# Propagation-Based Social-Aware Multimedia Content Distribution

ZHI WANG, WENWU ZHU, XIANGWEN CHEN, LIFENG SUN, JIANGCHUAN LIU, MINGHUA CHEN, PENG CUI AND SHIQIANG YANG

Online social network has reshaped the way how multimedia contents are generated, distributed and consumed on today's Internet. Given the massive number of user-generated contents shared in online social networks, users are moving to directly access these contents in their preferred social network services. It is intriguing to study the service provision of social 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 social content propagation. We have observed important propagation patterns, including social locality, geographical locality and temporal locality. Motivated by the measurement insights, we propose a propagation-based social-aware delivery framework using a hybrid edge-cloud and peer-assisted architecture. We also design replication strategies for the architecture, based on three propagation predictors designed by jointly considering user, content and context information. In particular, we design a propagation region predictor and a global audience predictor to guide how the edge-cloud servers backup the contents, and a local audience predictor to guide how peers cache the contents for their friends. Our trace-driven experiments further demonstrate the effectiveness and superiority of our design.

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

General Terms: Measurement, Design

Additional Key Words and Phrases: Social network, video service

**ACM Reference Format:**

Wang, Z., 2013. Propagation-Based Social-Aware Multimedia Content Distribution. *ACM Trans. Appl. Percept.* 2, 3, Article 1 (May 2013), 20 pages.
DOI:http://dx.doi.org/10.1145/0000000.0000000

This work was partially supported by the National Basic Research Program (973) under Grant No. 2011CB302206, the National Natural Science Foundation of China under Grant No. 60933013, 61272231 and 61003097, the "Thousand Talents Program" under Grant No. 553360001, the MOST International Cooperative Project under Grant No. 2013DFG12870, and the Tsinghua-Tencent Joint Lab.
Z. Wang*, W. Zhu*, L. Sun*, P. Cui and S. Yang are with the Department of Computer Science and Technology, Tsinghua University, Tsinghua National Laboratory for Information Science and Technology, and Beijing Key Laboratory of Networked Multimedia, Beijing, China (email: wangzhi04@mails.tsinghua.edu.cn, {wwzhu, sunlf, cuip, yangshq}@tsinghua.edu.cn).
X. Chen and M. Chen are with the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong (email: cxw1987@gmail.com, minghua@ie.cuhk.edu.hk).
J. Liu is with the School of Computing Science, Simon Fraser University, Canada (email: jcliu@cs.sfu.ca).


DOI:http://dx.doi.org/10.1145/0000000.0000000

## 1. INTRODUCTION

Recent years have witnessed the blossom of online social network service (*e.g.*, Facebook, Twitter) and online content sharing service (*e.g.*, YouTube, Flickr), as well as the rapid convergence of both services. *Social multimedia contents*, or *social contents* 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 online video contents [Foresee 2012]. It is fascinating to study how the social 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 how the shared contents can reach users in the online social network [Kwak et al. 2010]. The unique *propagation* properties make the content access pattern in the online social network quite different from that in the traditional centralized content sharing systems, in that (1) 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 contents among the users.

We are facing the following challenges in distributing the social contents with satisfactory QoE: (1) A huge number of user-generated contents (UGC) require a large amount of storage and network resource, *e.g.*, YouTube has hit a new record of $72$ hours' worth of videos uploaded by users per minute [YouTube 2013]; (2) Newly generated contents are the ones that tend to attract most of the users, but it is difficult to estimate their popularity for the service allocation, which is dynamically affected by the social network [Cha et al. 2009]; (3) Social contents have close-to-uniform [Mislove 2012] and highly-volatile popularity profiles, because a large portion of the contents are shared among small social groups (*e.g.*, family members).

Challenge (1) makes traditional service paradigms (*e.g.*, C/S based on private servers) not suitable — it is too expensive to replicate all contents to all servers, and a common practice to provision these content services is to replicate contents to severs at different geographic regions [Adhikari et al. 2011] by allocating resource from the geo-distributed CDN (Content Devilry Network) or cloud, where contents 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 contents with skewed and stable popularity profiles, not suitable in the context of online social network. Mislove *et al.* [Mislove et al. 2010] have observed a large deduction of cache hit ratio when traditional caching schemes are used to replicate social contents.

Our preliminary study [Wang et al. 2012] reveals a key observation that the social contents, unlike regular contents, do not propagate among users randomly — instead, they propagate along the social-network topology according to several rules determined by the social propagation. We now significantly extend our study to design a general framework with propagation patterns and predictions incorporated, based on which we develop a social-aware delivery system to effectively distribute social contents with superb QoE. In this paper, we use the most representative type of multimedia content — the social video content to study how the social contents can be effectively replicated based on social propagation — our design though can be used in the delivery of a variety of multimedia types.

*First*, we demonstrate that the statistic information obtained from the online social network can guide the video replication. We conduct large-scale measurement studies to explore the connection between the social propagation and the replication, and discover the propagation patterns 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 newly published videos can attract most of the viewers.

*Second*, based on the propagation patterns, we design propagation predictors to guide the content delivery. In particular, the propagation region predictor, global-audience predictor and local-audience predictor answer the following questions, respectively: (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?

*Third*, based on the propagation patterns (*i.e.*, localities in propagation), we propose a propagation-based social content distribution framework, in which a hybrid edge-cloud [Zhu et al. 2011] and peer-assisted video replication architecture is employed. Based on the propagation predictions, videos are replicated by both the edge-cloud servers and peers at different geographic locations as follows: (1) We design the edge-cloud replication strategies according to the region predictor and global-audience predictor, determining the region selection and bandwidth reservation; (2) We further design the peer-assisted replication according to the local-audience predictor, 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 measurement studies on social content propagation. We present the propagation pattern-based architecture in Sec. 4, and the detailed replication strategies based on propagation predictions in Sec. 5. In Sec. 6, we evaluate the performance of our design by trace-driven experiments. In Sec. 7, we discuss some implementation issues. Finally, we conclude the paper in Sec. 8.

