β_u is regarded as an altruism index for a user, and its value can be set according to the incentive mechanism of the system. For example, if the system is willing to pay users with credits if they contribute their upload, a large β_u is expected. The rationale of the D2D replication gain is that $Y_c^{(T)}$ reflects a *matching level* of users performing the replication for the predicted regional social popularities of the content items. A larger $Y_c^{(T)}$ indicates that requests are more likely to be successfully served by users.

The problem is then formulated as an optimization problem to find how contents generated and shared are assigned to potential D2D *peering* users so that the overall replication gain can be maximized. Our formulation is as follows:

$$\max_{\mathbf{K}^{(T)}} \sum_{c \in \mathcal{C}^{(T)}} Y_c^{(T)}, \tag{6}$$

$$\text{s.t.} \sum_{c \in \mathcal{C}^{(T)}} K_{u,c}^{(T)} \leq B_u, \forall u \in \mathcal{U}, \tag{7}$$

$$\sum_{u} \mathcal{Q}_{u,r}^{(T)} K_{u,c}^{(T)} \beta_u \leq A_{c,r}^{(T)}, \forall r \in \mathcal{R}, c \in \mathcal{C}^{(T)}, \tag{8}$$

$$\text{vars. } \mathbf{K}^{(T)},$$

where $\mathcal{C}^{(T)}$ is the set of content items generated or shared by users and B_u is the replication capacity of user u. Constraint (7) requires that replication of a user will not exceed its cache capacity. Constraint (8) requires that replication for a particular social content item will not exceed its popularity estimation.

5.4 Algorithm and Implementation

The problem is in nature difficult to solve and centralized; thus, we design a heuristic local algorithm to make a replication decision using local information.

5.4.1 *User Replication Algorithm*

Our solution for users to perform the propagation- and mobility-aware D2D replication is as follows. When a user is in an edge region, it is assigned to replicate content that will propagate over users that are nearby. Notice that the user itself may not be actually sharing in the propagation for which it replicates content. In our design, a user can replicate the social content either by directly receiving it from other users close to them or by receiving it from the content server. The user selects a subset of the candidate social content items to cache for D2D delivery only according to their local information.

Algorithm 1. D2D Content Replication

1: **procedure** Regional Content Replication
2: Let $\mathcal{Z} \leftarrow \mathcal{W}_u^{(T-1)} \bigcup \mathcal{C}_u^{(T)}$
3: **for all** content $c \in \mathcal{Z}$ **do**
4: *c.gain* $\leftarrow \sum_{r \in \mathcal{R}} \mathcal{Q}_{u,r}^{(T)} A_{c,r}^{(T)}$
5: **end for**
6: Let $\mathcal{W}_u^{(T)}$ contain at most B_u content items randomly selected from $\mathcal{Z}$, where each content has a probability to be chosen of $\frac{c.gain}{\sum_{c' \in \mathcal{Z}} c'.gain}$
7: User u caches content items in $\mathcal{W}_u^{(T)}$ and reports them to the coordinate server
8: **end procedure**

The details of the local replication strategy are illustrated in Algorithm 1. $\mathcal{W}_u^{(T-1)}$ denotes the set of content items user u has already replicated previously. $\mathcal{C}_u^{(T)}$ is defined as the set of candidate content items to replicate, which contains contents that are estimated to propagate to regions where u is predicted to move across in the near future time slot:

$$\mathcal{C}_u^{(T)} \equiv \{c | A_{c,r}^{(T)} > 0, \mathcal{Q}_{u,r}^{(T)} > 0\}. \tag{9}$$

The set $\mathcal{Z} \equiv \mathcal{W}_u^{(T-1)} \bigcup \mathcal{C}_u^{(T)}$ will then be the candidate content items for user u to replicate for the next time slot. Let $\mathcal{W}_u^{(T)}$ denote the set of content items that user u chooses to replicate. To determine which content items to carry for D2D delivery, a coordinate server calculates a local replication gain for content items in $\mathcal{Z}$. *c.gain* $\equiv \sum_{r \in \mathcal{R}} \mathcal{Q}_{u,r}^{(T)} A_{c,r}^{(T)}$ for user u to carry content c (line 0). After that, B_u content items

