Propagation-Based Social-Aware Replication for Social Video Contents

Zhi Wang[1], Lifeng Sun[1], Xiangwen Chen[2], Wenwu Zhu[1],
Jiangchuan Liu[3], Minghua Chen[2] and Shiqiang Yang[1]

[1]Tsinghua National Laboratory for Information Science and Technology
Department of Computer Science and Technology, Tsinghua University
[2]Department of Information Engineering, The Chinese University of Hong Kong
[3]School of Computing Science, Simon Fraser University

ABSTRACT

Online social network has reshaped the way how video contents are generated, distributed and consumed on today's Internet. Given the massive number of videos generated and shared in online social networks, it has been popular for users to directly access video contents in their preferred social network services. It is intriguing to study the service provision of social video contents for global users with satisfactory quality-of-experience. In this paper, we conduct large-scale measurement of a real-world online social network system to study the propagation of the social video contents. We have summarized important characteristics from the video propagation patterns, including social locality, geographical locality and temporal locality. Motivated by the measurement insights, we propose a propagation-based social-aware replication framework using a hybrid edge-cloud and peer-assisted architecture, namely PSAR, to serve the social video contents. Our replication strategies in PSAR are based on the design of three propagation-based replication indices, including a geographic influence index and a content propagation index to guide how the edge-cloud servers backup the videos, and a social influence index to guide how peers cache the videos for their friends. By incorporating these replication indices into our system design, PSAR has significantly improved the replication performance and the video service quality. Our trace-driven experiments further demonstrate the effectiveness and superiority of PSAR, which improves the local download ratio in the edge-cloud replication by 30%, and the local cache hit ratio in the peer-assisted replication by 40%, against traditional approaches.

Categories and Subject Descriptors

C.2.4 [**Distributed System**]: Distributed Applications; H.4 [**Information Retrieval**]: Social Network

General Terms

Measurement, Design

Keywords

Social network, video replication, hybrid edge-cloud and P2P

MM'12, October 29–November 2, 2012, Nara, Japan.

1. INTRODUCTION

Recent years have witnessed the blossom of online social network service (*e.g.*, Facebook, Twitter) and online video service (*e.g.*, YouTube, Youku), as well as the rapid convergence of the two services [28]. *Social video contents*, or *social videos* in short that are *generated* and *shared* by users in online social networks are becoming increasingly popular on today's Internet. ForeSee has reported that more than 18% users are influenced by the social network when accessing video contents [3]. It is fascinating to study how the social video contents can be served to users with satisfactory Quality-of-Experience (QoE).

In the online social network, users create and maintain different social connections, *e.g.*, *friending* their friends in real life, *following* celebrities or even *liking* virtual social entities. Such social connections determine which videos can reach a user in the online social network [16]. The unique *propagation* properties make the video access pattern in the online social network quite different from that in the traditional centralized video service systems, in that (1) video contents are no longer produced by a few centralized content providers, but by all individual users; and (2) social connections and social activities determine the propagation of the videos among the users.

We are facing the following challenges in distributing the social video contents with satisfactory QoE: (1) A huge number of user-generated videos require a large amount of storage and network resource, *e.g.*, YouTube has hit a new record of 60 hours' worth of videos uploaded by users per minute [4]; (2) Newly generated videos are the ones that tend to attract most of the users, but it is difficult to estimate their popularity for the video service allocation, which are dynamically affected by the social network [9]; (3) Social video contents have close-to-uniform and highly-volatile popularity profiles, because a large portion of the videos are shared among small social groups (*e.g.*, family members).

Challenge (1) makes traditional service paradigms (*e.g.*, C/S) not suitable, and a common practice to provision these video services is to replicate videos in severs at different geographic regions [5] by allocating resource from the CDN (Content Devilry Network) or cloud, where videos can be dynamically placed to serve users all over the world. Challenges (2) and (3) make the traditional replication approaches, which work well only for videos with skewed and stable popularity profiles, not suitable in the context of online social network. Mislove *et al.* [21] have observed a large deduction of cache hit ratio when traditional caching schemes are used to replicate social contents. In this paper, we reveal a key observation that the social videos, unlike regular videos, do not propagate among users randomly. Instead, they propagate along the social-network topology according to several rules determined by the social propagation. Exploiting the new design space enabled by this observation, we develop a social-aware replication system — PSAR, to effectively distribute social videos with superb QoE.

First, we demonstrate that the statistic information obtained from the online social network can guide the video replication. We conduct large-scale measurement to explore the connection between the social propagation and the replication, and discover the propagation characteristics of social video contents, including social locality that videos are generally shared among users who are socially connected, geographical locality that most of the videos are shared between users that are geographically close to each other, and temporal locality that most of the activities are issued to videos that are recently generated or shared.

Second, based on the measurement, we propose a hybrid edge-cloud and peer-assisted video replication framework, where videos are replicated by both the edge-cloud servers and peers at different geographic locations. In this framework, we are facing the following problems: (a) which videos should be replicated to which edge-cloud servers? (b) how much bandwidth should be reserved for each video by the edge-cloud? and (c) which videos should be served by which peers? We address these questions in the design of PSAR as follows. (1) We summarize three replication indices from the propagation patterns; (2) We design the edge-cloud replication strategies based on the geographic influence index and content propagation index, determining the region selection and bandwidth reservation for each video; (3) We further design the peer-assisted replication based on the social influence index, performing social-aware cache replacement at each peer.

The remainder of this paper is organized as follows. In Sec. 2, we discuss related work. In Sec. 3, we motivate our design by a measurement study on social video propagation. We present the architecture of PSAR in Sec. 4, and the detailed design in Sec. 5. In Sec. 6, we evaluate the performance of PSAR by trace-driven simulations. Finally, we conclude the paper in Sec. 7.

2. RELATED WORK

Propagation in Online Social Network. Online social network has become a popular Internet service. Based on traces from Flickr, YouTube, LiveJournal and Orkut, Mislove *et al.* [20] study the topology of the social graph, and confirm the power-law, small-world, and scale-free properties of the online social network. Krishnamurthy *et al.* [7] investigate Twitter, and identify the distinct classes of Twitter users and their behaviors, as well as the geographic growth patterns of the social network.

In an online social network, contents spread among users by their social activities. A number of research efforts have been devoted to studying the propagation of information in online social networks. Kwak *et al.* [16] investigate the impact of users' retweets on information diffusion in Twitter. Dodds *et al.* [12] use the epidemic model to study the information propagation, where a piece of information is regarded as an infective disease that spreads via the social connections. Kempe *et al.* [15] investigate how to maximize the spread of influence in an online social network, and Hartline *et al.* [13] utilize such maximum spread to achieve revenue maximization.