## 2. RELATED WORK

### 2.1 Propagation in Online Social Network

Online social network has become a popular Internet service. Based on traces from Flickr, LiveJournal and Orkut, Mislove *et al.* [Mislove et al. 2007] 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.* [Balachander Krishnamurthy 2008] 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.* [Kwak et al. 2010] investigate the impact of users' retweets on information diffusion in Twitter. Dodds *et al.* [Dodds and Watts 2005] 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.* [Kempe et al. 2003] investigate how to maximize the spread of influence in an online social network, and Hartline *et al.* [Hartline et al. 2008] utilize such maximum spread to achieve revenue maximization. Domingos *et al.* [Domingos and Richardson 2001] explore the value of social networks in estimating potential buyers of a product or a service, which can be influenced by an existing customer. Kempe *et al.* [Kempe et al. 2003] investigate how to maximize the spread of influence in an online social network, and Hartline *et al.* [Hartline et al. 2008] utilize such maximum spread to achieve revenue maximization, *i.e.*, the largest number of buyers can be attracted, by selecting the best set of initial users to push the information onto.

In this paper, we will study how to connect the social propagation and the social multimedia content replication, *i.e.*, how statistic information about the content propagation can be utilized to guide the content replication.

### 2.2 Social Content Replication

Many architectures have been proposed in large-scale content service systems, including (1) the server-based architecture, *e.g.*, CDN and cloud-based approaches [Peng 2004], (2) the client-based architecture, *e.g.*, the P2P content distribution [Liu et al. 2008], and (3) the hybrid architecture, *e.g.*, a hybrid CDN and P2P distribution framework [Xu et al. 2006]. For Internet-scale social content service, replicating the content at different geographic regions is a promising approach to provide good service quality to users [Adhikari et al. 2011]. Zhu *et al.* [Zhu et al. 2011] propose to allocate cloud servers at the network edges to distribute the multimedia contents to users.

*Content-based replication*. From the content aspect, traditional content distribution mainly takes content popularity into account, and allocates storage and bandwidth resource according to the popularity [Kangasharju et al. 2002]. After 2005, a huge number of contents are generated by users. Exploring the correlation between contents can effectively help users fetch the contents precisely [Cheng and Liu 2009]. However, online social network has greatly changed the assumptions in traditional replication algorithms [Benevenuto et al. 2009], *e.g.*, the distribution of contents is shifted from a "central-edge" manner to an "edge-edge" manner, resulting in the close-to-uniform popularity distribution. Li *et al.* [Li et al. 2012] study the content 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 contents, some social-aware content replications have been proposed.

*Social-based replication*. After online social network is widely used by people to access online contents. User relation and influence are studied, to reflect that after a content is shared by a person, it may be requested by his friends, so that contents can be distributed based on this inference. Pujol *et al.* [Pujol et al. 2010] 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.* [Tran et al. 2012] study the partition of contents in the online social network by taking social relationships into consideration. Nguyen *et al.* [Nguyen et al. 2011] study how to improve the system efficiency in case of server failures by taking social locality into consideration. Wang *et al.* [Wang et al. 2012] observe that a social network can be used to help predict the content access pattern in a standalone on-demand system. Wu *et al.* [Wu et al. 2012] study how to minimize the cost in social media migration among servers at different regions. Cheng *et al.* [Cheng and Liu 2011] study the social media partition to balance the server load and preserve the social relationship.

Understanding the content access patterns is a key to perform effective content replication. With respect to understanding how users access contents in the online social network, related works have been considering contents and users separately. However, content sharing is dynamically determined by both the contents and users. In order to address the problem of providing effective content distribution to users in the social network, we propose propagation-based social content distribution. This paper is an extension of our preliminary studies in [Wang et al. 2012] — we design a new general framework for propagation-based social content delivery, incorporating both the propagation patterns and propagation predictions.

## 3. MEASUREMENT OF SOCIAL CONTENT PROPAGATION

In this section, we present the logic framework of our study, our measurement setup and a glance of the propagation of social multimedia contents.

### 3.1 Framework

Fig. 1 illustrates the logic framework of our design. (1) *Input*. User, content and context information is the input to our design. In particular, we have the social relationship and influence between users, their preference of different contents, the content popularity and the server and user regional distributions, determining which users prefer to receive contents from which servers. (2) *Propagation pattern mining*. Based on the input, we have observed the propagation patterns, including the social locality, geographical locality and temporal locality. (3) *Propagation prediction*. Based on these propagation pat-

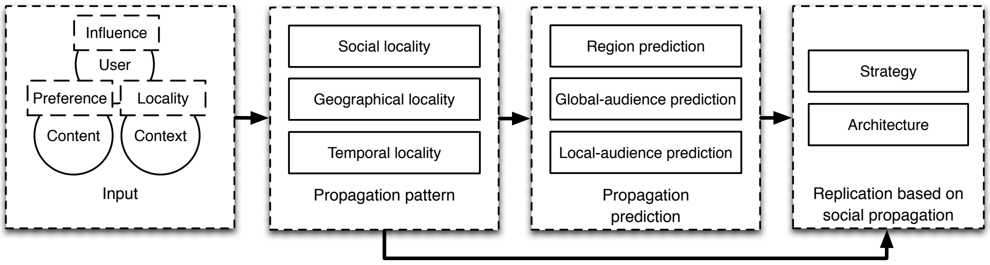



Fig. 1. Logic framework.

terns, we further design predictors for the propagation regions (*i.e.*, where a video will be shared), the global audience (*i.e.*, the global popularity of a video), and the local audience (*i.e.*, how a video is shared between friends). (4) *Propagation-based replication*. Our replication architecture and strategies are designed according to the propagation patterns and the propagation predictions, respectively. Edge-cloud servers and peers are allocated to strategically serve the contents so that users can download from the local servers.

