# Video Sharing Propagation in Social Networks: Measurement, Modeling, and Analysis

Xu Cheng, Haitao Li, Jiangchuan Liu

School of Computing Science, Simon Fraser University, British Columbia, Canada Email: {xuc,haitaol,jcliu}@cs.sfu.ca

Abstract—The social networking services (SNS) have drastically changed the information distribution landscape and people's daily life. With the development in broadband accesses, video has become one of the most important types of objects spreading among social networking service users, yet presents more significant challenges than other types of objects, not only to the SNS management, but also to the network traffic engineering. In this paper, we take an important step towards understanding the characteristics of video sharing propagation in SNS, based on the real viewing event traces from a popular SNS in China. We further extend the epidemic models to accommodate the diversity of the propagation, and our model effectively captures the propagation process of video sharing in SNS.

#### I. INTRODUCTION

Since the emergence of Facebook and Twitter, the information distribution landscape and even people's daily life have drastically changed. These online social networking services (SNS) directly connect people through cascaded relations, and information thus spread much faster and more extensively than through conventional web portals or newsgroup services, inverting the conventional 2.5-hour delay of online blogging after mainstream news report [1].

There have been studies on information propagation over social networks [2], [3], [4], yet their focuses have been largely on the conventional text or image objects and on their stationary coverage among users. With the development in broadband accesses and data compression, video has become an important type of objects spreading over the social networks, and today's video sharing websites have also enabled social feeds that automatically post video links to user's SNS personal pages. Video objects, as richer media, however possess quite different characteristics:

- Video objects are generally of much larger size, hence most videos are fed from external hosting sites, like YouTube, and then spread as URL links (together with titles and/or thumbnails). Video sharing thus involves not only internal information propagation in the social network, but also external data accesses.
- Unlike text and photo objects that can be viewed directly, video objects will not be played until the user click. Thus the user behavior on watching and sharing video in the social networks is quite different.
- From social perspective, diaries and photos often possess personal information, while videos are generally more "public". Together with the shorter links, videos often propagate more broadly than texts and images. This new

video propagation trend has brought up numerous well-known Internet memes .

In this paper, we take a first but important step towards understanding the characteristics of video sharing propagation in SNS, based on real data traces from a popular SNS in China. The data traces have recorded the user's video sharing and viewing events in the entire network for one week. Our contribution in this paper can be summarized as follows:

- 1) We examine the user behavior from diverse aspects and identify different types of users in video propagation and evaluate their activities.
- 2) We examine the video link propagation patterns and reveal more details beyond stationary coverage.
- 3) We introduce an SI<sup>2</sup>RP Model which extends the epidemic models to accommodate diverse types of users and their probabilistic behavior. We validate our model and show that it effectively captures the propagation process of video sharing in social networks.

#### **II. RELATED WORK**

There have been significant studies on information propagation in social networks in the recent years, but existing works have largely targeted on viral marketing, with an objective of maximizing the information coverage. Tang *et al.* [2] investigated multiple relationships among users in SNS. They proposed a random walk-based algorithm for relationship classification, which facilitates advertising specific products. Budak *et al.* [3] presented an Independent Cascade Model, and examined an eventual influence limitation problem. Bakshy *et al.* [4] investigated influence in Twitter, and they concluded that predictions of which particular user or URL will generate large cascades are unreliable.

Although in social network content, these works focused on the influence in the social network. The difference between influence and video sharing propagation in our work is that, influence studies focus on the final stationary coverage through social relations, while not the **dynamics** of the propagation process itself. The latter, as we will show, is critical toward understanding video sharing. In addition, our work will show that, video sharing in social networks possesses quite different characteristics. Therefore, the models in those works do not work to capture the video sharing propagation. In this paper, we will present an enhanced epidemic model that effectively captures the video propagation process.

# III. DATA TRACE

We have closely worked with engineers from a popular Facebook-like Chinese SNS to capture the necessary dataset for our study. Given the common structure with other social networks, and the large dataset available, we expect that our study captures certain general trend of the video propagation process, and the methodology can be extendable to other social networking services. In addition, we will focus on the measurement results for one week.

