Video Sharing in Online Social Networks: Measurement and Analysis

Haitao Li, Haiyang Wang, Jiangchuan Liu
School of Computing Science
Simon Fraser University, Vancouver, BC, Canada
Email: {haitaol, hwa17, jcliu}@sfu.ca

Ke Xu
Department of Computer Science & Technology
Tsinghua University, Beijing, China
Email: xuke@mail.tsinghua.edu.cn

ABSTRACT

Online social networks (OSNs) have become popular destinations for connecting friends and sharing information. Recent statistics suggest that OSN users regularly share contents from video sites, and a significant amount of requests of the video sites are indeed from them nowadays. These behaviors have substantially changed the workload of online video services. To better understand this paradigm shift, we conduct a long-term and extensive measurement of video sharing in RenRen, the largest Facebook-like OSN in China. In this paper, we focus on the video popularity distribution and evolution. In particular, we find that the video popularity distribution exhibits perfect power-law feature (while videos in YouTube exhibit a power-law waist with a long truncated tail). Moreover, we observe that the requests for the new published videos generally experience two or three days latency to reach the peak value, and then change dynamically with a series of unpredictable bursts (while in YouTube, videos reach the global peak immediately after introduction to the system, and then the accesses generally decrease overtime, except possibly on some special days). These differences can raise new challenges to content providers. For example, the video popularity is now hard to predict based on their historical requests. We further develop a simple yet effective model to simulate user requests process across videos in OSNs. Trace-based simulation shows that it can well capture the observed features.

Categories and Subject Descriptors

J.4 [**Social and Behavioral Sciences**]: Sociology; H.3.5 [**Information Storage and Retrieval**]: Online Information Services—*Web-based services*

General Terms

Measurement, Performance

NOSSDAV'12, June 7–8, 2012, Toronto, Ontario, Canada.

Keywords

Social network, video sharing, popularity distribution, popularity evolution, power-law

1. INTRODUCTION

Traditionally, users have discovered videos on the Web by browsing or searching [5]. Recently, word-of-mouth has emerged as a popular way of discovering the videos, particularly on social network sites such as Facebook and Twitter [11]. On these sites, users discover video contents by following their friends' shares. Such word-of-mouth based content discovery has become a major driver of traffic to many video sharing sites. YouTube statistics [2] reported that as of January 2011 more than 500 tweets per minute containing a YouTube link, and over 150 years worth of YouTube video is watched on Facebook every day. Besides Facebook(Twitter)/YouTube, we have seen similar trends in other OSNs/VSSes, for example, between RenRen [3], the biggest Facebook-like OSN in China, and Youku [4], one of the most popular video sharing sites in China. Our measurement shows that, as of July 2011, more than 54 million unique RenRen users have participated in video viewing and 20 million participated in sharing, generating 12.4 million views, and 1.64 million shares every day. 80% of these videos are hosted by Youku.

However, such characteristics have not yet been explored in real online social networks at large scales due to a number of challenges. First, privacy protection generally prevents crawling video viewing information as easily in OSNs (e.g., Facebook/RenRen) as in VSSes (e.g., YouTube/Youku); Second, unlike dedicated video sites, OSNs can rarely provide rich statistics about shared videos; Finally, given the wide distribution of OSN users, tracing traffic from a small set of network routers/switches can hardly reveal the geographic evolution of video sharing, not to mention the sheer volume of the mixed network traffic to be analyzed.

To understand video sharing in OSNs, we closely collaborate with RenRen to analyze its server access logs. Starting from March 24^{th}, 2011, we recorded the detailed user video viewing and sharing behaviors over three months. When a user started to view a video shared by her/his friend or further shares the video, a separate record was sent to the log server. The trace data records such information as the time, viewer, sharer, and video URL, which enable us to extract rich statistics. Our measurement unveils many distinctive features of video sharing through OSNs as compared to VSSes, especially on the video popularity distribution and evolution. For popularity distribution, we find that the plot of requests and video ranks exhibits perfect power-law feature (while previous study [9] showed that in VSSes, it exhibits a power-law waist with a long truncated tail). We also find the user requests are much more skewed across the videos in OSNs (top-0.5% videos account for 80% requests) than that in VSSes (10%-80%). To further understand these unique features, we design a model to simulate the user requests process in OSNs, and analyze whether the OSN-based spreading mechanism can result in the observed distribution. For popularity evolution, we observe that the requests for the new published videos generally experience two or three days latency to reach the peak value, and then change dynamically with a series of unpredictable bursts (while in YouTube, videos reach the global peak immediately, and then the accesses generally decrease overtime, except possibly on some special days).