In this paper, we will study how to connect the social propagation and the social video replication, *i.e.*, how statistic information about the video propagation can be utilized to guide the video content replication in a joint edge-cloud and peer-assisted architecture.

Social Video Replication. Many architectures have been proposed in large-scale video service systems, including (1) the server-based architecture, *e.g.*, CDN and cloud-based approaches [23], (2) the client-based architecture, *e.g.*, the P2P content distribution [18], and (3) the hybrid architecture, *e.g.*, a hybrid CDN and P2P distribution framework [30]. For Internet-scale social video service, replicating the videos at different geographic regions is a promising approach to provide good service quality to users [6].

However, online social network has greatly changed the assumptions in traditional replication algorithms [8], *e.g.*, the distribution of video contents is shifted from a "central-edge" manner to an "edge-edge" manner, resulting in the close-to-uniform popularity distribution. Li *et al.* [17] study the video sharing in the online social network, and observed the skewed popularity distribution of contents and the power-law activity of users. To better serve such social video contents, some social-aware video replications have been proposed. Pujol *et al.* [24] investigate the difficulties of scaling online social networks, and designed a social partition and replication middle-ware where users' friends' data can be co-located in the same server. Tran *et al.* [26] study the partition of contents in the online social network by taking social relationships into consideration. Nguyen *et al.* [22] study how to improve the system efficiency in case of server failures by taking social locality into consideration. Wang *et al.* [27] observe that a social network can be used to help predict the video access pattern in a standalone video-on-demand system. Wu *et al.* [29] study how to minimize the cost in social media migration among servers at different regions. Cheng *et al.* [11] study the social media partition to balance the server load and preserve the social relationship.

This paper focuses on effectively distributing videos generated and propagated inside the social network. We will study how to improve the user experience of watching social videos by exploring a joint edge-cloud and P2P design based on the propagation characteristics of social videos extracted from real-world measurement.

3. MEASUREMENT OF PROPAGATION

In this section, we investigate how videos are generated and distributed among users in the online social network.

3.1 Measurement Setup

In our measurement, we have collected traces from Tencent Weibo [2], which is a microblogging website, where users can broadcast a message including at most 140 characters

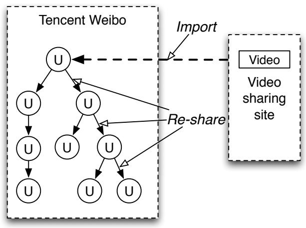

Figure 1: Connection between Weibo and video sharing systems.

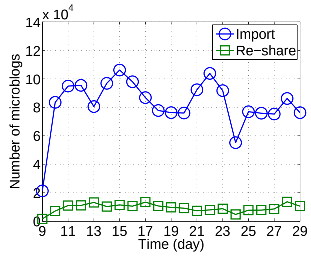

Figure 2: Imports and re-shares over time.

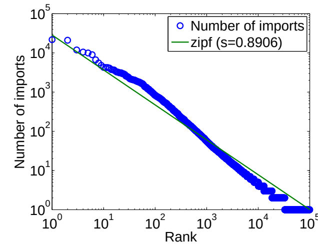

Figure 3: Number of imports of a video versus its rank.

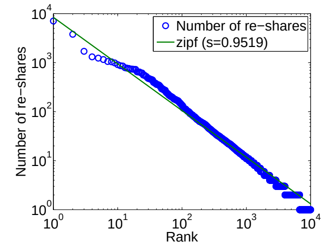

Figure 4: Number of re-shares of a video versus its rank.

to their friends. Tencent Weibo features several social activities in the system, *e.g.*, online chatting with friends who mutually follow each other. We obtained Weibo traces from the technical team of Tencent, containing valuable runtime data of the system in 20 days (October 9 – October 29) in 2011. Each entry in the traces corresponds to one microblog posted, including ID, name, IP address, geographic location of the publisher, time stamp when the microblog was posted, IDs of the parent and root microbloggers if it is a re-post, and contents of the microblog. The traces were recorded on an hourly basis.

We are focused on microblogs with video links which are imported from external video sharing websites. In particular, we have collected $350,860$ video links from 5 popular video sharing sites: Youku, Ku6, Tudou, Xunlei and Tencent Video. We then retrieve the microblogs which are related to these video links, *i.e.*, the microblogs either include the video links to these videos in the contents or they are re-shares of the ones that include the links. These video links cover $1,923,507$ microblogs in the time span. Besides, we also retrieve the profiles of users who have posted these microblogs, *e.g.*, their friend lists. In our measurement, we use the number of microblog posts to estimate the number of video views, in a sense that the microblog publishers can represent a sample of users who have watched the videos.

Fig. 1 illustrates how Tencent Weibo are connected with the video sharing sites. After a video is published on a video sharing site, the link to that video can be imported by users to Weibo. We will regard the import as the video generation by that user. Then users who are socially connected to that user can be reached by the imported video and further re-share the video.

3.2 Generation, Distribution and Popularity of Social Video Contents

Video Generation and Distribution. On Tencent Weibo, users generate videos by importing the links to the videos from the external video sharing sites, and distribute the videos by re-sharing the microblogs containing the links. Import and re-share are the most important activities that determine how videos reach users in the online social network. Fig. 2 illustrates the number of imports and re-shares of the targeted videos over time. We observe that (1) more users are generating videos instead of distributing them in the online social network, and (2) the number of imports shows more obvious weekly pattern than the number of re-shares, indicating more randomness in users' re-sharing of social video contents.

Popularity Profiles. Videos can reach many people in the online social network by users' importing and re-sharing, which determine the video propagation range. We observe that different videos attract quite different levels of imports and re-shares, resulting in a skewed popularity distribution of videos in the online social network. We study the popularity distribution of the social video contents, in terms of their imports and re-shares in a given time period of 1 day. In Fig. 3, videos are ranked in their import number's descending order. Each sample illustrates the number of imports of a video versus the rank of that video. We observe that the video import popularity is highly skewed, following a zipf-like distribution with a shape parameter of $s = 0.8906$. Similarly, Fig. 4 illustrates the number of re-shares versus the rank of the video, and we observe the video re-share popularity also follows a zipf-like distribution with a shape parameter of $s = 0.9519$. The popularity distributions of the import and re-share indicate that there are a dominate fraction of unpopular videos in the online social network — it is of great challenge to serve all the videos to users locally, with limited storage and network resources.