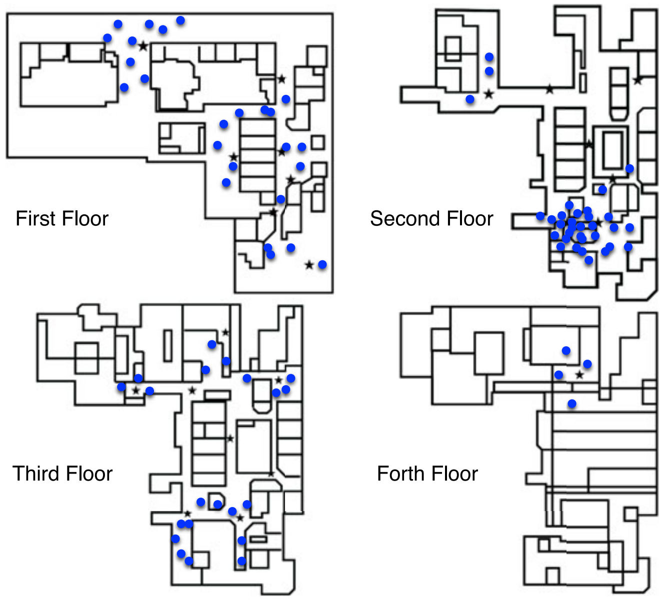


Fig. 9. Illustration of locations of users in one time slot in our experiments: stars represent the locations of Wi-Fi APs (regions), and circles represent users.

will be randomly chosen from the set $\mathcal{Z}$, where a content item has the probability to be chosen of $\frac{c.gain}{\sum_{c' \in \mathcal{Z}} c'.gain}$ (line 0).

Users that have replicated content items from others will then send a notification to the coordinate server, which can help users find potential edge peers from which to download social content items. Notice that traffic for such information communication at coordinator servers is substantially smaller than the traffic of content delivery.

5.4.2 On-Demand Service

Users replicating social content will serve other users in an on-demand manner: A requesting user who receives the *meta information* of a content in the social propagation will first contact the coordinator server to find which users in the same region have replicated the content. If a set of replicating users are found, the user will randomly select one user to download the content from. If the content cannot be downloaded from local peers, e.g., due to limitations presented by peers' upload capacities, the user will turn to the original content servers to download the content, guaranteeing that the user can download the requested content within a given delay.

5.4.3 Complexity Analysis

In our design, each user individually determines which content it will replicate using the algorithm above based on the social popularity indices provided by the centralized server, which periodically calculates $A^{(T)}_{c,r}, \forall c \in \mathcal{C}^{(T)}, r \in Regions$. Although $|\mathcal{C}^{(T)}||\mathcal{R}|$ can be too large for a single server to maintain the social popularity indices for all content items, the calculation itself can scale in a horizontal manner: (1) We partition the content items into several subsets. (2) Content items in each subset will be handled by one coordinate server, which maintains the social popularity index. (3) Users in the edge networks will retrieve the social popularity indices from different servers according to the content.

TABLE 2
Experimental Parameters Used in Both Scenarios

Parameter	Indoor	Outdoor
Social connections	40	40
λ_P	$[0.001, 0.02]$	$[0.001, 0.02]$
Re-share latency	10 hours	10 hours
Area size	$40 \times 100 \times 100$ m^2	5×5 km^2
Crowdedness	$[0, 15]$	$[0, 5]$

6 Performance Evaluation

In this section, we present our evaluation results of the propagation- and mobility-based replication strategy based on the trace-driven experiments.

6.1 Experimental Setup

We implemented our algorithms in C++ in a simulator based on an event-driven programming model, i.e., user mobility and social propagation activities; in addition, network transmissions are simulated as events with action times and handlers [40]. The activities are driven by the real traces.

Combining Mobility and Propagation Behaviors. We borrow the design idea from [41], the ONE simulator for DTN, and map users in the social propagation traces to the users in the mobility traces, i.e., the social behaviors and mobility behaviors from two traces are combined and assigned to one user in the simulation following different correlation levels. Note that we adjust the user mapping scheme in our experiments to vary the correlation between user mobility and social propagation and evaluate the correlation's impact on the effectiveness of our design.

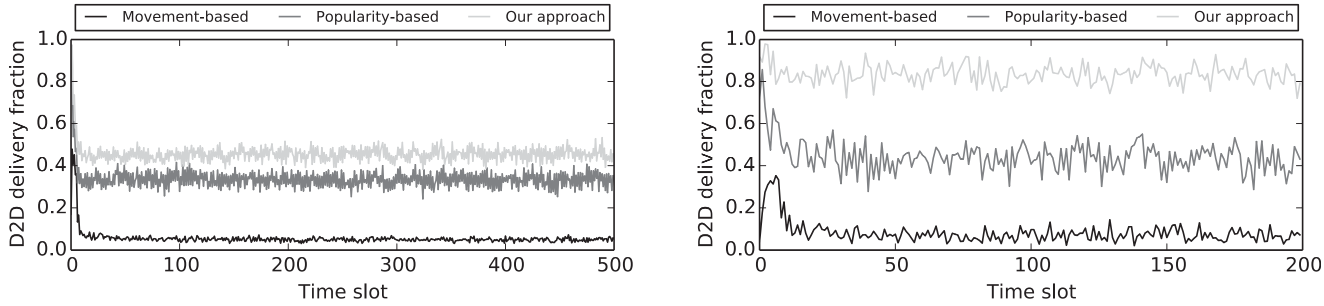
6.1.1 Mobility Behaviors

In our experiments, two types of scenarios are used to investigate user mobility patterns.