The rest of the paper is organized as follows. We present related works in Section 2. Sections 3 gives an overview of the RenRen OSN and our measurement methodology. We present measurement results on the video popularity distribution in Sections 4. In Section 5, we further design a model to analyze how OSN-based spreading mechanism can change the user requests across videos. We make a preliminary study of the video popularity evolution in Section 6. Finally, we conclude in Section 7.

2. RELATED WORK

To our best knowledge, our work is the first one on characterizing the patterns of video requests from OSNs, by measurement and model. There are some pioneer data-driven analysis of information spreading in OSNs. Cha et al. [12] conducted a large-scale measurement study on Flickr network, one of the most popular photo sharing social networks. They found that even popular photos spread slowly through the network. By contrast, we found that the videos in an OSN spread much faster. Rorigues et al. [11] studied the propagation of URL links posted in Twitter, using large data gathered from Twitter. They presented the distribution of height, width, and size of propagation trees and found that Twitter yields propagation trees that are wider than they are deep. They did not separate the video links from their dataset to give them an individual analysis. Scellato et al. [16] pointed that given the increasing size of Twitter and other OSNs, they may generate millions of accesses to YouTube, accounting for a consistent fraction of the total number of daily requests. Instead of studying the video popularity characteristics, they focused on the geographic property of social cascades of videos by tracking social cascades of YouTube links over Twitter.

There are also plenty of works on the user access patterns from video sharing sites (e.g., YouTube) either by crawling the webpages or tracing traffic from a set of network routers/switches. Cha et al. [9] presented an in-depth study of the static popularity distribution, and dynamic popularity evolution of videos in two large-scale VSSes, YouTube and Daum. They found that the video popularity in YouTube shows a power-law waist with a long truncated tail for huge unpopular videos. Cheng et al. [10] also studied the distribution and evolution of videos in YouTube, and found similar results. They further presented other statistics of YouTube video files such the length, bitrate, and size. More recently, Figueiredo et al. [14] made an in-depth analysis on how the popularity of individual videos evolves since the video's upload time. They found that popularity growth pattern depends on the choice of the video dataset. Besides those works that focused on the global nature of YouTube traffic by crawling YouTube webpages, and there are some complementary works by collecting YouTube traffic from local networks. Gill et al. [7] characterized the YouTube traffic collected at the University of Calgary campus network, comparing its properties with those previously reported for Web and streaming media workloads. They analyzed daily and weekly patterns as well as several videos characteristics such as duration, bitrate, age, ratings, and category. Another similar study [8] by Zink et al. also analyzed network traces for YouTube traffic at a campus network to understand the benefits of alternative content distribution strategies. Our work focuses on the distinguished features for videos shared in the RenRen OSN especially regarding video popularity distribution and evolution. And we demonstrate the word-of-mouth based social sharing can dramatically affect the pattern of user requests for videos.

3. BACKGROUND AND MEASUREMENT

This section gives an overview of the RenRen online social network and our measurement methodology.

3.1 The RenRen Social Network

Launched in 2005, RenRen is the earliest and so far the largest OSN in China. RenRen can be best characterized as Facebook's Chinese twin, implementing Facebook's features, layout, and a similar user interface. Like Facebook, RenRen's users can post video links from VSSes. Unlike Facebook, RenRen has two unique features that make it an attractive platform for our study. First, while RenRen users have full privacy control over their private profiles, their shared videos are public and thus can be crawled. For example, each individual user has a page that list all shared videos with their statistics, including the number of views and shares within RenRen. Second and perhaps more importantly, RenRen provides certain proprietary information about users' viewing behaviors.

Video sharing in RenRen is based on the friend relationships. Initially, a user shares a video link from a VSS in RenRen; This link immediately appears in her/his friends' main page as a "News Feed" in chronological order; Meanwhile, this shared video is also listed in the sharer's home page, which lists all her/his ever shared contents. Then her/his friends will probably click the shared video appeared in "News Feed"; or they may regularly visit friends' home pages to watch those shared videos, though this frequency is much lower than the first way. A video can be further propagated if some viewers share the link again.