We further investigate in which types of social groups are these unpopular videos propagating. By randomly collecting 50 videos with different propagation size (the number of users involved in a video's propagation), we explore the correlation between the propagation size and the clustering coefficient of the social group formed by the users involved in the propagation. In Fig. 5, each sample illustrates the video propagation size versus the clustering coefficient of the corresponding social group [1]. We observe a relatively strong correlation between the propagation size and the clustering coefficient. The reason is that the unpopular videos tend to be shared among small social groups that are relatively closely connected (socially). The trend of many unpopular videos to be shared among small social groups results in a close-to-uniform popularity distribution, which makes the replication extremely challenging.

3.3 Characteristics of Social Video Propagation

The above measurement has demonstrated the challenges in the replication of social video contents. Next, we study the characteristics of the social video propagation to guide the replication design.

3.3.1 *Social Locality in Propagation*

The generation and re-share of a video on Weibo form a propagation tree which is rooted by the user who generates the video. Any user who re-shares the video will become

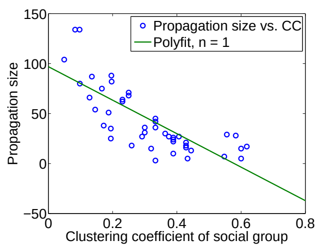

Figure 5: Propagation size versus the clustering coefficient of social group.

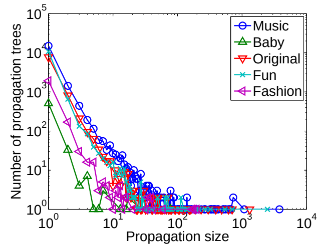

Figure 6: The number of propagation trees versus the propagation size.

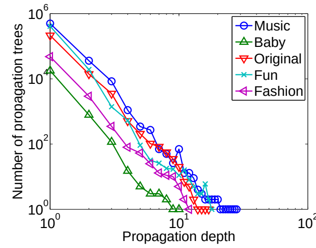

Figure 7: The number of propagation trees versus the propagation depth.

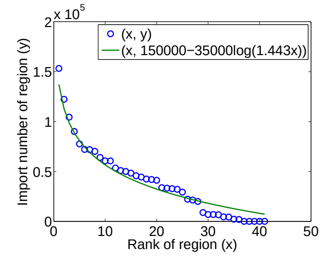

Figure 8: The import number of a region versus the rank of the region.

a new leaf node in the propagation tree. Fig. 6 shows the propagation size of videos in 5 different categories. Each sample illustrates the number of propagation trees (with the same propagation size) versus the size of these propagation trees. We observe that the size of most propagation trees is very small, *e.g.*, about 90% of the propagation trees are smaller than 100. Next, we study the propagation depth, which is defined as the average number of social hops between users in the propagation tree and the root user. Fig. 7 illustrates the propagation depth of videos in the same 5 categories. Each sample represents the number of propagation trees (with the same propagation depth) versus their propagation depth. We observe that in most of the propagation trees, the depth does not exceed 10, *i.e.*, users who re-share the video are socially close to the root user (with a small number of social hops between them).

The limited propagation size and propagation depth indicate that in each propagation tree, only users within a *limited social range* will be reached by the video. This observation motivates us to design the peer-assisted replication so that users who are both socially and geographically close to each other, can help distribute the video contents among themselves effectively.

3.3.2 Geographical Locality in Propagation

Users who generate, re-share and view the social videos are located in a variety of regions over the world. For Internet-scale video service providers, when performing replication for the social video contents, they need to strategically determine the regions (where datacenters are deployed) where the videos should be stored and served. To this end, we investigate how a social video content propagates among different geographic regions.

First, we observe that the popularity of different regions is quite different. We define the import number of a region as the number of total imports issued by users in the region. In Fig. 8, we rank 41 regions in their import number's descending order. Each sample in this figure illustrates the import number of a region versus the rank of the region. We observe that the popularity distribution of regions with respect to their import numbers follows a logarithm function $y = 150000 - 35000\log(1.443x)$. This observation indicates that it is not necessary to replicate each video to all the regions. A video should be replicated to a region only when the region is in the video's propagation range.

Next, we explore how to utilize the propagation information to estimate the geographical range of the video propagation. We observe that the propagation size can be used

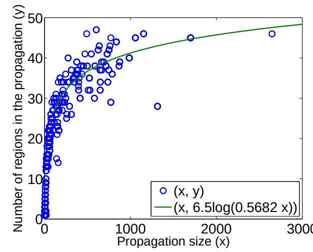

Figure 9: Number of regions in the propagation versus the size of the propagation.

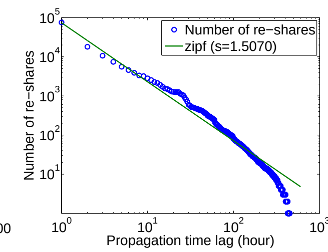

Figure 10: Number of re-shares versus the time lag.

to predict the geographical propagation range. Fig. 9 illustrates the correlation between the number of regions involved in the video propagation and the propagation size for different videos. We observe that a large propagation size generally results in more regions involved in the propagation. The relationship follows a logarithm function $y = 6.5\log(0.5682x)$. In PSAR, the propagation size is utilized to determine whether a video will be replicated to more regions. Intuitively, a video should be replicated to more regions when the predicted number of regions involved in the propagation is larger than the number of regions it has already been replicated to.

3.3.3 Temporal Locality in Propagation

In the online social network, we observe that users are more likely to re-share new video contents, *i.e.*, videos that are recently imported or re-shared. Fig. 10 illustrates the number of re-shares of a video in a time slot (1 hour) versus the time lag since it is generated. We observe that most of the re-shares happen in the recent hours, and the re-share number against the time lag follows a zipf-like distribution with a shape parameter $s = 1.5070$. This observation indicates that new videos in the online social network can attract more re-shares, leading to more viewers of the videos. We will also incorporate the temporal locality into the design of PSAR.

4. ARCHITECTURE OF PSAR

Based on the characteristics of the social video propagation, in this section, we present the conceptual architecture of PSAR and its key components, respectively.

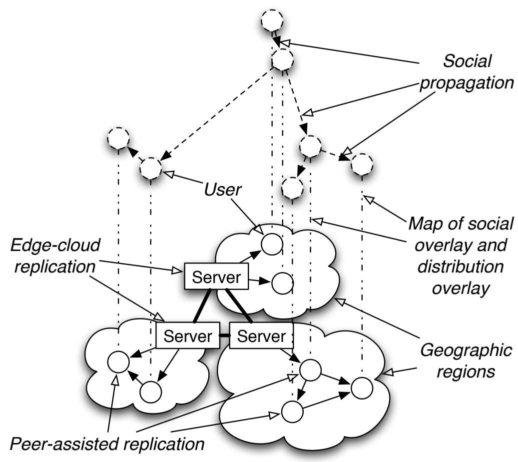

Figure 11: Conceptual architecture of PSAR.