For ease of exposition, we first list some key terminologies. When a user posts a video link from an external video sharing website to the social networking website, we refer to the action as *initiate*, and the user as *initiator*. A user can *share* a video after watching it; unless otherwise specified, "share" does not include "initiate", yet "shared contents" includes "initiated contents". Two types of datasets are obtained for our study:

- In *sharing dataset*, a record was logged when a user clicks the "share" button for a content in user's portal. We interested in 12.8 million video sharing records.
- viewing dataset, contains over 115 million records, each of which was logged when a user started watching a video.

#### IV. USER BEHAVIOR ON VIDEO SHARING PROPAGATION

We first investigate the behavior of individual users in propagating videos across social networks. We are particularly interested in the following three key questions:

- 1) How users initiate video sharing;
- 2) How users react upon receiving shared videos from friends, i.e., to watch or not;
- How users react upon finishing watching videos, i.e., to share or not.

## A. Initiating

We start from examining the initiators, each of which triggers the first share of a video. The rank distribution of the initiators (in terms of the number of initiated videos) is plotted in Figure 1. Without surprise, it is long-tail scale-free, suggesting that most users initiate few videos, but a few *active users* have initiated a remarkable number of videos. The most active user indeed has initiated over two thousand videos in one week. From the figure, we find that the top-10 data clearly differs from the rest, implying the existence of two possible types of users with different initiating behavior. Since the top active initiators serve as hubs that draw much more attention than the general users, they are worth particular attention in system optimization.

## B. Receiving and Watching

When a user shares a video, her/his friends will be notified in the news feed on the SNS site. Different from text or images that can be instantly viewed, a shared video will not be really watched until the recipient clicks. How users react to the shared videos in the news feed, i.e., to watch or not, is thus a crucial step to the propagation of video sharing.



Fig. 1: Number of initiated videos against rank

Fig. 2: CDF of share rate

We first examine the number of videos a user has watched in our dataset. Not surprisingly, the distribution is highly skewed. To understand user behavior on watching video, we also consider the number of the user's received videos from friends, given that the user's friends' activenesses are various. We compute the ratio of the number of viewed videos and that of the received videos from friends, defined as the *reception rate* for each user. On average, we find that a user watches 16% of videos shared from friends.

Whether a particular video will be viewed depends on the attractiveness of the video to the particular user, including the topic, title, thumbnail, among others, and the activeness of the particular user. The former is beyond the scope of this paper. To this end, we evaluate the correlation between the reception rate and the number of initiated videos. To our surprise, the two characters are nearly non-correlated. This finding suggests us that activeness is not necessarily a factor that determines user behavior on initiating and watching videos in SNS.

### C. Spreading

We next examine the user behavior on sharing videos after watching, a key step toward propagation. To understand how a user reacts upon finishing watching a video, that is, whether or not to further spread the video, we calculate the ratio of the number of shared videos against the number of viewed videos, defined as the *share rate*, for each user.

The CDF of share rate is shown as blue solid line in Figure 2. We notice that there are over half of the users do not share any video. Among these users, most of them have only watch a few videos, but we do find the most "selfish" users have watched more than one thousand videos without sharing any. Such *free-riders*, like those in peer-topeer systems, largely hinder the propagation. Note that, most users generally consume video rather than interact with it. As a result, for those users who have watched a few number of videos without sharing one, we do not consider them as the free-riders. We use a simple "95/5" rule to identify free-rider, i.e., 95% users have watched fewer than 5% videos.

To further investigate whether user behavior on watching and sharing videos are correlated, we generate a series of samples, each eliminates the users that have shared videos less than a threshold. We compute the *correlation coefficient* (CC) between the two rates, shown by the blue solid line in Figure 3, and we also show the series of CCs between the number of





Fig. 3: Correlation coefficient between share/reception rate, share/view count

Fig. 4: CDF of time span from share to view



## D. Summary

The above observations suggest that the users have diverse activenesses, but they are not necessarily correlated. Though in this case it is difficult to find a universal model for characterizing the behavior of all users, we can roughly distinguish three types of the users.