In our experiments, we also use an outdoor mobility dataset provided by Tencent Wi-Fi [42], a mobile app that asks users to respond to questions on how they use Wi-Fi networks. In particular, we collected information on how 1.2 million users move across the urban areas and associate with approximately 1 million Wi-Fi hotspots in Shenzhen over 10 days. Using these traces, we are able to infer user mobility in outdoor areas.

An indoor area (i.e., the shopping mall in our previous measurement study) is used to simulate typical indoor usage. An illustrative example of these locations is selected from 4 floors in the building. As illustrated in Fig. 9, a star represents a Wi-Fi AP covering a certain edge-network region, and a circle represents a user, who is associated with the closest AP.

In both scenarios, we divide the area into 100×100 m^2 regions in which users can communicate with each other. In our simulation, the mobility traces of over 10 k users are used in different areas. Each region has a popularity for users to visit based on the statistics of users staying in these regions. In Table 2, we present the important parameters used in our experiments.



(a) Outdoor scenario (b) Indoor scenario

Fig. 10. Comparison of fraction of requests successfully delivered by D2D replication over time in the indoor scenario.

6.1.2 Social Behaviors

We also select users from the social propagation traces and randomly map them to the users in the mobility traces. First, we build social connections between these users according to the distribution of social connections in the real world recorded in our traces. In our experiments, each user has approximately 40 social connections on average.

Then, we simulate the social propagation. Each time slot is set to be five minutes, which is approximately the average time that a user stays in a region according to our measurement study. The number of content items for a user u to post in each time slot in the online social network follows a Poisson distribution, which has a probability mass function $f(k;\lambda_P)=\frac{\lambda_P^k e^{-\lambda_P}}{k!}$. λ_P follows the distribution of the average number of content items posted by users in each time slot in the social propagation traces, in the range of $[0.001, 0.02]$ per time slot (about $[0.2, 5]$ per day), which is summarized from our traces. After a content item is posted/shared by a user, it will be re-shared by other users. The average latency of the re-shares is 10 hours.

In the online social network, users can reach the content generated by other users through the social connections. We observe that the number of content items propagating over different social connections is different. Based on our previous observation that the distribution of the social propagation intensity of different social connections follows a power-law distribution [43], we assign the re-share probability to each of the social connections following a power-law distribution learned from the social connections between the sampled users.

6.2 Baselines

In our study, we compare our design with baselines as follows.

⊳ *Movement-based* replication. In this scheme, contents are replicated only based on the mobility patterns of users. Many previous studies fall into this category. We implement the movement-based strategy according to the spatial mobility analysis surveyed in [18]. Our implementation of the movement-based scheme is as follows: (1) We build a contact measurement index for each user, i.e., a large index is assigned to a user who has contacted more other users. (2) Based on this contact measurement, we prioritize users with their contact indices. (3) When there is a content item to replicate, a user with a larger contact index is more likely to be selected to carry the content item. The rationale is that such a replication scheme greedily makes use of peers that have a large possibility to distribute the content to more users.

⊳ *Popularity-based* replication. In this scheme, content items are replicated to peers according to their popularities. In conventional peer-assisted content delivery, such a popularity-based approach has been widely used [44]. In our implementation of the popularity-based replication, the strategies are as follows: (1) A user ranks the popularity of content items according to their received requests. (2) When there are other users moving around, the user will choose a content item randomly from its local storage (a more popular content item has a larger chance to be chosen) and ask other users to replicate the content items. (3) A user, when receiving multiple replication requests from other users, will also determine which requests to replicate according to their popularity.

6.3 Metrics

We study the impact of running factors in the real world, including the intensity of social propagation, the crowdedness of users in the regions (e.g., the number of users in each 100×100 m^2 region), and the impact of the popularity distribution of the social contents. We use the following metrics to verify our design: (1) D2D delivery fraction, which is the fraction of traffic load that our D2D mechanism carries over all traffic served by both users and servers for the whole system, and (2) The delivery load distribution of users who perform the D2D delivery.

6.4 Experiment Results

6.4.1 Performance Improvement

We study the effectiveness of our design in terms of the amount of social contents that can be delivered by D2D replication, compared to the pure movement-based and pure popularity-based approaches. As illustrated in Fig. 10, the curves represent the fraction of D2D delivered content over time in both outdoor and indoor scenarios.