4.1 Conceptual Architecture

We employ a joint edge-cloud and P2P architecture to replicate the social video contents, where the edge-cloud can support the time-varying bandwidth and storage allocations requested by different regions, while the peers are able to help contribute to each other in similar social groups. Fig. 11 illustrates the conceptual architecture of our design. In this figure, two overlays are presented as follows: (1) *social propagation overlay* based on the social graph, which determines the video propagation among friends, *i.e.*, users after generating a video, can share the video with their direct friends, who will further re-share the video to more people, and (2) *delivery overlay* which determines how video contents are delivered from edge-cloud servers to users or among themselves in a P2P paradigm. In this architecture, on one hand, we make use of the edge-cloud servers distributed at different geographic regions, to serve the social videos to users from different regions; on the other hand, we let peers cache the video contents in their local storage, so that they can help each other to download the videos. In the design of PSAR, we will study the edge-cloud replication on how videos are replicated to edge-cloud servers, as well as the peer-assisted replication on how videos are cached at a peer.

4.2 Framework and Key Components

In our measurement, we have shown that the social video propagation demonstrates social locality, geographical locality and temporal locality. In PSAR, we design three indices based on the social propagation to guide the replication: the geographic influence index which represents a video's propagation range among different regions, the content propagation index which represents the video's ability to attract new users in the online social network, and the social influence index which represents the possibility for a video to be requested by a peer's local friends. Based on the three replication indices, we design the edge-cloud replication and the peer-assisted replication, respectively.

4.2.1 Edge-Cloud Replication

Purpose. In the edge-cloud video replication, video contents are generally replicated to servers located in different geographic regions. The main purpose of the edge-cloud replication is for users at different locations to download the wanted videos from their local servers, which are located in the same regions with the users, to improve the video service quality [6].

Redesign. We redesign the edge-cloud replication by taking the social propagation into account. We first select the videos that are the most likely to propagate across geographic regions, by evaluating the videos' geographic influence index we design. Since the selected videos are more likely to attract users from more regions in the future, we replicate them to more regions so that users can be better served by the local servers. After that, based on the social influence index, which reflect their popularity in the near future, we determine which regions to replicate these videos to and how much bandwidths to allocate for the videos. We will present the detailed design in Sec. 5.2.

4.2.2 Peer-Assisted Replication

Purpose. The reason we propose a joint edge-cloud and peer-assisted paradigm in the social video replication lies two-folds. (1) Social videos are generally shared in small social groups, resulting in the close-to-uniform popularity distribution of the videos, which cost a huge amount of server resource to be distributed to users. To scale the delivery system, peers' resource is in demand. (2) Users typically share videos with their friends, who are observed geographically close to each other [25] — these socially connected users tend to have good Internet connectivity between each other to perform the peer-assisted video download [14].

Redesign. In traditional peer-assisted video distribution, LRU and LFU-based cache replacement algorithms are widely used. Such algorithms only depend on the static popularity of the video contents, which cannot achieve good performance when the access patterns of videos are affected by the social activities in the online social network. Based on the social influence index summarized from the propagation pattern, we redesign the peer cache replacement algorithm. In particular, we let peers cache videos that not only improve the general peer contribution (*i.e.*, the fraction of video contents upload by peers over all videos uploaded), but also improve the possibility for peers to serve the unpopular videos to their local friends. These friend users can benefit from the good Internet connectivity to the local peers. We will present the detailed design in Sec. 5.3.

5. SYSTEM DESIGN OF PSAR

In this section, we first present the design challenges in PSAR. Then, we establish the connection between the social video propagation and the video replication. After that, we present the detailed design of PSAR based on the connection.

5.1 Challenges in the Design of PSAR

In PSAR, the replication of social video contents is facing great resource-allocation challenges in the presence of multiple video propagations. Fig. 12 illustrates an example when there are only two videos. In this figure, the circles represent users in the online social network, which are located in different geographic regions, *e.g.*, region 1 and region 2. User A generates and shares video a in time slot T, then the video is re-shared by his friends C and D in time slot $T+1$. At the same time, another user B generates a different video b. Video a and video b will propagate across the social connections, and the two propagation trees may intersect in

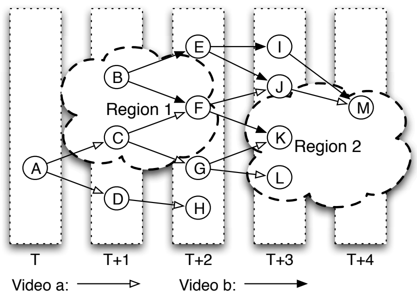

Figure 12: Example of two propagation trees.

the same region or at the same peer, *e.g.*, both region 1 and region 2 are involved in the two propagation trees, and both videos can reach user K in time slot $T+3$. The resource allocation has to determine (1) how to serve video a and b by the edge-cloud servers in region 1 and 2, and (2) how to cache video a and b at the peers to help others. It is of great challenge when many videos are propagating at the same time. The two problems will be discussed in the edge-cloud replication and the peer-assisted replication, respectively.

5.2 Edge-Cloud Replication

The edge-cloud provides the backup servers when peers are not able to download videos from the neighboring peers. We design two replication indices to guide the edge-cloud replication. When performing the edge-cloud replication, we first select the videos to be replicated and determine regions they should be replicated to, then we reserve upload bandwidth at edge-cloud servers for them.

5.2.1 Social-Aware Edge-Cloud Replication Indices

To perform social-aware replication at the edge-cloud servers, we design two replication indices based on the social propagation.

Geographic Influence Index. When performing video replication, we need to find out the videos that may propagate to more regions in the future. We deign the geographic influence index for that. In our measurement study, we observe that the number of propagation regions can be predicted from the propagation size of the video. Based on our measurement, we design the geographic influence index, which is calculated as follows:

$$g_v^{(T)} = c_1 \log(c_2 s_v^{(T)}),$$

where $s_v^{(T)}$ is the propagation size of the propagation tree of video v in time slot T. Large $g_v^{(T)}$ indicates that more regions will be involved in the propagation of the video. To achieve a better video download quality, a video with a larger $g_v^{(T)}$ should be replicated to more regions to serve users locally. In our experiments, we use $c_1 = 6.5, c_2 = 0.5682$ which are consistent with our measurement results. Based on the geographic influence index, we can predict whether the regions where the video has been replicated are enough.