## 3.2 Measurement Setup

In our measurement, we have collected traces from Tencent Weibo [Tencent Weibo 2013], which is a microblogging website, where users can broadcast a message including at most $140$ characters to their friends. 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 is 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, which are posted or re-shared by $1,465,328$ users, from more than $200$ regions in the propagation (each region is defined by Tencent as a geographical area). 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. 2 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.

## 3.3 A Glance of Generation, Distribution and Popularity of Social Contents

*Content 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. 3 illustrates the number of imports and re-shares of the targeted videos over time. We observe that (1) more users are generating videos instead

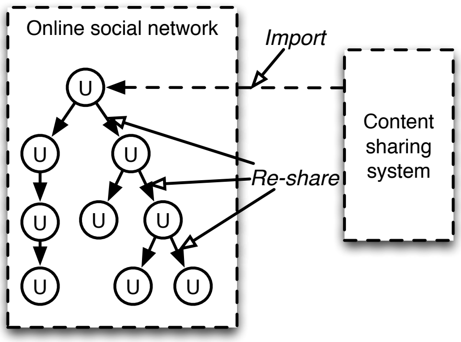



Fig. 2. Connection between an online social network and a content sharing system.

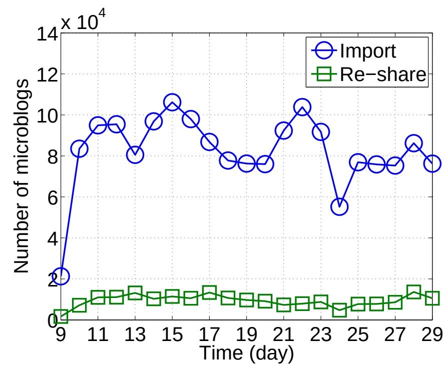



Fig. 3. Imports and re-shares over time.

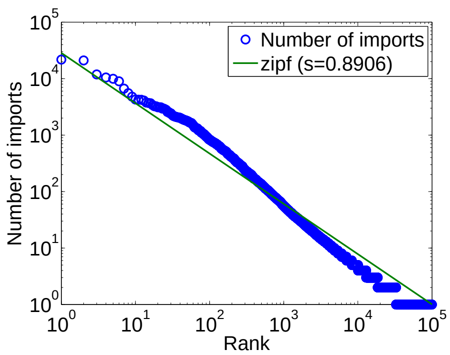



Fig. 4. Number of imports of a video versus its rank.

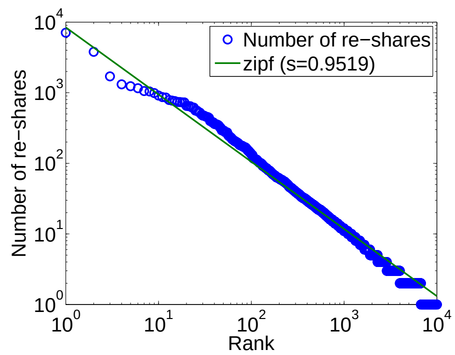



Fig. 5. Number of re-shares of a video versus its rank.

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. 4, 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. 5 illustrates the number of re-shares versus the rank of the video, and we observe that 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 (*e.g.*, users can download the videos from the servers located at the same region), with limited storage and network resource.

*Social groups*. We further investigate in which types of social groups these unpopular videos are propagating. By randomly collecting $50$ videos with different propagation size (the number of users

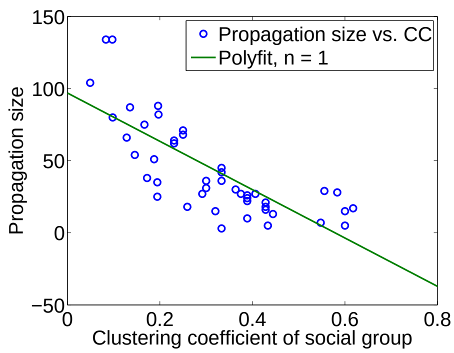



Fig. 6. Propagation size versus the clustering coefficient of social group.

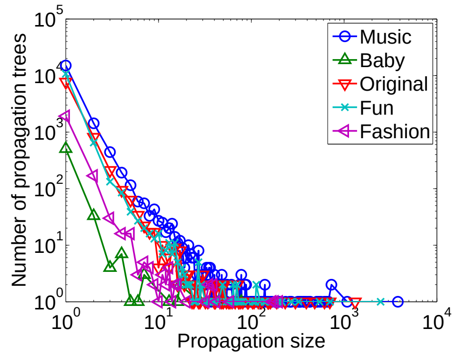



Fig. 7. Number of propagation trees versus the propagation size.

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. 6, each sample illustrates the video propagation size versus the clustering coefficient of the corresponding social group [Wiki: Clustering/Coefficient 2013]. 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.

## 4. REPLICATION ARCHITECTURE BASED ON PROPAGATION PATTERNS

In this section, we first present the propagation patterns observed in the measurement studies. Then, we present the architecture of our design and its key components, based on the propagation patterns.

### 4.1 Social Propagation Patterns

We study the patterns of the social content propagation to guide our social content delivery architecture design.

4.1.1 *Social Locality*. 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 a new leaf node in the propagation tree. Fig. 7 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.*, the size of over $90\%$ of the propagation trees is 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. 8 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 same 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.