3.2 Measurement Methodology

To understand the video sharing in OSNs, we closely collaborate with RenRen to analyze its server access logs. Starting from March 24^{th}, 2011, RenRen had been recording the detailed user video viewing and sharing behaviors over three months. When a user starts to view a video shared by her/his friend or further shares the video, a separate record will be sent to the log server. The data record of each viewing action includes: (*Starting Time*, *Viewer ID*, *Video URL*, *Direct Sharer ID*, *Original Sharer ID*). We use an example to explain the data format. Initially, $User_A$ shared $Video_1$

(denoted by URL_1) from a video sharing site; At $Time_1$, $User_B$ watched URL_1 through the share link created by $User_A$, and $User_B$ further shared URL_1 after watching it; At the $Time_2$, $User_C$ watched URL_1 through the share link created by $User_B$. For the viewing behaviors of $User_B$ and $User_C$, two records are reported: ($Time_1$, $User_B$, URL_1, $User_A$, $User_A$) and ($Time_2$, $User_C$, URL_1, $User_B$, $User_A$). Similarly, the format of sharing action is (*Creating Time, Video URL, Creating User, Direct Sharer, Original Sharer*). Table 1 summarizes our dataset with basic statics in one-day period (March 24^{th}, 2011). In this paper, we use both short-term traces (from one day to one week) to analyze the video popularity distribution, and long-term traces (several months) to explore the video popularity evolution. Since all trace data are within 2011, we omit the year index in the later sections.

Table 1: Summary of trace in one-day period

Views	Shares	Users	Videos	NewVideos
12,432,708	1,628,852	3,514,461	201,517	71,236

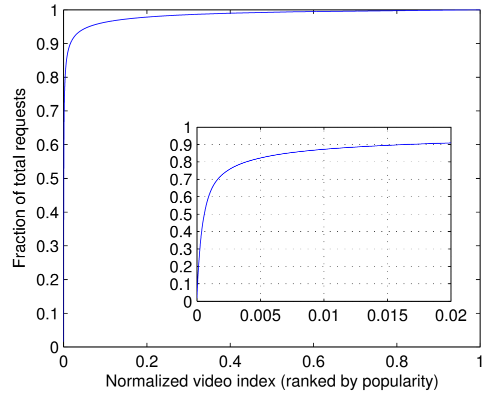

Figure 1: Skewness of requests across all videos

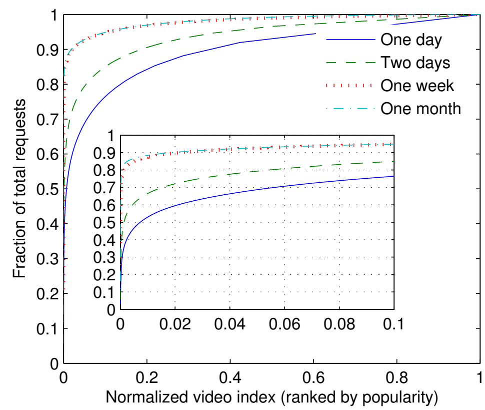

Figure 2: Requests of videos initially shared on the same day

4. VIDEO POPULARITY DISTRIBUTION

In this section, we present the measurement results on video popularity[1] distribution in the RenRen OSN from two perspectives: Pareto principle and Power-law behavior, and compare them with the corresponding results in VSSes, which were studied by a previous work of Cha et al. [9].

4.1 Pareto Principle

The Pareto principle [15] (also known as the 80-20 rule) is widely used to describe the skewness in distributions. For example, the analysis of YouTube shows that 10% of the most popular videos account for 80% of user requests [9]. It is interesting to see whether the social-network-based sharing amplifies or smooths this skewness. As shown in Fig. **??**, we can see a dramatically skewed result that 0.5% videos account for more than 80% of the total requests (the x-axis of this figure represents the videos sorted from the most popular videos to the least popular ones, with video ranks being normalized between 0 and 1); and top-2% videos account for 90% of the total requests. This suggests that OSNs amplify the skewness of video popularity. For attractive videos, more friends would view them if some users shares them; and again with higher probability these viewers will further share them. For unattractive videos, few users want to view them and are also not likely to share them after the viewing. Such difference in videos' attractiveness[2] can be further amplified over the cascading process along friend links. Therefore, attractive videos become more popular and unattractive videos become more unpopular and fade out quickly. An immediate implication of this skewed distribution is that caching can be made very efficient since storing only a small set of objects can produce high hit ratios.