Content Propagation Index. We design a content propagation index to evaluate the strength of a video's propagation in time slot T based on the propagation information as follows: (1) the current propagation size ($s_v^{(T)}$); (2) the current propagation depth ($h_v^{(T)}$); and (3) the time lag since the propagation tree is formed ($\tau_v^{(T)}$). The content propagation index is defined as follows:

$$e_v^{(T)} = z_s(\tau_v^{(T)})(s_v^{(T)}/h_v^{(T)}),$$

where $z_s(\tau_v^{(T)})$ is a decreasing function to make use of the temporal locality, which can adjust the content propagation index according to $\tau_v^{(T)}$ so that more recently generated or shared videos will have larger content propagation index. Based on our observation in Sec. 3.3.3, $z_s(t)$ is defined as follows:

$$z_s(t) = 1/(t^s \sum_{k=1}^{N} \frac{1}{k^s}),$$

where s is the zipf shape parameter and N is the number of hours between the initial time of the earliest video and the latest video. In our experiments, we let $s = 1.507$ and $N = 600$, which are the same as used in our measurement of Tencent Weibo's traces. In our design, $e_v^{(T)}$ will be used to guide the replication. Larger $e_v^{(T)}$ indicates that more users can join the propagation tree in time slot T. The rationale of $e_v^{(T)}$ lies as follows: (1) Larger $s_v^{(T)}$ indicates that more users can be reached by the video, and these users are the potential viewers (downloaders) of video v; (2) According to the social locality, small $h_v^{(T)}$ indicates that users in the propagation tree are still socially close to the root user and the video can propagate more; (3) According to the temporal locality, large $\tau_v^{(T)}$ slows down the propagation. Based on the content propagation index, we will determine how much bandwidth we will reserve for a video in the future time slot in PSAR.

5.2.2 Video and Region Selection

Initial Replication. After video v is first generated by a user in the online social network, it will be stored by a server which is closest to the user's friends. Let $d_{r,i}, i \in \mathcal{F}_v$ denote the geographic distance between region r and user i, where $\mathcal{F}_v$ is the set of friends of the root user of video v ("distance" based on Internet connectivity measurement can also be used, *e.g.*, bandwidth or RTT). The initial region is then selected by solving the problem as follows:

$$r_v = \arg\min_{r \in \mathcal{R}} \sum_{i \in \mathcal{F}_v} d_{r,i},$$

where $\mathcal{R}$ is the set of regions that can be used for the replication, and r_v is the region selected for the replication.

Selecting Existing Videos for Replication. According to our measurement study, we observe that although there are a massive number of videos in the online social network, in each time slot, only limited videos are shared among users. In particular, we observe that among 350,860 videos that we study in our measurement, only 1919 of them are re-shared in one time slot (1 hour) on average. Thus, in each time slot, only a little fraction of existing videos need to be replicated to improve the service quality. How should we select the candidate videos for replication? We observe that the overlapped fraction of the common videos that are re-shared in time slot T and $T-1$ over all videos re-shared in time slot T can be as large as 49%. In our design, the replication video set $\mathcal{V}^{(T)}$ is constructed as follows. (1) We build a candidate video set $\mathcal{W}^{(T)}$ by selecting videos that are imported or re-shared in the previous time slot. In particular, we randomly choose 80% of the videos that have been imported or re-shared in the previous time slot and 20% of the videos among the most popular ones in history. (2) We choose the videos in $\mathcal{W}^{(T)}$ that have the geographic influence index $g_v^{(T)}$ larger than $\theta_v^{(T)}$, which is a control parameter depending on the current replication status of video v, to form the video replication set $\mathcal{V}^{(T)}$. In our experiments, we let $\theta_v^{(T)} = 0.8|\mathcal{R}_v^{(T)}|$, where $\mathcal{R}_v^{(T)}$ is the set of regions that v has been replicated to. The rationale is that a video should be replicated to more regions if its current replication is under the requirement estimated from the geographic influence index.

Selecting Replication Regions for Videos in $\mathcal{V}^{(T)}$. After $\mathcal{V}^{(T)}$ has been constructed, the videos in $\mathcal{V}^{(T)}$ need to be replicated to more regions. Since these videos are the candidates that can attract users from more regions, we have to determine which videos need to be replicated to which regions. In our design, we extend the replication of a video to one more region each time. The selection of the region is similar to the approach used in the initial region selection. We minimize the geographic distance between the region and the potential users who may join the propagation tree. Let $\mathcal{L}_v^{(T)}$ denote the set of users who join the propagation tree in the previous time slot. The selection is as follows:

$$r_v = \arg \min_{r \in \mathcal{R} - \mathcal{R}_v^{(T)}} \sum_{i \in \bigcup_{k \in \mathcal{L}_v^{(T)}} \mathcal{F}_k} d_{r,i},$$

where $\mathcal{F}_k$ is the friend set of user k. The rationale is that users in $\mathcal{L}_v^{(T)}$ are the ones who join the propagation tree in the previous time slot, and it is likely for them to attract new users of the video, due to the temporal locality of the propagation. We utilize these users' friends' locations as a sample of all the users that can join the propagation tree, and select the region that is closest to all the users. The benefit of always extending a video to a new region in the replication (*i.e.*, r_v is selected from $\mathcal{R} - \mathcal{R}_v^{(T)}$) is that users in a popular propagation tree can choose more regions to download the video contents from, and our scheme improves the possibility for them to select the preferred regions.

5.2.3 Bandwidth Reservation

In each schedule round, we need to allocate upload bandwidths at the edge-cloud servers for the videos replicated. In our design, the bandwidth reservation depends on the social propagation strength, which can be evaluated by the content propagation index $e_v^{(T)}$. Let $\mathcal{V}_r$ denote the set of videos that are replicated in region r, the bandwidth reservation is then performed as follows:

$$b_{v,r_v} = B_{r_v} e_v^{(T)} / \sum_{v \in \mathcal{V}_{r_v}} e_v^{(T)}, \forall v \in \mathcal{V}^{(T)},$$

where b_{v,r_v} is the amount of bandwidth to be reserved for video v in the selected replication region r_v when the region is fully requested by users of different videos; and a video can extend to use more than b_{v,r_v} when the region is not fully loaded. B_r is the upload capacity of region r. The rationale of the bandwidth reservation is that videos with larger $e_v^{(T)}$ tend to attract more users in the propagation in the near future, and more upload bandwidth should be allocated for these videos' propagation to benefit the potential downloaders. Our edge-cloud replication algorithm is illustrated in Algorithm 1.