Our results are as follows. (1) In the outdoor scenario, the D2D delivery performance of our design is slightly better than that of the popularity-based approach, and approximately 45 percent of content items are delivered via D2D delivery under both strategies, as illustrated in Fig. 10a. (2) In the indoor scenario, as illustrated in Fig. 10b, our design can obtain a performance improvement of 4 times compared to the pure movement-based approach and of 2 times

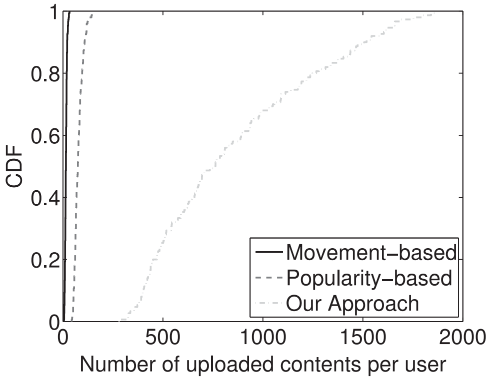


Fig. 11. CDF of users' contributions in D2D social content delivery in the indoor scenario.

compared to the pure popularity-based approach. This result indicates that, in edge-network content delivery, our proposal based on regional propagation and mobility predictions is more capable of fully utilizing edge-network users' peering resources, especially when propagation is intensive in the local area.

6.4.2 Users' D2D Contribution

We investigate the actual contributions of users in terms of the number of social contents served by them. In Fig. 11, each curve is the CDF of users' contributions (the number of uploaded social contents). We observe that, in our design, peers serve substantially more social contents to their neighbors than under the other two schemes. In particular, in our design, the users who can contribute the least are more utilized than under other strategies. In our experiments, greater than 50 percent of the content items are downloaded by users that are not requested by themselves; rather, they are only requested for other users. This indicates that good incentive mechanisms are required in such D2D systems.

6.4.3 Social Propagation Intensity

We study the impact of social propagation intensity by varying the number of social contents generated and shared by users per time slot. As illustrated in Fig. 12, each sample is the contribution of a particular user versus their user index. At a high propagation level, the number of social contents generated or shared by users per time slot is approximately 10; at the medium propagation level, this number is approximately 6; and at the low propagation level, this number is approximately 2. We observe that our design can well adapt to the social propagation intensity, all under the same storage and upload capacities. We also observe that, when the level of the propagation intensity is high, users tend to make more similar contributions.

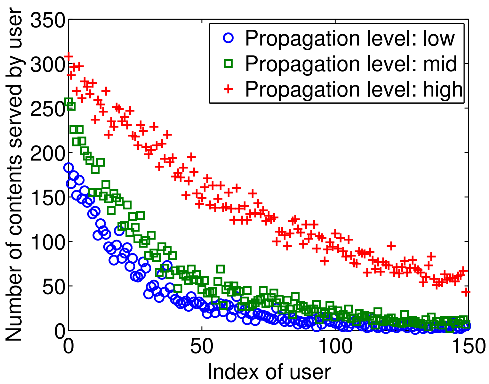


Fig. 12. Users' D2D contributions under different social propagation intensities in our design.

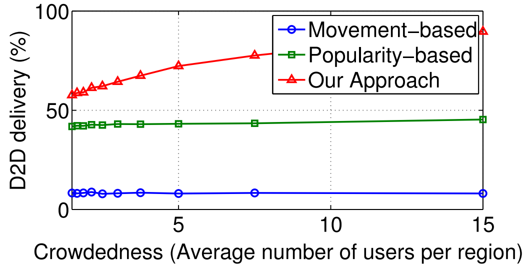


Fig. 13. Comparison of fraction of requests successfully delivered via D2D replication versus crowdedness.

6.4.4 Regional Crowdedness

D2D content delivery is by nature based on device-to-device communication; we thus evaluate the impact of the *crowdedness* of regions, which captures the average number of users per region. By varying the number of users in our experiments, we are able to change the regional crowdedness. As illustrated in Fig. 13, the curves represent the D2D delivery fraction against the average number of users per region during an experiment. We observe that, in contrast to the movement-based and popularity-based approaches, the performance of our design is found to be sensitive to the change in crowdedness. In particular, the D2D delivery fraction in our design increases when the number of users per region increases, especially in regions where there are few users.

6.4.5 Distance between Friends

By mapping social propagation users to mobility users in different scenarios, we obtain different distances between friends. We calculate the average distance between each pair of friends and study the impact of the average distance on the performance of these algorithms. As illustrated in Table 3, we observe that such D2D delivery performs well when social propagation occurs between friends that are close to each other, which is consistent with our previous results. When friends are moving at a city level, i.e., the average distance is greater than 5 km, the D2D delivery decreases to approximately 30 percent, which is only slightly better than the popularity-based approach.