First, a small number of users initiates a lot of videos, and also have many friends, being hub-like. These *spreaders* (SU) are critical to the start of video propagation.

Second, *free-riders* (FU) that watch many videos without sharing any, which noticeably hinders the video propagation.

We call the rest *ordinary users* (OU), as they sometimes initiate a few videos, watch some of their friends' shared videos, and occasionally share some videos they watched. The three behaviors can be different that some users may be only active in watching videos while some may be only active in sharing (users active in initiating videos are likely to be SU).

# V. TEMPORAL AND SPATIAL CHARACTERISTICS OF VIDEO SHARING PROPAGATION

We now examine the temporal and spatial characteristics of video sharing propagation from a global view, beyond the behavior of individual users.

# A. Temporal Locality

We examine the time span between sharing a video and the actual view of this shared video by the sharer's friends. We define the view from the first friend that watches the video as "first view", and if a shared video has not been watched in our examing period, we set the "first view" value to infinity. All the views by friends are defined as "all view". The respective



Fig. 5: Illustration of propagation structures for popular videos

CDFs of the time spans for both are plotted in Figure 4. We observe that 13% of the shared videos will not be watched over the period, and for those videos that have been watched, 68% can be watched within one hour. This indicates that videos can quickly propagate to friends in SNS, exhibiting strong temporal locality. By examine the data of "all view", we find that less than 1% after 5 days. This implies that the life span of video propagating in SNS is in general of short durations, and thus our one-week dataset is suitable for the study.

We then compute the time span between watching a video and sharing it, i.e., how long it takes a user to share a video to friends after clicking to watch it. We find that over 90% of views are not followed by sharing. For the rest of the records, they are clearly affected by the video lengths: 88% of shares are created within 10 minutes after the users starting watching a video, which can be well explained that the videos shared in SNS are mostly short user-generated-content (UGC) [5].

#### **B.** Spatial Structures

We next study the spatial structures of video sharing propagation. Consider each user that has shared a video as a node, and if user a shares a video that is previously shared by user b, then a directed edge forms from b to a. It is easy to see that the shares propagate along a tree structure. By associating viewing and sharing events, we obtain all the relations to form the tree structures. Note that the users that only watch but have not shared are not included in the propagation trees. As a result, we obtain over 23 thousand propagation trees.

We first examine the tree size, which reflects the popularity of the shared video. We observe that most of the trees have small sizes, indicating most propagations are in small range, showing social relation locality. Yet there also exist trees with size greater than one thousand, and we further find out that the most popular video has been watched by near 70 thousand users. This shows the great coverage of video propagation.

The tree height is the largest length from root to a tree node, which corresponds to the maximum number of hops that the video propagates, and also indicates the "liveness" of the video. The observation is quite different from other information propagation structures; for example, email forwarding have ultra-shallow trees, among which 95% are of height 2 and no trees are higher than 4 [6]. The height of the video sharing propagation trees however can reach to 30. Finally, we investigate the tree shape of the popular video propagations. We observe two typical propagation structures for popular videos, and a visual illustration (generated by Pajek [7]) of two typical examples is shown in Figure 5: the left one has moderate depth, but most of the nodes are the children of the root node; the other one has large depths, and the number of children is not very diverse. This observation indicates that the propagation structures are highly diverse, and this is mainly due to the diverse user behavior. Therefore, those deterministic model trying to structurally and statically capture the propagation might not be unsuitable. Characterizing user's status evolution in the propagation system is thus essential to understand the video sharing propagation.

# VI. PROPAGATION MODEL

As we discussed, previous models are trying to capture the static structure of propagation. Therefore, they are not suitable to model video sharing propagation, due to the diverse user behaviors in the social networks. An alternative method is to utilize empirical knowledge to capture the user's status evolution, and the widely-used epidemic model is an appropriate method for modeling video sharing propagation in SNS.

An Epidemic model describes the transmission of communicable disease through individuals. One of the classical epidemic model is the *SIR (Susceptible-Infectious-Recovered) model*. It considers a fixed population with three compartments: *Susceptible* (S), *Infectious* (I), and *Recovered* (R). More details about SIR model can be referred to reference [8].