Algorithm 1 Edge-Cloud Replication Algorithm.

procedure VIDEO AND REGION SELECTION
  $\mathcal{V}^{(T)} \leftarrow \Phi$
  if v is newly published **then**
    $\mathcal{V}^{(T)} \leftarrow \mathcal{V}^{(T)} \cup \{v\}$
    $r_v \leftarrow \arg\min_{r \in \mathcal{R}} \sum_{i \in \mathcal{F}_v} d_{r,i}$
  else
    if $v \in \mathcal{W}^{(T)}$ **and** $g_v^{(T)} > \theta_v^{(T)}$ **then**
      $\mathcal{V}^{(T)} \leftarrow \mathcal{V}^{(T)} \cup \{v\}$
      $r_v \leftarrow \arg\min_{r \in \mathcal{R} - \mathcal{R}_v^{(T)}} \sum_{i \in \bigcup_{k \in \mathcal{L}_v^{(T)}} \mathcal{F}_k} d_{r,i}$
    end if
  end if
end procedure
procedure BANDWIDTH RESERVATION
  for all $v \in \mathcal{V}^{(T)}$ **do**
    if v is replicated at region r_v **then**
      $b_{v,r_v} \leftarrow B_{r_v} e_v^{(T)} / \sum_{v \in \mathcal{V}_{r_v}} e_v^{(T)}$
    end if
  end for
end procedure

5.2.4 Reduction of Replications

In our measurement study, we have shown the temporal locality of the social video propagation, *i.e.*, after a period of time since its publication, a video content will not be able to attract as many users as before. Though the bandwidth reservation can adapt to reduce the upload capacity allocated for a video that becomes less popular, the replications of the video still occupy the storage at edge-cloud servers. Thus, we need to reduce a video's replications on the edge-cloud servers to make room for new videos generated by users in the system. We let each edge-cloud server determine the replication reduction as follows: (1) the region of a video's initial replication acts as a permanent backup of the video; (2) other regions of a video's extended replications locally dump the video contents according to their geographic influence index, *i.e.*, videos with a smaller geographic influence index are more likely to be removed from an edge-cloud server to make room for new ones.

5.3 Peer-Assisted Replication

In Sec. 4, we have justified that the unique propagation pattern makes it very promising to utilized the peer-assisted paradigm to allocate certain amount of resource from the users to replicate the video contents, and peers (users) can serve their social neighbors with good Internet connectivity. In our peer-assisted replication, we assume users download video contents according to their own demands, and we design the social-aware cache replacement strategy for peers to determine which videos are cached to help other users, since peers' cache strategy can greatly affect the performance of a P2P system [19].

5.3.1 Social-Aware Cache Replacement

In PSAR, a peer locally performs the cache replacement using not only the perceived video popularity, but also the local social factors. In particular, the following information is used at peer i: (1) the local popularity which is the number of requests for video v received by peer i, denoted as p_i^v; (2) the fraction of peer i's friends that can join the propagation tree of video v, denoted as f_i^v. f_i^v is calculated by historical records for different video categories, *i.e.*, peer i keeps a record of the fraction of friends that have been attracted in each category in the history; and (3) the time lag between the propagation tree is constructed and the time when the peer re-shares the video, *i.e.*, $\tau_v^{(T)}$.

Social Influence Index. Based on the social propagation, we design a social influence index as follows:

$$q_v = z_s(\tau_v^{(T)})(p_i^v f_i^v).$$

In the peer-assisted replication, videos with smaller social influence index are more likely to be dumped by the peer. The rationale of the social influence index is that larger $p_i^v f_i^v$ indicates that peer i can potentially attract more users to re-share video v from its friends in the future, and $\tau_v^{(T)}$ is utilized to reflect the temporal locality. Large social influence index indicates that the video can be potentially downloaded by more local friends, and the peer should keep it to serve these friends. Thus, in our cache replacement algorithm, the peer will try to dump videos with the smallest q_v's until the capacity is enough for new videos.

5.3.2 Collaboration with Edge-Cloud Servers

In the distribution of social video contents, edge-cloud servers compensate the upload capacities when peers are not able to fully serve all the requests. The collaboration between the edge-cloud servers and the peers in PSAR is based on the replication indices as follows: (1) for the popular videos with large geographic influence and content propagation indices, they can be replicated to many regions to serve different users, who can download the contents from either the neighboring peers or the local edge-cloud servers; (2) for the unpopular videos, they are usually not widely replicated by the edge-cloud servers because their geographic influence index is small. However, a large fraction of such videos are shared in small social groups according to our measurement results, in which users can download the videos from their local friends, according to the social influence index. In both cases, users can achieve a good local download quality.

6. PERFORMANCE EVALUATION

6.1 Experiment Setup

Based on the same traces used in our measurement, we select 9318 videos from the original traces in the last 10 days for our experiments. These videos propagate among the regions captured by Tencent Weibo. Peers are located in the regions according to their profiles, and an edge-cloud server is deployed in each region. In our experiments, we assume the records of imports and re-shares indicate users' downloads of these videos. Thus, these records are used to drive users' downloads in the simulation. We also assume the user-generated videos have the same short duration [10], and we let the replication unit be a whole video for both servers and peers. We normalize the geographic distance between peers and servers in the evaluation. In the peer-assisted replication, peers exchange their cache states with socially connected neighbors periodically, so that they are aware of what can be downloaded from these social neighbors. When downloading from other peers, a tracker server is employed to help peers find each other. A peer downloads a video according to the following rules: (1) It first tries to download the video from neighboring peers, where peers that are socially connected in the same region are prioritized; (2) If no peer is able to serve the video, it will resort to the edge-cloud servers in the same region; (3) If the local servers are not able to serve it, it will try other servers with the smallest geographic distances.

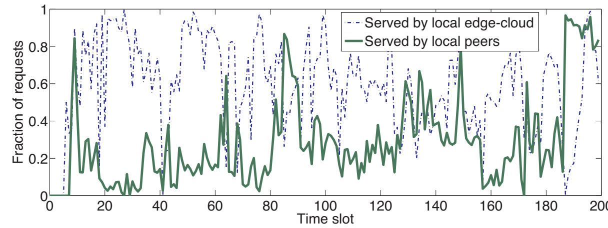

Figure 13: Number of requests served by local edge-cloud servers and local peers.