TABLE 3
Impact of Average Distance Among Friends on D2D Delivery Fraction

Average distance (m)	D2D delivery fraction
$[0, 500)$	0.76
$[500, 1500)$	0.65
$[1500, 2500)$	0.59
$[2500, 5000)$	0.48
$[5000, \inf)$	0.31

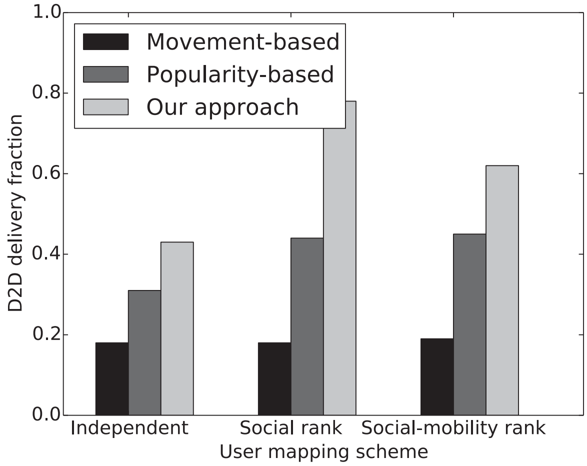


Fig. 14. Impact of social behavior and mobility behavior.

6.4.6 Correlation Between Social Behavior and Mobility Behavior

In practice, social propagation, including content generation and propagation behaviors, is not independent of user mobility, e.g., users may post more content items when they are waiting for a bus. We investigate the impact of the correlation between social behaviors and mobility behaviors using different user mapping schemes as follows. (1) Independent: We randomly map social users (users collected from the social network traces) to mobility users (users collected from the mobility traces). (2) Social rank: We rank users according to their social behavior intensity (i.e., the number of content posts and re-shares) and map the ranked users to mobility users in a random rank of regions. (3) Social-mobility rank: We rank users in the social network according to the propagation intensity, and we rank users from the mobility traces according their mobility intensity. Then, we map the ranked social users to the ranked mobility users.

As illustrated in Fig. 14, we plot the D2D delivery fraction for different strategies under different user mapping schemes. We observe that our approach is more sensitive to different user mapping schemes than are the movement-based and popularity-based approaches. We also observe that our design performs well when social behaviors occur in nearby regions, and social propagation and mobility are not highly correlated. These observations indicate that the performance of our design is affected by whether the social behaviors are independent of mobility behaviors. In particular, the performance is better when social behaviors are not uniformly distributed in different regions.

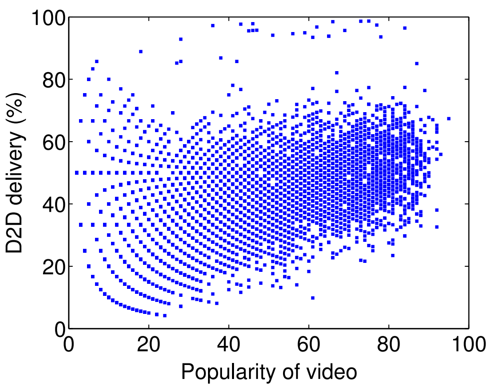


Fig. 15. Fraction of D2D delivery versus the popularity of contents.

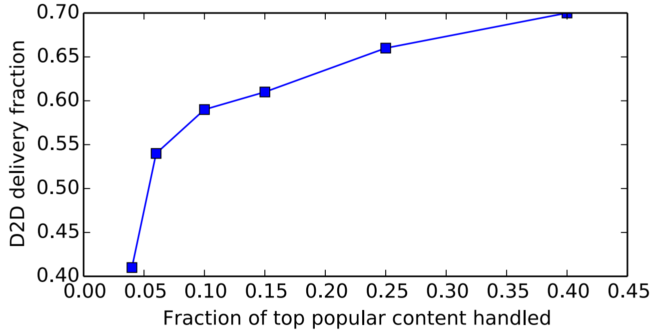


Fig. 16. Impact of fraction of the top popular content handled.

6.4.7 Content Popularity

We study the impact of the popularity of social contents, which is calculated as the number of all previous requests from users for a content. In Fig. 15, each sample is the D2D delivery fraction of a particular content during the experiments versus the popularity of the content. In general, we observe that our design can handle social contents with different popularities. In particular, we observe a slight trend that the D2D delivery fraction is larger when a content's popularity is higher. This is because, it is more possible for contents of high popularity (which are likely to be already cached by many users) to be better replicated.

6.4.8 Fraction of Content Handled

In a practical implementation, the system might not be able to monitor information about all the content items shared between users. We thus investigate the impact of the fraction of content items handled by the D2D delivery. We vary the fraction of the most popular content items that are considered to be delivered by peers and let the server serve the other content items. In Fig. 16, we plot the D2D delivery fraction versus the fraction of the most popular content handled by D2D delivery. We observe that, due to the high skewness of the popularity distribution of social content, our design can achieve a relatively high performance by only handling a small fraction of the most popular content.