To further analyze user requests distribution, we also take a closer look at the videos that are initially shared on the same day (March 24^{th}). Since mostly users are more interested in newly updated videos, this analysis will avoid the possible bias due to different video ages. We count the cumulative requests of those videos within one day, two days, one week, and one month respectively since March 24^{th}, and plot the results in Fig. 2. Similarly, the popularity of those videos also exhibits such a high skewness that the top-2% popular videos account for 90% of the total requests. We also notice that the skewness increases as the time-window increases, and almost converges after one week.

4.2 Power-law Behavior

The Power-law model [15] has been increasingly used to explain various statistics appearing in the computer science and network systems. To check the power-law pattern for the videos in OSNs, Fig. 3 plots the requests versus video ranks of all videos initially shared on the same day. We find that the plot exhibits perfect power-law (the exponent value is also given in the figure) pattern[3], and the curves of different days are very similar except for some top videos. As a comparison, the video popularity in YouTube shows a power-law waist, with a long truncated tail for huge un-

[1]We define a video's *popularity* as the amount of requests to this video.

[2]We use the term of *attractiveness* to reflect whether a video is likely to be watched and shared by users when they see it in their "News Feed" pages.

[3]A distinguished feature of power-law is a straight line in the log-log plot.

popular videos and sharp decay for popular videos [9]. It indicates that OSNs provide chance for all videos (including niche videos) to become popular, and they also amplify the effect of difference in videos' attractiveness along the spreading process. Next we will propose a model to further analyze the reason under this power-law distribution.

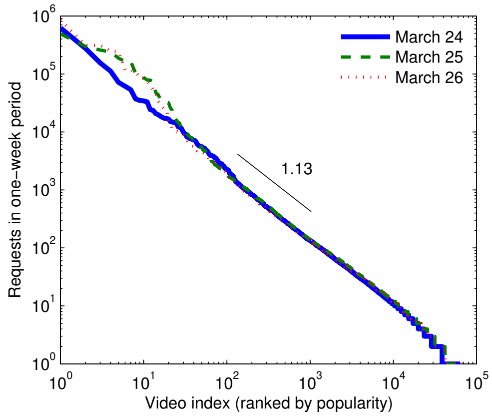

Figure 3: Requests versus video ranks (log-log)

5. MODEL ANALYSIS

Our measurement has shown distinctive popularity distribution pattern for video sharing in OSNs. To further testify whether the OSN-based spreading mechanism is the underlying reason for these features, we develop a simple yet effective model to make some preliminary analysis.

5.1 Modeling Video Spreading Process

Preferential attachment process is widely used to simulate the processes whose underlying mechanism is *rich-get-richer*. The most common example is Yule-Simon process [13], which was first introduced by Yule to study the growth in the number of species per genus. A general form of this process can be described as follows: balls are added to the system at an overall rate of m new balls for each new urn. Each newly created urn starts out with k_0 balls and further balls are added to urns at a rate proportional to the number that they already have (v_i) plus a constant $c > k_0$. In other words, when an existing entity has to be incremented by one, the i^{th} entity is chosen with probability $P(i)$:

$$P(i) = \frac{v_i + c}{\sum_{j=1}^{n} v_j + nc} \quad (1)$$

where n is the current number of urns in the system.

It is intuitive that the videos that have gained more requests have more chance to gain more requests, and this is the reason why we choose Yule-Simon Process as the basics of of model. However, it is not enough to directly apply this process to capture the video spreading process in OSNs, because it is not precise that the rate for a new request to be assigned to a video is proportional to the number of users that have already watched this video. In fact, at a given moment the number of potential requests for a video is determined by the number of users who can find this video in their "News Feed" pages but have not yet watched it. And the number of users who can find a video is mainly determined by the number of shares of this video, because in OSNs almost all videos that a user can find come from the shares of their friends. Now we formulate the preferential attachment mechanism in video spreading process by the following equation:

$$P(i) = \frac{E_i - V_i}{\sum_{j=1}^{n}(E_i - V_i)} \quad (2)$$

where $P(i)$ indicates the probability that a new request will be assigned to video i; n is the current number of videos in the system; E_i is the expected total number of requests for current shares of video i; V_i is number of requests that have happened. Thus, the value of $E_i - V_i$ reflects the number of expected requests in the future for current shares.