We first show the performance of PSAR over time. Fig. 13 illustrates the fraction of requests served by their local edge-cloud servers over all server-served requests and the fraction of requests served by local peers over all peer-served requests, respectively. We observe that our replication can reach relatively high levels of requests that are served by local edge-cloud servers and local peers (58.7% and 28.5%, respectively). Meanwhile, in PSAR, we observe that the local edge-cloud servers and peers can compensate each other to serve the users over time. Next, we will evaluate the detailed performance of PSAR.

6.2 Efficiency of Edge-Cloud Replication

In the edge-cloud replication, we compare PSAR with the following algorithms that are widely used in real-world video service systems. (1) A popularity-based replication, where videos are prioritized to be replicated or removed according to the videos' historical popularity, *i.e.*, the number of total imports and re-shares in the recent period. The videos selected for replication are assigned to regions so that the load (overall popularity of videos) can be balanced among the regions. In each region, the edge-cloud server allocates upload bandwidth for a video proportionally to its recent popularity. (2) A random approach where videos are replicated randomly in different regions and reserved with a random amount of upload bandwidth. Note that these two algorithms are also executed periodically in each time slot.

Fraction of Locally Served Requests. We first evaluate how many requests of videos can be served by local servers using different replication algorithms. We define a local download ratio as the fraction of requests that are served by users' local servers, *i.e.*, servers in the same geographic region with the users issuing the requests. Fig. 14 illustrates the local download ratio versus the average upload capacity allocated at an edge-cloud server. We observe that our edge-cloud replication in PSAR can greatly improve the local download ratio, and as the available server bandwidth capacity grows, the local download ratio in PSAR increases faster than the popularity-based and random algorithms, indicating that users can benefit more from increased server resources in PSAR.

Normalized Download Geo-Distance from Servers. We also evaluate the normalized download distance, which is define as the average normalized geographic distance between the users and the servers from which they download the videos. Fig. 15 illustrates the normalized download distance versus the average server capacity. We observe that PSAR

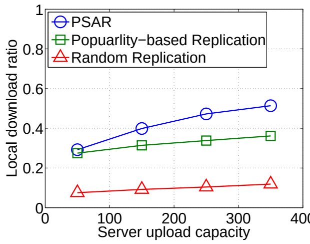

Figure 14: Local download ratio versus the server capacity.

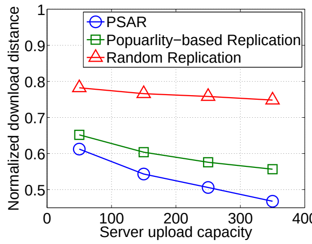

Figure 15: Normalized download distance versus the server capacity.

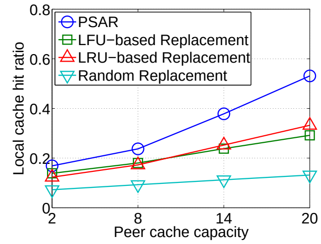

Figure 18: Local cache hit ratio versus peer's capacity.

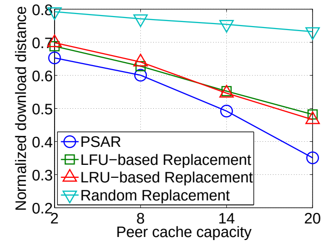

Figure 19: Normalized download distance versus peer's capacity.

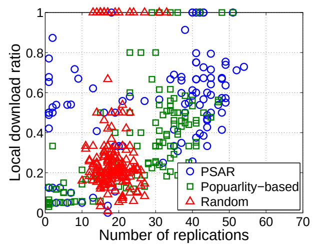

Figure 16: Local download ratio of a video versus the number of its replications.

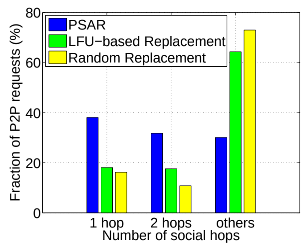

Figure 17: Peer contribution versus the number of social hops between the pair of peers.

achieves a smaller download distance than the other two algorithms. The reason is that by inferring the geographic influence index, the content propagation index and users' social connections, better prediction of a video's propagation range can be utilized to perform the region selection. Similarly, we observe that the normalized download distance in PSAR decreases faster than other algorithms when server capacity increases.

Number of Replications. We further study the impact of the number of replications of a video on the service quality. Fig. 16 compares the local download ratio of a video in the three strategies in terms of different numbers of video replications. We observe that in the random replication, all the videos have the similar number of replications — this is the reason for its inefficiency for contents that are either very popular or only propagated among small social groups. The replication number in the popularity-based replication is similar to that in PSAR; however, PSAR is more effective to replicate videos that are propagated in small social groups and the ones that are highly propagating across many regions.

6.3 Efficiency of Peer-Assisted Replication

We also evaluate the efficiency of the peer-assisted replication. We compare PSAR with (1) an LFU-based peer cache replacement algorithm where videos least requested recently (a reference time window of 24 hours) are dumped to make room for new ones, (2) an LRU-based cache replacement algorithm where videos that have not been recently requested are dumped, and (3) a random replacement algorithm where randomly selected videos are dumped.

Local Cache Hit Ratio. We first evaluate the local cache hit ratio, which is defined as the fraction of videos that can be directly downloaded from the socially connected peers. Higher local cache hit ratio indicates better local download performance, since we have already justified that peers which are socially connected to each other are also geographically close to each other, resulting in a better Internet connectivity. Fig. 18 illustrates the local cache hit ratio versus the storage capacity at each peer (the number of videos that can be stored). We observe that our design significantly outperforms other strategies. As the cache capacity increases, the cache hit ratio in our design improves much faster than other algorithms. The reason for the inefficiency of LFU and LRU is that many unpopular videos cannot be efficiently cached according to users' historical requests; while they can be addressed in our design where peers actively cache them for their friends based on the social influence index. We also observe that LFU and LRU have achieved the similar ratio, because LFU with a small reference time window acts similar to LRU, since requests are distributed close to the time when the user generates the content in the online social network.

Normalized Download Geo-Distance from Peers. We also evaluate the normalized geographic distance between the neighboring peers. Fig. 19 illustrates the average normalized download distance between peers who upload videos to each other versus the cache storage capacity at each peer. We observe that our social-aware cache replacement achieves a much smaller geographic download distance than the other algorithms, meaning that a peer is more likely to find a close neighbor to download the videos from, thereby achieving a better Internet connectivity for both sides. We also observe that when a large cache capacity is allocated at a peer, our design benefits more than other algorithms.

Social-Aware Contribution. We further investigate the P2P networks by studying which type of peers upload contents to the users. Fig. 17 illustrates the fraction of requests served by peers versus the number of social hops between the pair of peers. We observe that in our design, much more requests are served by their direct friends or two-hop friends in the social network, so that more local peers can be used to upload the contents. The reason is that in PSAR, videos are cached according to not only the requests of users, but also the level of friends that can be influenced by the video in the future.