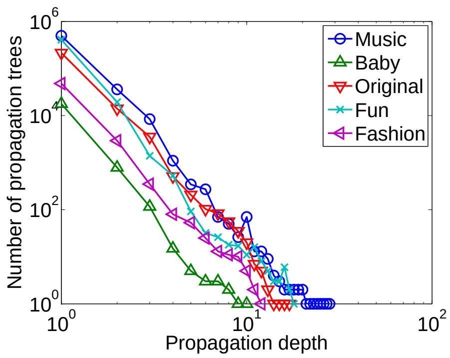



Fig. 8. Number of propagation trees versus the propagation depth.

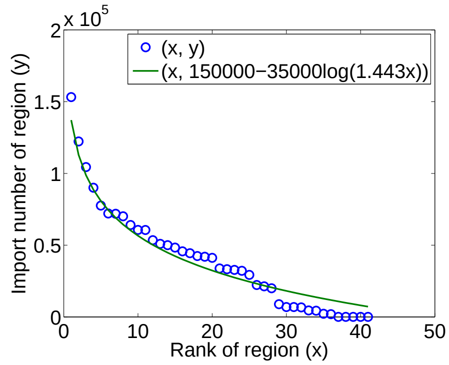



Fig. 9. Import number of a region versus the rank of the region.

4.1.2 *Geographical Locality*. 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. 9, 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.

*Second*, we observe that most of the contents are propagating in a very small number of regions. In Fig. 10, each sample represents the number of regions involved in a content's propagation versus the rank of the content. We observe that most of the videos are shared by users from only a very few number of regions. More than $90\%$ of the contents are propagating in less than $5$ regions. This indicates that in social video sharing, users in a video's propagation can be geographically close to each other.

4.1.3 *Temporal Locality*. 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. 11 illustrates the number of re-shares of a video in a timeslot ($1$ hour) versus the time lag since the propagation tree 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 $s = 1.5070$. More than $95\%$ of the re-shares happen within the first 24 hours. This indicates that in social video sharing, users' behaviors are highly crowded around the time point when it is imported. We will also incorporate the temporal locality into our design.

### 4.2 Edge-Cloud and Peer-Assisted Replication Architecture

According to our observations, users in a video propagation are socially and geographically close to each other, and their social actions are crowed in a short period. Accordingly, we propose to use a *hybrid edge-cloud and peer-assisted* architecture for social video distribution, 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. 12 illustrates the conceptual

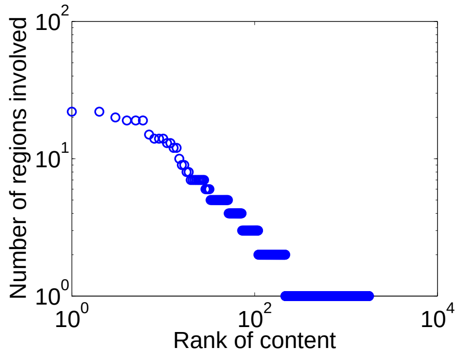



Fig. 10. Number of regions involved versus the rank of content.

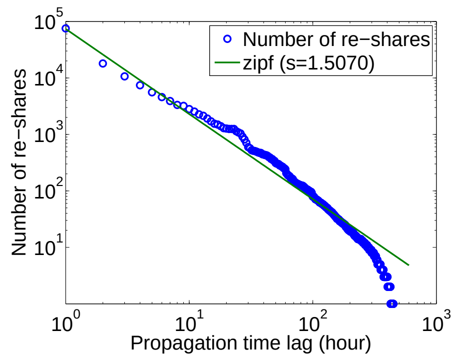



Fig. 11. Number of re-shares versus the time lag.

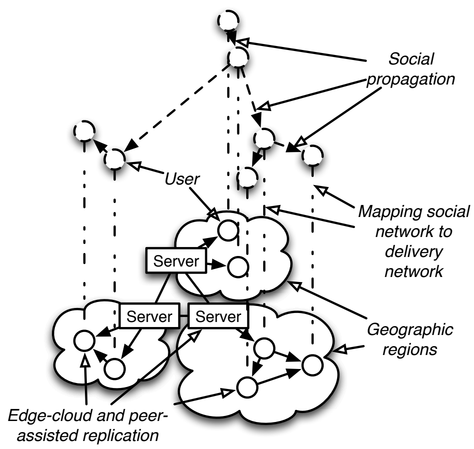



Fig. 12. Conceptual architecture of PSAR.

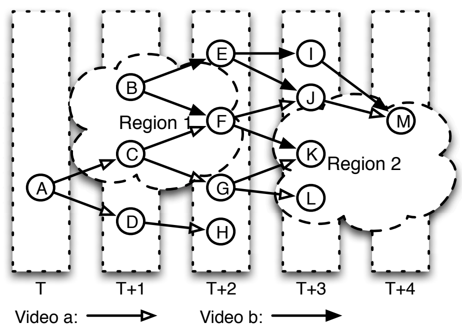



Fig. 13. Resource allocation for two propagation trees.

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* (Propagation-Based Social-Aware Replication for social contents), 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.

4.2.1 *Edge-Cloud Replication.* 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 [Adhikari et al. 2011].

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 local-audience 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.* 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 [Scellato et al. 2010] — these socially connected users tend to have good Internet connectivity between each other to perform the peer-assisted video download [Huffaker et al. 2002].

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 local-audience 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. 5.2.3.

## 5. REPLICATION STRATEGIES BASED ON PROPAGATION PREDICTION

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, using the propagation prediction. After that, we present the detailed design of PSAR based on the connection.

*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. 13 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 timeslot $T$, then the video is re-shared by his friends $C$ and $D$ in timeslot $T+1$. 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 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 timeslot $T+3$. 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.