# A. The SI<sup>2</sup>RP Model

There is a natural mapping between object sharing propagation in a social network and the compartments of the SIR model. For a particular object, all the users in the social network are Susceptible at the beginning; the users accessing the object are Infectious, indicating that they are able to infect others by sharing the object. They can be Recovered if they choose not to share. For video sharing, the mapping however is far from being complete:

- Most of the videos propagate in relatively small ranges, covering only a small portion of the users in the entire social networks. We introduce a new compartment, *Immune* (Im, and thus Infectious is abbreviated to In), to indicate those users who have not watched the video.
- 2) In the classical SIR model, the transition is timedependent, that is, at any time, there is a chance that the stage transits to the next one. While for video sharing propagation in SNS, the transition of the stages depends on decisions at a certain time, e.g, the user needs to choose watch or not, and share or not share. Therefore, we introduce two temporary decision stage, D1 and D2;
- 3) We also need to differentiate the users who have shared the video and those who have not after watching the video. Therefore, we introduce a new compartment, *Permanent* (P), indicating users who have shared the video, and otherwise Recovered.



Fig. 6: SI<sup>2</sup>RP model

TADLE I. Vanualion of SI KI WIU	TABL
---------------------------------	------

statistic	$R^2$
reception rate	0.9958
share rate	0.9535
time to watch	0.9346
time to share	0.9811

The enhanced SI<sup>2</sup>RP (Susceptible-Immune-Infectious-Recovered-Permanent) model is illustrated in Figure 6. The transition rate  $\beta$  and  $\gamma$  can be inferred from the measurement on temporal characteristics in Section V, and probability  $p_V$  and  $p_S$  can be inffered from the measurement on user behavior in Section IV.

# **B.** Model Validation

We have ran our SI<sup>2</sup>RP model multiple times to validate its accuracy. We generate 1000 users participating in 1000 video sharing propagations for 144 hours. We then extract a series of statistics, such as number of received, watched, shared videos for each user, time span from share to watch, and time span from watch to share. We examine these statistics with the real dataset, specifically, we compute the *coefficient* of determination  $R^{21}$  of the generated data and the real data. We list those goodness of fit in Table I. The high values of  $R^2$ (above 0.9) indicates that our model accurately characterizing the user behavior in video propagation.

## C. Model Discussion

We next investigate the number of each compartment, and this will give us the knowledge of the amount of video propagating in SNS. We calculate the average and standard deviations of In, R, and P in Figure 7. Note that, we do not show statistics of S and Im, because users in these two compartment has no impact on the network traffic. From the figure, we can see that R and P are quite diverse, this is because of the diverse user behavior, and our stochastic model can well capture this diversity. It is also easy to observe that the growth trend of R and P are decreasing, as they are almost

<sup>&</sup>lt;sup>1</sup>The coefficient of determination  $R^2$  is a goodness of fit describing how well it fits a set of observations.

unchanged after 24 hours. This further confirms the temporal locality of video sharing propagation in SNS that, a video is usually watched and shared in short time, say a few hours, and then fewer and fewer users will watch and share the video.

It is worth noting that In is quite small in the figure. This is because, first, for most of the videos, the propagation range is extremely small, and thus the number of users that have chances to watch the video is not large; second, the time span from watching a video to sharing it is generally very short, and thus In will transit to R or P very quickly. This observation, along with the temporal locality observed, indicates that for most videos propagating in SNS, the number of concurrent viewers is extremely small.

It is a trend that video sharing sites are attempting to utilize peer-to-peer (P2P) technique to reduce the server workload, as in P2P system, users contribute resources. The size of P2P overlay directly affect the efficiency of P2P delivery. Therefore, if the number of concurrent viewers is small, the traditional overlay for single video will be too small to achieve satisfying performance. This also confirms our conclusion in the previous work [9]. As a result, enlarged overlay that includes a series of videos is required, and social network can in turn help this case.