7. CONCLUDING REMARKS

This paper addresses the challenges in the replication of social video contents, resulting from the massive number of the videos generated by users and the close-to-uniform popularity distribution. By conducting extensive measurement of traces obtained from a representative online social network system, we observe unique characteristics, which demonstrate social, geographical and temporal localities in the propagation. Based on the social propagation characteristics in the propagation, we propose a propagation-based social-aware replication strategy to serve such social video contents to users. Specifically, we design three replication indices: a geographic influence index, a content propagation index and a social influence index, which can guide the region selection, bandwidth reservation and cache replacement in the joint edge-cloud and peer-assisted replication. Extensive experiments driven by the real-world traces further demonstrate the effectiveness and superiority of our design.

Acknowledgment

We would like to thank Chong Yuan from Tsinghua University to conduct the experiments on the peer-assisted replication, and the Tsinghua-Tencent Joint Lab to provide the traces used in our research. This work is supported by the National Basic Research Program (973) under Grant No. 2011CB302206, the National Natural Science Foundation of China under Grant No. 60833009 and 60933013, and the Tsinghua National Laboratory for Information Science and Technology under Grant No. 916-042003091.

8. REFERENCES

[1] http://en.wikipedia.org/wiki/clustering_coefficient.

[2] http://t.qq.com.

[3] http://www.foreseeresults.com.

[4] http://www.youtube.com.

[5] V. Adhikari, S. Jain, Y. Chen, and Z. Zhang. Reverse Engineering the YouTube Video Delivery Cloud. In *Proc. of IEEE Hot Topics in Media Delivery Workshop*, 2011.

[6] V. Adhikari, S. Jain, and Z. Zhang. Where Do You Tube? Uncovering YouTube Server Selection Strategy. In *Proc. of IEEE International Conference on Computer Communications and Networks (ICCCN)*, 2011.

[7] M. A. Balachander Krishnamurthy, Phillipa Gill. A Few Chirps About Twitter. In *Proc. of ACM WOSN*, 2008.

[8] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida. Characterizing User Behavior in Online Social Networks. In *Proc. of ACM IMC*, 2009.

[9] M. Cha, A. Mislove, and K. P. Gummadi. A Measurement-driven Analysis of Information Propagation in the Flickr Social Network. In *Proc. of ACM WWW*, 2009.

[10] X. Cheng, C. Dale, and J. Liu. Statistics and Social Network of YouTube Videos. In *Proc. of IEEE IWQoS*, 2008.

[11] X. Cheng and J. Liu. Load-Balanced Migration of Social Media to Content Clouds. In *Proc. of ACM NOSSDAV*, 2011.

[12] P. Dodds and D. Watts. A Generalized Model of Social and Biological Contagion. *Journal of Theoretical Biology*, 232(4):587–604, 2005.

[13] J. Hartline, V. Mirrokni, and M. Sundararajan. Optimal Marketing Strategies over Social Networks. In *Proc. of ACM WWW*, 2008.

[14] B. Huffaker, M. Fomenkov, D. Plummer, D. Moore, and K. Claffy. Distance Metrics in the Internet. In *Proc. of IEEE International Telecommunications Symposium (ITS)*, 2002.

[15] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the Spread of Influence Through a Social Network. In *Proc. of ACM SIGKDD*, 2003.

[16] H. Kwak, C. Lee, H. Park, and S. Moon. What Is Twitter, a Social Network or a News Media? In *Proc. of ACM WWW*, 2010.

[17] H. Li, H. Wang, and J. Liu. Video Sharing in Online Social Network: Measurement and Analysis. In *Proc. of ACM NOSSDAV*, 2012.

[18] Y. Liu, Y. Guo, and C. Liang. A Survey on Peer-To-Peer Video Streaming Systems. *Peer-to-peer Networking and Applications*, 1(1):18–28, 2008.

[19] J. Luo, Q. Zhang, Y. Tang, and S. Yang. A Trace-Driven Approach to Evaluate the Scalability of P2P-Based Video-On-Demand Service. *IEEE Transactions on Parallel and Distributed Systems*, 20(1):59–70, 2009.

[20] A. Mislove, M. Marcon, K. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and Analysis of Online Social Networks. In *Proc. of ACM IMC*, 2007.

[21] A. Mislove, B. Viswanath, K. Gummadi, and P. Druschel. You Are Who You Know: Inferring User Profiles in Online Social Networks. In *Proc. of ACM WSDM*, 2010.

[22] K. Nguyen, C. Pham, D. Tran, and F. Zhang. Preserving Social Locality in Data Replication for Social Networks. In *Proc. of IEEE ICDCS Workshop on Simplifying Complex Networks for Practitioners*, 2011.

[23] G. Peng. CDN: Content Distribution Network. *Arxiv preprint cs/0411069*, 2004.

[24] J. M. Pujol, V. Erramilli, G. Siganos, X. Yang, N. Laoutaris, P. Chhabra, and P. Rodriguez. The Little Engine(s) That Could: Scaling Online Social Networks. *SIGCOMM Comput. Commun. Rev.*, 40:375–386, Aug. 2010.

[25] S. Scellato, C. Mascolo, M. Musolesi, and V. Latora. Distance Matters: Geo-Social Metrics for Online Social Networks. In *Proc. of Conference on Online Social Networks*. USENIX Association, 2010.

[26] D. Tran, K. Nguyen, and C. Pham. S-CLONE: Socially-Aware Data Replication for Social Networks. *Computer Networks*, 2012.

[27] Z. Wang, L. Sun, C. Wu, and S. Yang. Guiding Internet-Scale Video Service Deployment Using Microblog-Based Prediction. In *Proc. of IEEE INFOCOM Mini-Conference*, 2012.

[28] Z. Wang, L. Sun, S. Yang, and W. Zhu. Prefetching Strategy in Peer-assisted Social Video Streaming. In *Proc. of ACM Multimedia*, 2011.

[29] Y. Wu, C. Wu, B. Li, L. Zhang, Z. Li, and F. C. Lau. Scaling Social Media Applications into Geo-Distributed Clouds. In *Proc. of IEEE INFOCOM*, 2012.

[30] D. Xu, S. Kulkarni, C. Rosenberg, and H. Chai. Analysis of a CDN–P2P Hybrid Architecture for Cost-Effective Streaming Media Distribution. *Multimedia Systems*, 11(4):383–399, 2006.