Therefore, the user requests process in an OSN can be described as follows: initially, all videos have one share and zero request; when a new user request comes, the model chooses video i by Eq. 2 and adds one user access to current V_i; after that it determines whether this user will further share this video by a probability–$ShareRate(ShR)$. Then if the user shares, it uses a random variable–$BranchingFactor$ (BrF) to determine the expected number of requests for this share and add BrF to current E_i. We can find that two main inputs (BrF and ShR) determine this process. BrF reflects the number of requests that follow a share, and is determined by both the video's attractiveness and the number of the sharer's friends. ShR reflects the probability a viewer will further share the viewed video, and is simply determined by the video's attractiveness to the viewer.

5.2 Validation and Analysis

We first validate whether our model can reflect the real video spreading process in OSNs by inputting the parameters extracted from RenRen. For the number of videos and requests, we configure the same values (63,591 and 2,905,276) as those in Fig. 3. To get the distribution of BrF in RenRen, we collect all 1628852 shares created on March 24^{th} and count the followed requests separately over three months. The distribution along with the fitting function are shown in Fig. 4. We also notice that the average BrF does not have obvious correlation with the total requests of a video ($\rho_p = -0.001$ and $\rho_s = -0.15$). We thus configure all videos with the same BrF distribution. To get the distribution of ShR, we collect all 12,432,708 views on March 24^{th} and record whether there is a following share behavior after the view. We count the average ShR for each video separately and show the distribution of ShR along with the fitting function in Fig. 5. One key observation in our measurement is that the plot of requests versus video ranks shows perfect power-law distribution. As shown in Fig. 6, we can see the simulation result and real-world data are pretty matched. We also count the skewness of the video popularity distribution, and the simulation result shows that the top-2% videos account for 85% of the total requests, which is very close to our observation (2%-90%). In summary, these results thus verify the validity of our model.

Given the video popularity distribution in a VSS, we now analyze whether the OSN-based spreading mechanism can amplify the skewness of such popularity distribution. To do this, we first collect all Youku (YouTube-like VSS) videos shared in RenRen in one-day period, and crawl the number of their requests in Youku VSS. We then translate the num-