### 5.1 Propagation Prediction

5.1.1 *Propagation Region Prediction.* Based on the dataset used in our measurement studies, Fig. 15 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. In PSAR, the propagation size is utilized to determine whether a video will be replicated to more regions. In particular, we design a geographic influence index as follows:

$$g_v^{(T)} = c_1 \log(c_2 s_v^{(T-1)}),$$

where $s_v^{(T-1)}$ is the propagation size of the propagation tree of video $v$ in timeslot $T-1$. Large $g_v^{(T)}$ indicates that more regions will be involved in the propagation of the video. 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.

5.1.2 *Global-Audience Prediction.* In order to allocate bandwidth to serve a social video content, we design a global-audience predictor, based on a global-audience index to evaluate the strength of a video's propagation in timeslot $T$, using 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 global-audience index is defined as follows:

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

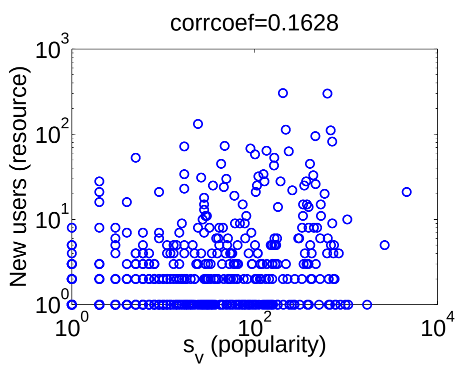
where $z_s(\tau_v^{(T)})$ is a decreasing function to make use of the temporal locality, which can adjust the global-audience index according to $\tau_v^{(T)}$ so that more recently generated or shared videos will have a larger global-audience index. Based on our observation in Sec. 4.1.3, $z_s(t)$ is defined as follows:

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

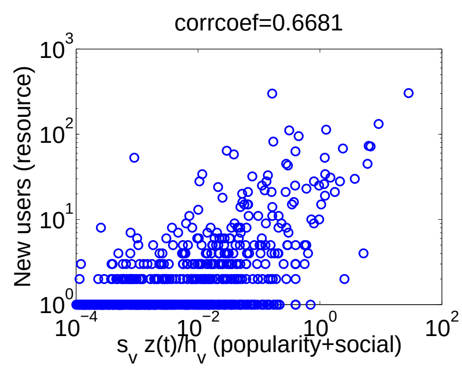
where $s$ is the zipf shape parameter and $N$ is the number of hours between the publication time of the earliest video and the publication time of the latest video. In our design, $e_v^{(T)}$ will be used to guide the replication. Larger $e_v^{(T)}$ indicates that more users can join the propagation tree in timeslot $T$. 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 reach more users; (3) According to the temporal locality, large $\tau_v^{(T)}$ slows down the propagation. Based on the global-audience index, we will determine how much bandwidth we will reserve for a video in the future timeslot in PSAR.

Fig. 14 compares our social-aware global-audience prediction and the traditional popularity estimation using only the historical popularity. The effectiveness of our global-audience prediction is verified by our dataset. In Fig. 14(a), each sample is a content's current popularity versus its popularity in the previous timeslot. We observe that the video's global audience are highly violating overtime, with a very small correlation between the current audience number and previous number. Our prediction is illustrated in Fig. 14(b). Each sample represents the current popularity versus the global-audience index of the previous timeslot. We observe that after incorporating the propagation patterns, the correlation coefficient is increased by $4$ times, indicating that the popularity can be better predicted by our design.

5.1.3 *Local-Audience Prediction.* In our architecture, a peer locally performs the cache replacement using not only the perceived video popularity, but also the local social factors. In order to determine which videos should be kept for peering neighbors, we design a local-audience predictor, based on the




(a) New popularity versus historical popularity.




(b) New popularity versus the global-audience index.

Fig. 14. Global-Audience Prediction.

following information 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 time when propagation tree is constructed and the time when the peer re-shares the video, *i.e.*, $\tau_v^{(T)}$. Based on the social propagation patterns, we design a local-audience index to perform the prediction as follows:

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

In the peer-assisted replication, videos with a smaller local-audience index are more likely to be dumped by the peer. The rationale 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.

The effectiveness of the local-audience predictor is verified by our data as well. Fig. 16 illustrates the CDF of the correlation coefficient between a friend's video category preference (calculated as a category preference vector) at time $T$ and the category preference at time $T-1$. We observe that most of the friends' preference can be inferred from their historical preference. In our dataset, the correlation coefficient for $80\%$ of the user preference in two consecutive timeslots can be larger than $0.8$.

### 5.2 Replication Strategies based on Propagation Prediction

5.2.1 *Region Selection in Edge-Cloud Replication.* Based on the region prediction and the global-audience prediction, 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. When performing video replication, we need to find out the videos that may propagate to more regions in the future. We use the geographic influence index in the region prediction for that. 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. Parameters $c_1$ and $c_2$ are selected according to the measurement. Based on the geographic influence index, we can predict whether the regions where the video has been replicated are enough.

*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

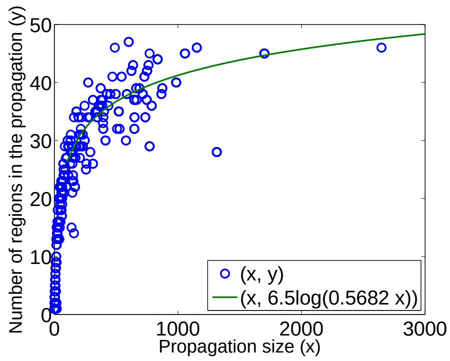



Fig. 15. Number of regions in the propagation versus the size of the propagation.

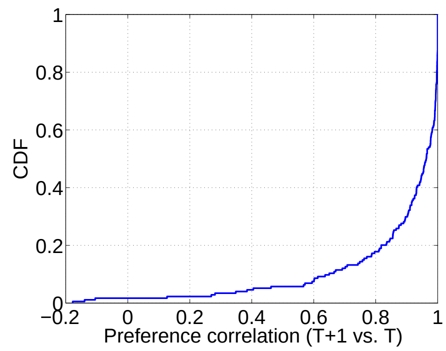



Fig. 16. CDF of preference of local friends.