7 Concluding Remarks

The massive number of user-generated bandwidth-intensive social contents and their highly skewed popularity distribution in practice make conventional content delivery based on a static and hierarchical infrastructure inefficient. In particular, it is difficult to serve users with network resources close to them for every social content shared. Motivated by the development of device-to-device communication, we propose a D2D replication for social contents. Based on extensive large-scale measurement studies, we find the local sharing and delay-tolerant characteristics of social content sharing and the regional propagation and mobility patterns of users. Based on these insights, we design regional propagation and mobility predictive models to estimate *where a social content may propagate during social propagation* and *which user can replicate it as they move*. We formulate the problem and design a heuristic algorithm based on only historical and local information to solve said problem. Trace-driven experiments further verify the effectiveness of our design, which not only outperforms conventional movement-based and popularity-based approaches by $2-4$ times but also is adaptive to different levels of social propagation intensity, regional crowdedness and content popularity.

ACKNOWLEDGMENTS

This work is supported in part by the National Natural Science Foundation of China under Grant No. 61402247, 61472204, and 61531006, the National Basic Research Program of China (973) under Grant No. 2015CB352300, and the research fund of Tsinghua-Tencent Joint Laboratory for Internet Innovation Technology. Z. Wang and L. Sun contributed equally to this study.

REFERENCES

[1] L. Zhang, F. Wang, and J. Liu, "Understand instant video clip sharing on mobile platforms: Twitter's vine as a case study," in *Proc. ACM Netw. Oper. Syst. Support Digit. Audio Video Workshop*, 2014 pp. 85:85–85:90.

[2] Youtube. [Online]. Available: http://www.youtube.com/yt/press/statistics.html

[3] F. T. Leighton and D. M. Lewin, "Content delivery network using edge-of-network servers for providing content delivery to a set of participating content providers," U.S. Patent 6553413, Apr. 22, 2003.

[4] A. Mislove, "Rethinking Web Content Distribution in the Social Media Era," in *Proc. NSF Workshop Soc. Netw. Mobility Cloud*, pp. 23–24, 2012.

[5] G. Fodor "Design aspects of network assisted device-to-device communications," *IEEE Commu. Mag.*, vol. 50, no. 3, pp. 170–177, Mar. 2012.

[6] 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. IEEE Int. Conf. Comput. Commun.*, 2014, pp. 2346–2354.

[7] E. Bakshy, J. Hofman, W. Mason, and D. Watts, "Everyone's an Influencer: Quantifying Influence on Twitter," in *Proc. ACM Int. Conf. Web Search Data Min.*, 2011, pp. 65–74.

[8] H. Li, H. Wang, and J. Liu, "Video Sharing in Online Social Network: Measurement and Analysis," in *Proc. ACM Netw. Oper. Syst. Support Digit. Audio Video Workshop*, 2012, pp. 83–88.

[9] G. Comarela, M. Crovella, V. Almeida, and F. Benevenuto, "Understanding factors that affect response rates in twitter," in *Proc. 23rd ACM Conf. Hypertext Soc. Media.*, 2012, pp. 123–132.

[10] J. Pujol, "The little engine(s) that could: Scaling online social networks," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 40, no. 4, pp. 375–386, 2010.

[11] S. Scellato, C. Mascolo, M. Musolesi, and J. Crowcroft, "Track globally, deliver locally: Improving content delivery networks by tracking geographic social cascades," in *Proc. ACM Int. Conf. Multimedia*, pp. 457–466, 2011.

[12] Z. Wang, L. Sun, C. Wu, and S. Yang, "Guiding Internet-scale video service deployment using microblog-based prediction," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2012, pp. 2901–2905.

[13] Z. Wang et al., "Propagation-based social-aware replication for social video contents," in *Proc. ACM Int. Conf. Multimedia*, 2012, pp. 29–38.

[14] Y. Wen, X. Zhu, J. J. Rodrigues, and C. W. Chen, "Cloud mobile media: Reflections and outlook," *IEEE Trans. Multimedia*, vol. 16, no. 4, pp. 885–902, Jun. 2014.

[15] H. Wang, F. Wang, J. Liu, C. Lin, K. Xu, and C. Wang, "Accelerating peer-to-peer file sharing with social relations," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 66–74, Sep. 2013.

[16] M. Kim, D. Kotz, and S. Kim, "Extracting a mobility model from real user traces," in *Proc. IEEE Int. Conf. Comput. Commun.*, vol. 6, 2006, pp. 1–13.

[17] J. Yoon, B. D. Noble, M. Liu, and M. Kim, "Building realistic mobility models from coarse-grained traces," in *Proc. 4th Int. Conf. Mobile Syst. Appl. Serv.*, 2006, pp. 177–190.