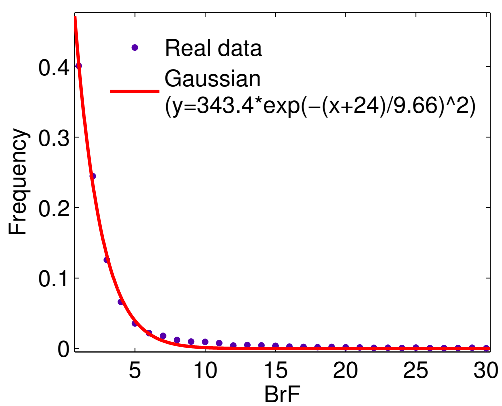

Figure 4: Distribution of BrF

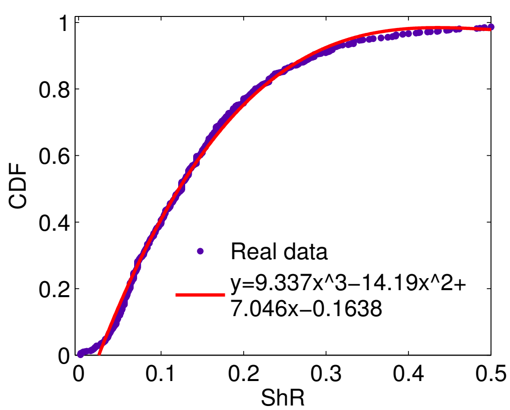

Figure 5: Distribution of ShR

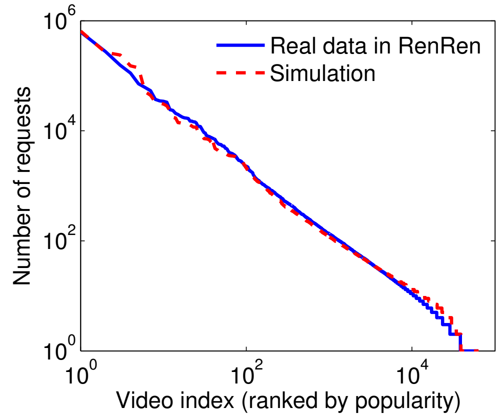

Figure 6: Model validation

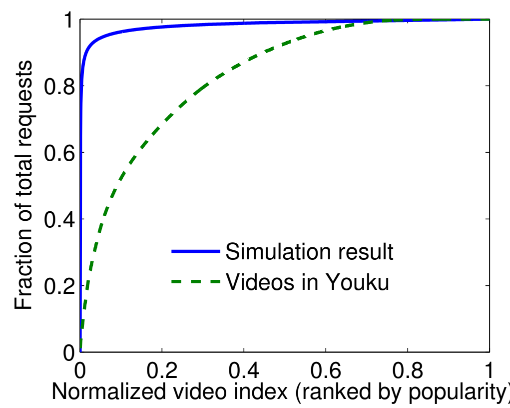

Figure 7: Popularity comparison

ber of video requests in YouTube to the value of ShR in our model by a linear function [4]. Finally, we simulate our model taking such a ShR distribution as the input parameter. For BrF, we configure the same distribution as that in Fig. 4. The comparison result is shown in Fig. 7. The result indicates that the difference in videos' attractiveness is indeed amplified over their spreading in OSNs.

6. VIDEO POPULARITY EVOLUTION

So far we have studied the static properties of video popularity. In this section, we make some preliminary analysis on video popularity evolution since they are initially shared in OSNs. When a video is shared in an OSN, it will start to attract users' attention and the number of requests will change over time. Fig. 8 shows the popularity evolution of three representative groups of videos over three months, with each group consisting of all videos with identical age (we sample three sets of videos that were initially shared on March 24^{th}, 25^{th}, 26^{th} perspectively). We observe that the requests for the new published videos generally experience two or three days latency to reach the peak value, and then change dynamically with a series of unpredictable bursts (while in YouTube, videos reach the global peak immediately after introduction to the system, and then the accesses generally decrease overtime, except possibly on some special days). An intuitive explanation for the local bursts is that, when the video is shared by a *super spreader* (the user who has a great number of friends), the video's popularity is very likely to increase again in the OSN. The fact that the local bursts for different groups do not appear on the same day also indicates the dynamic of popularity evolution. Note that the evolutions are based on the overall videos (around 0.2 million on each day). Thus the evolution for individual videos could be more dynamic.

This dynamics of popularity evolution can raise significant challenges to content providers. For example, a video's popularity is now harder to predict based on their historical requests. We compare the first few days' video requests with those after some period of time (e.g., 1, 3, and 7 days) and calculate both Pearson correlation coefficient (ρ_p) [1] and Spearman's rank correlation coefficient (ρ_s) [3] [5]. As shown in Table 2, we can see that the historical requests only have

[4]Actually their relationship is much more complicated and need further study. Here we choose the linear function as a simplified case.

[5]ρ_p has been widely used for measuring the strength of linear dependence between two variables, and ρ_s assesses how well the relationship between two variables can be described using a monotonic function. The ranges of both ρ_p and ρ_s are from -1 to 1, where a value greater than 0 indicates positive correlation, and less than 0 indicates negative correlation.

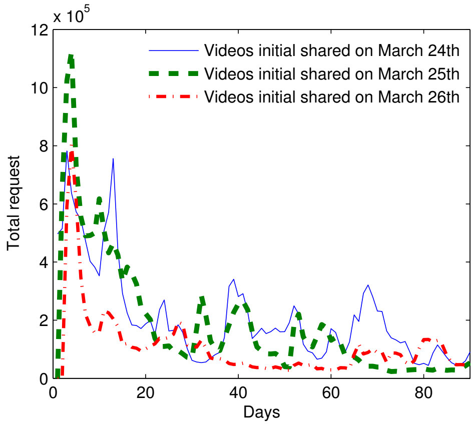

Figure 8: Popularity evolution of videos initially shared on different days

correlation with the requests in the next day, but no obvious correlation with the requests after one week. This is different from earlier study on the YouTube videos [9], where the historical requests can be effectively used to predict more distant future popularity (e.g., three months afterwards). This result suggests that some more sophisticated OSN-based models are needed to provide a better popularity prediction, which we will examine in the future work.