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 (determined by the cloud providers), 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 timeslot, 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 timeslot ($1$ hour) on average. Thus, in each timeslot, 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 timeslot $T$ and $T-1$ over all videos re-shared in timeslot $T$ 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 timeslot. In particular, we randomly choose $80\%$ of the videos that have been imported or re-shared in the previous timeslot 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 $\theta_v^{(T+1)}$, 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 timeslot. 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 timeslot, and it is likely for them to attract new users of the video,

**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** $g_v^{(T)} > \theta_v^{(T+1)}$ **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**

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.

5.2.2 *Bandwidth Reservation for Social Contents at Edge-Cloud Servers.* 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 global-audience 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 loaded with requests; 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.

5.2.3 *Cache Replacement in 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 that 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 [Luo et al. 2009].

*Peer Cache Replacement*. Large local-audience 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. In Sec. 6, we will present the effectiveness of our cache replacement for social contents.

## 6. PERFORMANCE EVALUATION

In this section, we conduct experiments using traces from Tencent Weibo to verify our design and evaluate its performance.

### 6.1 Experiment Setup

Based on the same traces used in our measurement, we randomly select $9,318$ videos from the original traces in the last $10$ days for our experiments. These videos propagate among the regions captured by Tencent Weibo — the propagation traces are used for video replication using our design. The length of a timeslot is set to $1$ hour, as used in our measurement studied. 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 that 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 [Cheng et al. 2008], 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.

We first show the performance of PSAR over time. Fig. 17 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

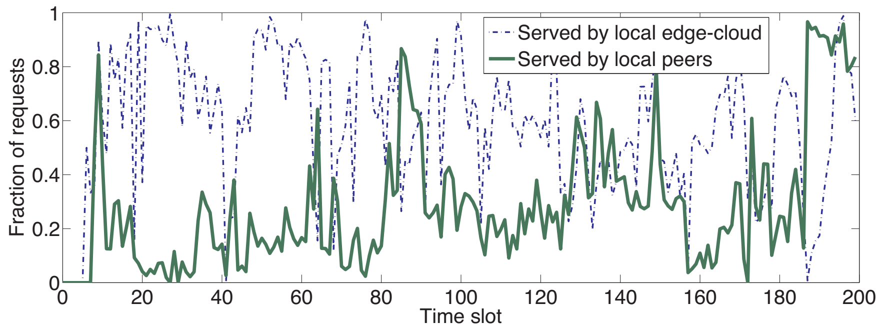



Fig. 17. Fraction of requests served by local edge-cloud servers and local peers.

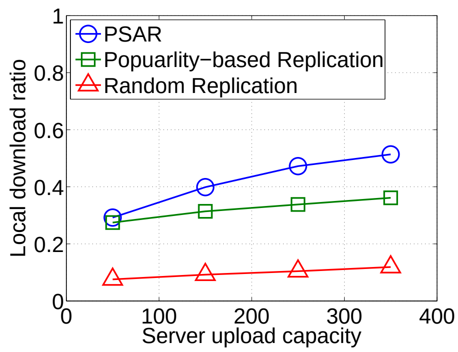



Fig. 18. Local download ratio versus the server capacity.

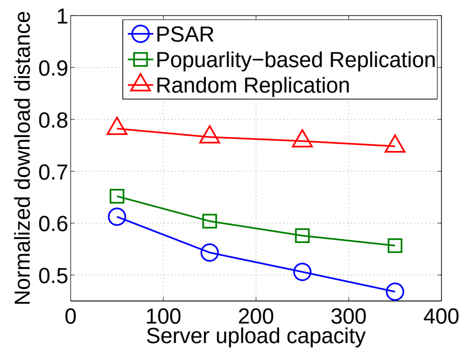



Fig. 19. Normalized download distance versus the server capacity.

random amount of upload bandwidth. Note that these two algorithms are also executed periodically in each timeslot.

*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. 18 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 improve the local download ratio by over $30\%$ comparing with the popularity-based scheme, 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 the increased server resource 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. 19 illustrates the normalized download distance versus the average server capacity. We observe that PSAR achieves a smaller download distance than the other two algorithms. The reason is that by inferring the geographic influence index, the global-audience 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.

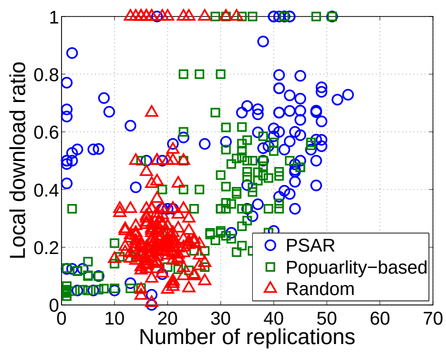



Fig. 20. Local download ratio of a video versus the number of its replications.

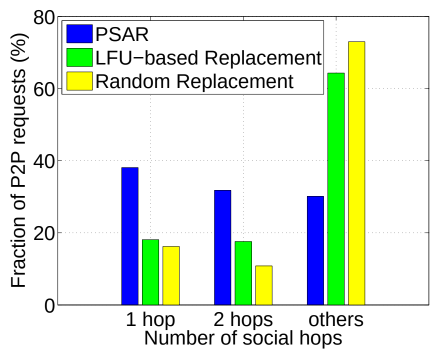



Fig. 21. Peer contribution versus the number of social hops between the pair of peers.

*Number of Replications*. We further study the impact of the number of replications of a video on the service quality. Fig. 20 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 shared 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 shared in small social groups and the ones that are highly propagating across many regions — over $4$ times of the unpopular requests can be served by local servers or peers.