[18] D. Karamshuk, C. Boldrini, M. Conti, and A. Passarella, "Human mobility models for opportunistic networks," *IEEE Commun. Mag.*, vol. 49, no. 12, pp. 157–165, Dec. 2011.

[19] I. Rhee, M. Shin, S. Hong, K. Lee, S. J. Kim, and S. Chong, "On the levy-walk nature of human mobility," *IEEE/ACM Trans. Netw.*, vol. 19, no. 3, pp. 630–643, Jun. 2011.

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

[21] H. Zhuang, A. Chin, S. Wu, W. Wang, X. Wang, and J. Tang, "Inferring geographic coincidence in ephemeral social networks," in *Machine Learning and Knowledge Discovery in Databases*. Berlin, Germany: Springer, 2012, pp. 613–628.

[22] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: User movement in location-based social networks," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Min.*, 2011, pp. 1082–1090.

[23] T. Wang, G. Xing, M. Li, and W. Jia, "Efficient wifi deployment algorithms based on realistic mobility characteristics," in *Proc. 7th IEEE Int. Conf. Mobile Ad hoc Sens. Syst.*, 2010, pp. 422–431.

[24] D. Wang, D. Pedreschi, C. Song, F. Giannotti, and A.-L. Barabasi, "Human mobility, social ties, and link prediction," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Min.*, 2011, pp. 1100–1108.

[25] K. Fall, "A delay-tolerant network architecture for challenged internets," in *Proc. Conf. Appl. Technol., Archit. Protocols Comput. Commun.*, 2003, pp. 27–34.

[26] S. Jain, K. Fall, and R. Patra, "Routing in a delay tolerant network," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 34, no. 4, pp. 145–158, 2004.

[27] E. M. Daly and M. Haahr, "Social network analysis for routing in disconnected delay-tolerant manets," in *Proc. 8th ACM Int. Symp. Mobile Ad hoc Netw. Comput.*, 2007, pp. 32–40.

[28] Ó. R. Helgason, E. A. Yavuz, S. T. Kouyoumdjieva, L. Pajevic, and G. Karlsson, "A mobile peer-to-peer system for opportunistic content-centric networking," in *Proc. 2nd ACM SIGCOMM Workshop Netw. Syst. Appl. Mobile Handhelds*, 2010, pp. 21–26.

[29] J. Haillot and F. Guidec, "A protocol for content-based communication in disconnected mobile ad hoc networks," *Mobile Inf. Syst.*, vol. 6, no. 2, pp. 123–154, 2010.

[30] P. Hui, J. Crowcroft, and E. Yoneki, "BUBBLE rap: Social-based forwarding in delay-tolerant networks," *IEEE Trans. Mobile Comput.*, vol. 10, no. 11, pp. 1576–1589, Nov. 2011.

[31] L. Pelusi, A. Passarella, and M. Conti, "Opportunistic networking: Data forwarding in disconnected mobile ad hoc networks," *IEEE Commun. Mag.*, vol. 44, no. 11, pp. 134–141, Nov. 2006.

[32] Y. Li, Y. Jiang, D. Jin, L. Su, L. Zeng, and D. O. Wu, "Energy-efficient optimal opportunistic forwarding for delay-tolerant networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 9, pp. 4500–4512, Nov. 2010.

[33] B. Han, P. Hui, V. A. Kumar, M. V. Marathe, J. Shao, and A. Srinivasan,"Mobile data offloading through opportunistic communications and social participation," *IEEE Trans. Mobile Comput.*, vol. 11, no. 5, pp. 821–834, May2012.

[34] X. Bao, Y. Lin, U. Lee, I. Rimac, and R. R. Choudhury, "Dataspotting: Exploiting naturally clustered mobile devices to offload cellular traffic," in *Proc. IEEE INFOCOM*, 2013, pp. 420–424.

[35] Z. Wang, W. Zhu, M. Chen, L. Sun, and S. Yang, "CPCDN: Content delivery powered by context and user intelligence," *IEEE Trans. Multimedia*, vol. 17, no. 1, pp. 92–103, Jan. 2015.

[36] M. Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. Moon, "I tube, you tube, everybody tubes: Analyzing the world's largest user generated content video system," in *Proc. 7th ACM SIGCOMM Conf. Internet Meas.*, 2007, pp. 1–14.

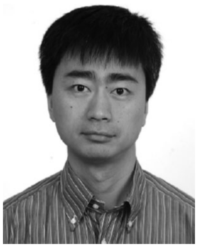
[37] M. Zhang, Q. Zhang, L. Sun, and S. Yang, "Understanding the power of pull-based streaming protocol: Can we do better?" *IEEE J. Select. Areas Commun.*, vol. 25, no. 9, pp. 1678–1694, Dec. 2007.