The SI<sup>2</sup>RP model captures user behavior on watching and sharing videos, as well as the latency of watching and sharing videos in SNS. As such, the model has diverse applications; in particular, it can serve as a request generator/predictor for video accesses from a social network. Specifically, Infectious that can be derived from the model gives the number of users that are downloading and watching one particular video, thus the total number of users that are downloading all videos can be estimated. Given the video bit-rate, the traffic volume and its temporal distribution for individual users or videos and that from the entire social network can also be synthesized.

Peer-to-peer video streaming system and content distribution network (CDN) can also benefit from our study on temporal, spatial, and social localities found in video sharing propagation. Specifically, Our measurement on propagation structure also shows that *flash crowd*, a critical challenge to existing video servers or peer-to-peer streaming system, may not be very severe with video accesses gradually propagated through friends. Yet certain hub-like initiators, as identified in our measurement, may need to be carefully dealt with.

As emerged in the recent years, cloud services, such as Amazon EC2, Microsoft Azure, Google App Engine, offer reliable, elastic and cost-effective resource provisioning. Our  $SI^2RP$  model further provides an valuable tool to predict the video request from a social network, thus helps with server load provisioning and balancing in cloud. It may also facilitate video content providers migrating to their services to a cloud platform, through effectively forecasting traffic demands for elastically leasing resources.

#### VII. CONCLUSION AND FUTURE WORK

This paper presented an initial attempt in understanding the characteristics of video sharing propagation in social



Fig. 7: In, R, and P along time

networking services. From 12.8 million video sharing and 115 million viewing event records from a popular SNS, we revealed the user behavior from diverse aspects. We identified different types of users in video propagation and evaluate their activities. We also examined the temporal distribution during propagation as well as the typical propagation structures, revealing more details beyond stationary coverage. We further extended the conventional epidemic models to accommodate diverse types of users and their probabilistic behavior.

In this paper, we took the first step towards understanding video sharing propagation in the social networks. There are a number of issues and directions worth exploring. For example, we studied one popular Chinese SNS, and our dataset includes the entire networks. Although the particular parameters might not be identical for different SNS, our model captures the general trend of the video propagation process, and we provides the methodology to extend to other SNS. Yet, such worldwide SNS as Facebook and Twitter worth exploring, as there might be culture differences.

#### ACKNOWLEDGEMENT

This research was supported in part by a Canada NSERC Discovery Grant, an Engage Grant, a MITACS intership grant, and a China NSF international collaboration grant 61120106008.

#### REFERENCES

- J. Leskovec, L. Backstrom, and J. Kleinberg, "Meme-tracking and the Dynamics of the News Cycle," in *Proceedings of ACM SIGKDD*, 2009.
- [2] S. Tang, J. Yuan, X. Mao, X.-Y. Li, W. Chen, and G. Dai, "Relationship Classification in Large Scale Online Social Networks and Its Impact on Information Propagation," in *Proceedings of IEEE INFOCOM*, 2011.
- [3] C. Budak, D. Agrawal, and A. E. Abbadi, "Limiting the Spread of Misinformation in Social Networks," in *Proceedings of WWW*, 2011.
- [4] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts, "Everyones an Influencer: Quantifying Influence on Twitter," in *Proceedings of ACM* WSDM, 2011.
- [5] X. Cheng, C. Dale, and J. Liu, "Statistics and Social Network of YouTube Videos," in *Proceedings of IEEE IWQoS*, 2008.
- [6] D. Wang, Z. Wen, H. Tong, C.-Y. Lin, C. Song, and A.-L. Barabasi, "Information Spreading in Context," in *Proceedings of WWW*, 2011.
- [7] "Networks / Pajek," http://vlado.fmf.uni-lj.si/pub/networks/pajek/.
- [8] D. J. Daley, J. Gani, and J. M. Gani, *Epidemic Modelling: An Introduction*, ser. Cambridge Studies in Mathematical Biology. Cambridge University Press, 2001.
- [9] X. Cheng and J. Liu, "NetTube: Exploring Social Networks for Peer-to-Peer Short Video Sharing," in *Proceedings of IEEE INFOCOM*, 2009.