## 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. 22 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 improves the local cache hit ratio by over $40\%$, compared with the LRU and LFU schemes. Besides, 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 local-audience index. We also observe that LFU and LRU have achieved the similar ratio.

*Normalized Download Geo-Distance from Peers*. We also evaluate the normalized geographic distance between the neighboring peers. Fig. 23 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

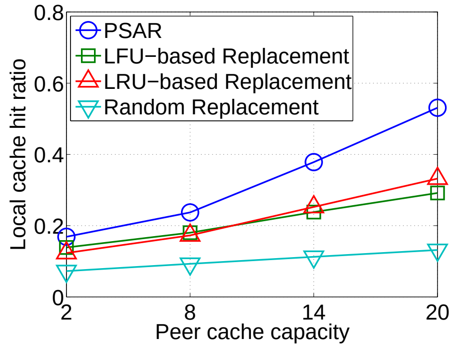



Fig. 22. Local cache hit ratio versus peer's capacity.

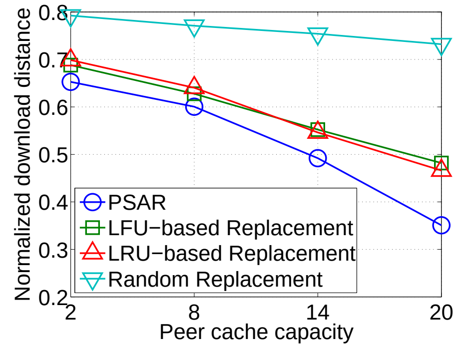



Fig. 23. Normalized download distance versus peer's capacity.

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. 21 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, over 2 times of 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. DISCUSSIONS

In this section, we discuss the reduction of replications of the social contents, the collaboration of edge-cloud and peers, how our design can be effectively implemented, and the potential content information to aid the content replication.

*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. Intuitive strategies for video reduction can be designed as follows: (1) the region where a video is mostly accessed can be selected as a permanent backup of the video; (2) other regions where the video is replicated can 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.

*Collaboration between Edge-Cloud Servers and Peers*. In the distribution of social video contents, edge-cloud servers and peers both serve the video download requests. The collaboration between them in PSAR is as follows: (1) Edge-cloud is mainly used to serve users with their preferred servers, and peers only assist in the unpopular videos; (2) In the online social network, friends are incentivized to upload contents to each other, *i.e.*, the social preference [Fehr and Fischbacher 2002]; (3) According to our design, peers are usually located close to each other. Given that edge-cloud servers are deployed closer and closer to users by the cloud providers, it is a trend that peer distribution may only take place when users very close to each other, where energy-saving transfer approaches can be used.

*Real-World Implementation*. The propagation-based social-aware delivery of social contents is a design incorporating both the content delivery system and the online social network. We discuss how the design can be implemented in a real system. (1) Data aspect (how social propagation is collected and mined). On one hand, APIs are opened by the online social networks for third-parties to retrieve not only how contents propagate between users, but also how users are socially connected to each other. Besides, they also provide mining and learning systems for the data processing (*e.g.*, Google Dremel) — a video distribution system can make use of such propagation information. (2) Delivery aspect (cloud-based and peer-assisted distribution). On one hand, cloud providers like Amazon have provided interfaces for the video service providers to elastically, dynamically and geographically allocate the cloud servers; on the other hand, content providers generally deploy clients (*e.g.*, Apps for mobile devices) which allow the peer-assisted delivery.

*Content information*. In our peer-assisted delivery, video category is used for friend preference mining. More content details of videos can still be explored, *e.g.*, videos uploaded for sharing are often accompanied with metadata that describes the content, which can be leveraged to better aid the preference inference [Melville et al. 2002].

## 8. CONCLUDING REMARKS

In this paper, we design a propagation-based social content delivery framework using a data-driven approach. By conducting extensive measurement of traces obtained from a representative online social network system, we observe unique propagation patterns, which demonstrate social, geographical and temporal localities in the social propagation. Based on the propagation patterns, we design propagation predictors to enable the propagation-based social-aware replication strategies to serve such social contents to users. Specifically, we invent three replication indices: a geographic influence index, a global-audience index and a local-audience index, which can guide the region selection, bandwidth reservation and cache replacement in the joint edge-cloud and peer-assisted replication framework proposed. Extensive experiments driven by the real-world traces further demonstrate the effectiveness and superiority of our design, which improves the local download ratio in the edge-cloud delivery by $30\%$, and the local cache hit ratio in the peer-assisted delivery by $40\%$, against traditional approaches.

## REFERENCES

ADHIKARI, V., JAIN, S., CHEN, Y., AND ZHANG, Z. 2011. Reverse Engineering the YouTube Video Delivery Cloud. In *Proc. of IEEE Hot Topics in Media Delivery Workshop*.

ADHIKARI, V., JAIN, S., AND ZHANG, Z. 2011. Where Do You Tube? Uncovering YouTube Server Selection Strategy. In *Proc. of IEEE International Conference on Computer Communications and Networks (ICCCN)*.

BALACHANDER KRISHNAMURTHY, PHILLIPA GILL, M. A. 2008. A Few Chirps About Twitter. In *Proc. of ACM WOSN*.

BENEVENUTO, F., RODRIGUES, T., CHA, M., AND ALMEIDA, V. 2009. Characterizing User Behavior in Online Social Networks. In *Proc. of ACM IMC*.

CHA, M., MISLOVE, A., AND GUMMADI, K. P. 2009. A Measurement-driven Analysis of Information Propagation in the Flickr Social Network. In *Proc. of ACM WWW*.

CHENG, X., DALE, C., AND LIU, J. 2008. Statistics and Social Network of YouTube Videos. In *Proc. of IEEE IWQoS*.