[38] P. Cui, F. Wang, S. Liu, M. Ou, S. Yang, and L. Sun, "Who should share what?: Item-level social influence prediction for users and posts ranking," in *Proc. ACM Int. Conf. Inf. Retr.*, 2011, pp. 185–194.

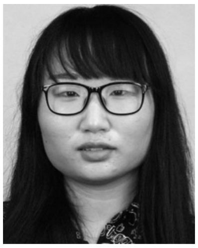
[39] G. Szabo and B. A. Huberman, "Predicting the popularity of online content," *Commun. ACM*, vol. 53, no. 8, pp. 80–88, 2010.

[40] A. Montresor and M. Jelasity, “Peersim: A scalable P2P simulator,” in *Proc. IEEE 9th Int. Conf. Peer-to-Peer Comput.*, 2009, pp. 99–100.
[41] A. Keränen, J. Ott, and T. Kärkkäinen, “The one simulator for DTN protocol evaluation,” in *Proc. 2nd Int. Conf. Simul. Tools Tech.*, 2009, Art. no. 55.
[42] Tencent. [Online]. Available: http://www.tencent.com/
[43] Z. Wang, B. Li, L. Sun, and S. Yang, “Dispersing instant social video service across multiple clouds,” *IEEE Trans. Parallel Distrib. Syst.*, vol. 27, no. 3, pp. 735–747, Mar. 2016.
[44] Y. Huang, T. Fu, D. Chiu, J. Lui, and C. Huang, “Challenges, design and analysis of a large-scale P2P-VoD system,” in *Proc. ACM SIGCOMM 2008 Conf. Data Commun.*, 2008, pp. 375–388.

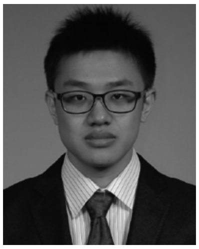
Zhi Wang (S’10-M’14) received the BE and PhD degrees in computer science from Tsinghua University, Beijing, China, in 2008 and 2014, respectively. He is currently an assistant professor with the Graduate School at Shenzhen, Tsinghua University. His research interests include online social networks, mobile cloud computing, and large-scale multimedia systems. He is a recipient of the China Computer Federation (CCF) Outstanding Doctoral Dissertation Award (2014), ACM Multimedia Best Paper Award (2012), and MMM Best Student Paper Award (2015).



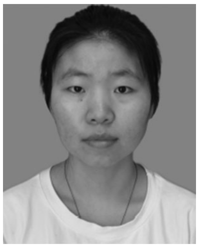
Lifeng Sun (M’05) received the BS and PhD degrees in system engineering from the National University of Defense Technology, Changsha, Hunan, China, in 1995 and 2000, respectively. He was an postdoctoral fellow from 2001 to 2003, an assistant professor from 2003 to 2007, an associate professor from 2007 to 2013, and currently a professor, all in the Department of Computer Science and Technology, Tsinghua University. His research interests include areas of online social network, video streaming, interactive multi-view video, and distributed video coding. He is a member of the IEEE and ACM.



Miao Zhang received the bachelor’s degree in computer science and technology from Sichuan University, Chengdu, China, in 2015. She is currently working toward the master’s degree in computer science from Tsinghua University, Beijing, China. Her research interests include content delivery network and data analysis.



Haitian Pang (S’16) received the BE degree from the Department of Automation, Tsinghua University, Beijing, China, in 2014. He is working toward the PhD degree in computer science from Tsinghua University. His research interests include network game modeling, cellular-WiFi networking, and mobile networking optimizations.



Erfang Tian received the BE degree in soft engineering from Xidian University, Xi’an, China, in 2013. She is working toward the master’s degree in computer science from Tsinghua University in 2013. Her main research interests include QoS improvement in Wi-Fi networks.

Wenwu Zhu (M’97-SM’01-F’10) received the PhD degree in electrical and computer engineering from the New York University Polytechnic School of Engineering in 1996. He is with the Computer Science Department at Tsinghua University as professor of “1,000 People Plan” of China. His current research interests include area of multimedia cloud computing, social media computing, multimedia big data, and multimedia communications and networking. He served(s) on various editorial boards, such as guest editor for the *Proceedings of the IEEE*, the *IEEE Transactions on Circuits and Systems for Video Technology*, and the *IEEE Journal on Selected Areas in Communications*; associate editor for the *IEEE Transactions on Mobile Computing*, the *IEEE Transactions on Multimedia*, and the *IEEE Transactions on Circuits and Systems for Video Technology*; leading editor of the area “Computer Networks and Distributed Computing” of the *Journal of Computer Science and Technology*. He received the Best Paper Award in ACM Multimedia 2012, the Best Paper Award in the *IEEE Transactions on Circuits and Systems for Video Technology* in 2001, and the other four international Best Paper Awards.