Table 2: Correlation (ρ_p, ρ_s) between video requests in early days and in near future

Age (x_0)	x_0+1 days	x_0+3 days	x_0+7 days
1st day	(0.48,0.53)	(0.25,0.29)	(0.13,0.21)
2nd day	(0.93,0.97)	(0.79,0.88)	(0.11,0.19)
3rd day	(0.97,0.99)	(0.80,0.89)	(0.10,0.18)

7. CONCLUSIONS AND FUTURE WORK

In this paper we presented an extensive data-driven analysis on video sharing in the RenRen OSN. Our measurement showed that videos exhibit different popularity distribution pattern compared with that in VSSes. Particularly, it shows much more popularity skewness in the OSN. We further developed a model to simulate the video spreading process in OSNs, and validated that the OSN-based spreading mechanism is the fundamental reason under such new video popularity distribution. We also made some preliminary measurement on the video popularity evolution in OSNs and revealed some distinctive features, such as the randomness, unpredictability, and multiple peaks. To capture such popularity evolution features, some enhancements are needed for our current model, and we will take this for the future work.

8. ACKNOWLEDGMENTS

Part of this work was done when Haitao Li interned at the Oak Pacific Interactive Corporation (the operator of RenRen) during the summer vacation, 2011. We want to thank the engineers from RenRen, especially Song Wen, Liang Zhang, Zhiliang Wang, Yunlong Bai, and Jing Huang. They helped us collect the data and also provided many valuable suggestions during the writing of this paper. This research was also supported by a Canada NSERC Discovery Grant, an NSERC DAS grant, a Nokia University Relation Fund, and an NSFC Major Program of International Cooperation.

9. REFERENCES

[1] J. S. Maritz. Distribution-free Statistical Methods. Chapman & Hall, 1981.

[2] YouTube Statistics. http://www.youtube.com/t/press_statistics

[3] J. Jing, W. Christo, X. Wang, P. Huang, W. Sha, Y. Dai, and B. Y. Zhao. Understanding Latent Interactions in Online Social Networks. In Proc. of IMC, 2010.

[4] K. Lai and D. Wang. Towards Understanding the External Links of Video Sharing Sites: Measurement and Analysis. In Proc. of NOSSDAV, 2010.

[5] R. Zhou, S. Khemmarat, and L. Gao. The Impact of YouTube Recommendation System on Video Views. In Proc. of IMC, 2010.

[6] X. Wu, V. Kumar, and et al.. Top 10 Algorithms in Data Minining. Journal of Knowledge and Information Systems, 2007.

[7] P. Gill, M. Arlitt, Z. Li, and A. Mahanti. YouTube Traffic Characterization: a View from the Edge. In Proc. of IMC, 2007.

[8] M. Zink, K. Suhb, Y. Gu, and J. Kurosea. Characteristics of YouTube Network Traffic at a Campus Network - Measurements, Models, and Implications. Computer Networks, 2009.

[9] M. Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. B. Moon. I Tube, You Tube, Everybody Tubes: Analyzing the Worldąŕs Largest User Generated Content Video System. In Proc. of IMC, 2007.

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

[11] T. Rodrigues, F. Benvenuto, M. Cha, K. P. Gummadi, and V. Almeida. On Word-of-Mouth Based Discovery of the Web. In Proc. of IMC, 2011.

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

[13] G. Yule, A Mathematical Theory of Evolution Based on the Conclusions of Dr. J. C. Willis. Philosophical Transactions of the Royal Society of London, 1925.

[14] F. Figueiredo, F. Benevenuto, and J. Almeida. The Tube over Time: Characterizing Popularity Growth of YouTube Videos. In Proc. of WSDM, 2011.

[15] M. Newman. Power Laws, Pareto Distributions and Zipfąŕs Law. Contemporary Physics, 2004.

[16] S. Scellato, C. Mascolo, M. Musolesi and J. Crowcroft. Track Globally, Deliver Locally: Improving Content Delivery Networks by Tracking Geographic Social Cascades. In Proc. of WWW, 2011.