CHENG, X. AND LIU, J. 2009. NetTube: Exploring Social Networks for Peer-To-Peer Short Video Sharing. In *Proc. of IEEE INFOCOM*.

CHENG, X. AND LIU, J. 2011. Load-Balanced Migration of Social Media to Content Clouds. In *Proc. of ACM NOSSDAV*.

DODDS, P. AND WATTS, D. 2005. A Generalized Model of Social and Biological Contagion. *Journal of Theoretical Biology 232,* 4, 587–604.

DOMINGOS, P. AND RICHARDSON, M. 2001. Mining the Network Value of Customers. In *Proc. of ACM SIGKDD on Knowledge Discovery and Data Mining*.

FEHR, E. AND FISCHBACHER, U. 2002. Why Social Preferences Matter–The Impact of Non-Selfish Motives on Competition, Cooperation and Incentives. *The Economic Journal 112,* 478, C1–C33.

FORESEE. 2012. http://www.foreseeresults.com.

HARTLINE, J., MIRROKNI, V., AND SUNDARARAJAN, M. 2008. Optimal Marketing Strategies over Social Networks. In *Proc. of ACM WWW*.

HUFFAKER, B., FOMENKOV, M., PLUMMER, D., MOORE, D., AND CLAFFY, K. 2002. Distance Metrics in the Internet. In *Proc. of IEEE International Telecommunications Symposium (ITS)*.

KANGASHARJU, J., ROBERTS, J., AND ROSS, K. 2002. Object Replication Strategies in Content Distribution Networks. *Computer Communications 25,* 4, 376–383.

KEMPE, D., KLEINBERG, J., AND TARDOS, É. 2003. Maximizing the Spread of Influence Through a Social Network. In *Proc. of ACM SIGKDD*.

KWAK, H., LEE, C., PARK, H., AND MOON, S. 2010. What Is Twitter, a Social Network or a News Media? In *Proc. of ACM WWW*.

LI, H., WANG, H., AND LIU, J. 2012. Video Sharing in Online Social Network: Measurement and Analysis. In *Proc. of ACM NOSSDAV*.

LIU, Y., GUO, Y., AND LIANG, C. 2008. A Survey on Peer-To-Peer Video Streaming Systems. *Peer-to-peer Networking and Applications 1,* 1, 18–28.

LUO, J., ZHANG, Q., TANG, Y., AND YANG, S. 2009. 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.

MELVILLE, P., MOONEY, R., AND NAGARAJAN, R. 2002. Content-Boosted Collaborative Filtering for Improved Recommendations. In *Proc. of the National Conference on Artificial Intelligence*.

MISLOVE, A. 2012. Rethinking Web Content Distribution in the Social Media Era. In *NSF Workshop on Social Networks and Mobility in the Cloud*.

MISLOVE, A., MARCON, M., GUMMADI, K., DRUSCHEL, P., AND BHATTACHARJEE, B. 2007. Measurement and Analysis of Online Social Networks. In *Proc. of ACM IMC*.

MISLOVE, A., VISWANATH, B., GUMMADI, K., AND DRUSCHEL, P. 2010. You Are Who You Know: Inferring User Profiles in Online Social Networks. In *Proc. of ACM WSDM*.

NGUYEN, K., PHAM, C., TRAN, D., AND ZHANG, F. 2011. Preserving Social Locality in Data Replication for Social Networks. In *Proc. of IEEE ICDCS Workshop on Simplifying Complex Networks for Practitioners*.

PENG, G. 2004. CDN: Content Distribution Network. *Arxiv preprint cs/0411069*.

PUJOL, J. M., ERRAMILLI, V., SIGANOS, G., YANG, X., LAOUTARIS, N., CHHABRA, P., AND RODRIGUEZ, P. 2010. The Little Engine(s) That Could: Scaling Online Social Networks. *SIGCOMM Comput. Commun. Rev. 40*, 375–386.

SCELLATO, S., MASCOLO, C., MUSOLESI, M., AND LATORA, V. 2010. Distance Matters: Geo-Social Metrics for Online Social Networks. In *Proc. of Conference on Online Social Networks*. USENIX Association.

TENCENT WEIBO. 2013. http://t.qq.com.

TRAN, D. A., NGUYEN, K., AND PHAM, C. 2012. S-CLONE: Socially-Aware Data Replication for Social Networks. *Computer Networks 56,* 7, 2001 – 2013.

WANG, Z., SUN, L., CHEN, X., ZHU, W., LIU, J., CHEN, M., AND YANG, S. 2012. Propagation-based Social-aware Replication for Social Video Contents. In *Proc. of ACM Multimedia*.

WANG, Z., SUN, L., WU, C., AND YANG, S. 2012. Guiding Internet-Scale Video Service Deployment Using Microblog-Based Prediction. In *Proc. of IEEE INFOCOM Mini-Conference*.

WIKI: CLUSTERING/COEFFICIENT. 2013. http://en.wikipedia.org/wiki/Clustering_coefficient.

WU, Y., WU, C., LI, B., ZHANG, L., LI, Z., AND LAU, F. C. 2012. Scaling Social Media Applications into Geo-Distributed Clouds. In *Proc. of IEEE INFOCOM*.

XU, D., KULKARNI, S., ROSENBERG, C., AND CHAI, H. 2006. Analysis of a CDN–P2P Hybrid Architecture for Cost-Effective Streaming Media Distribution. *Multimedia Systems 11,* 4, 383–399.

YOUTUBE. 2013. http://www.youtube.com/t/press_statistics.

ZHU, W., LUO, C., WANG, J., AND LI, S. 2011. Multimedia Cloud Computing. *IEEE Signal Processing Magazine 28,* 3, 59–69.