A Peep on the Interplays between Online Video Websites and Online Social Networks

Junzhou Zhao, Pinghui Wang, Jing Tao, Xiaobo Ma and Xiaohong Guan
MOE KLINNS Lab, Xi’an Jiaotong University, Xi’an, P.R. China
Email: {jzzhao, phwang, jtao, xbma, xhguan}@sei.xjtu.edu.cn

Abstract—Many online video websites provide the shortcut links to facilitate the video sharing to other websites especially to the online social networks (OSNs). Such video sharing behavior greatly changes the interplays between the two types of websites. For example, users in OSNs may watch and re-share videos shared by their friends from online video websites, and this can also boost the popularity of videos in online video websites and attract more people to watch and share them. Characterizing these interplays can provide great insights for understanding the relationships among online video websites, OSNs, ISPs and so on.

In this paper we conduct empirical experiments to study the interplays between video sharing websites and OSNs using three totally different data sources: online video websites, OSNs, and campus network traffic. We find that, a) there are many factors that can affect the external sharing probability of videos in online video websites. b) The popularity of a video itself in online video websites can greatly impact on its popularity in OSNs. Videos in Renren, Qzone (the top two most popular Chinese OSNs) usually attract more viewers than in Sina and Tencent Weibo (the top two most popular Chinese microblogs), which indicates the different natures of the two kinds of OSNs. c) The analysis based on real traffic data illustrates that 10% of video flows are related to OSNs, and they account for 25% of traffic generated by all videos.

I. INTRODUCTION

Video traffic is rapidly growing in the Internet. It is reported that 15% to 25% of all the inter-autonomous system traffic today is video[1]. Recently, according to comScore's report released in February 2013[2], besides the online video websites, online social networks (ONSs) such as Facebook are the second largest platforms for people to watch videos. Since the majority of these videos are still hosted in online video websites such as YouTube and Hulu. Hence, it is an interesting topic to study how these videos transfer to OSNs, which also forms one of our motivations in this paper.

In fact, many online video websites provide people the sharing buttons to facilitate the video sharing to other external websites such as Facebook, Twitter and personal blogs. An example is shown in Fig. 1. Youku, a popular online video website in China, which is just as famous as YouTube in U.S., provides such video sharing function for each of its video. When a Youku user finds that a video is interesting, he can click on the sharing button, choose an OSN, then the video will be shared to that OSN immediately. At the same time, Youku provides brief summary information of external links that have referred this video.

Besides Youku, there are many other online video websites in China competing with each other, e.g., Tudou, Ku6.com, 56.com and so on. Many of them announced that they are the top bananas of the online video industry of China, which usually makes a third person confused to these unbelievable declarations. Similar to online video websites, there are also many OSNs co-existing in China, e.g., Renren, Douban, Qzone, Sina Weibo, Tencent Weibo and so on. Their situations are just similar to or even worse than the online video websites. Hence, there are requirements to bring orders to this disordered online ecosystem, which forms our second motivation to study them and provide people a more realistic knowledge about their positions in the online ecosystem.

In order to study the complex relationships between them, we conduct an in depth analysis based on datasets collected from multiple perspective of views of data sources:

1) Data form online video websites, Youku and Tudou.
2) Data from OSNs, Renren and Sina Weibo.
3) Network traffic data from a campus network.

We summarize our findings as follows:

- Based on the dataset collected from online video websites, we find that many factors can affect the external sharing probability of videos in online video websites. The nature of a video, e.g., its category, is also an important factor.

- Based on the dataset collected from Renren and Sina Weibo, we find that the popularity of a video itself in online video websites can greatly impact on its popularity in OSNs. Videos in Renren and Qzone usually attract more viewers than in Sina and Tencent Weibo.

Fig. 1. A video profile in Youku.
Weibo, which indicates the different natures of the two kinds of OSNs.

- The analysis based on real traffic data shows that 10% of video flows are associated with OSNs, and they account for 25% of traffic generated by all videos.

The outline of this paper is as follows. Section II summarizes related works. The data collection process is described in Section III. Section IV presents the results of an in-depth analysis of the collected datasets. Concluding remarks then follow.

II. RELATED WORK

Cha et al. [3] crawled the YouTube and Daum UCC, the most popular UGC service in Korea, and presented an extensive analysis of the video popularity distribution, popularity evolution, user behavior analysis, and content duplication in YouTube. Further, Cheng et al. [4] investigated the social networks in YouTube videos. Chatzopoulou et al. [5] analyzed popularity by looking at properties and patterns in time and considering various popularity metrics, and studied the relationship of the popularity metrics. [6] and [7] collected traces at the edge of a single campus network and studied usage patterns, video properties, popularity and referencing characteristics, and transfer behaviors of YouTube from the perspective of an edge network. Saxena et al. [8] analyzed and compared the underlying distribution frameworks of three video sharing services—YouTube, Dailymotion and Metacafe, based on traces collected from measurements performed in a PlanetLab environment.

Using traffic trace collected at multiple PoPs of the ISP, Adhikari et al. [9] inferred the load balancing strategy used by YouTube to serve user requests. Torres et al. [10] employed state-of-the-art delay based geolocation techniques to find the geographical location of YouTube servers, and performed analysis on groups of related YouTube flows. Compared with a location-agnostic algorithm to map users to data centers analyzed in [9], they found that the YouTube infrastructure has been completely redesigned and now primarily uses a nearest RTT mapping policy.

In [11], [12], the authors studied the impacts of external links on the YouTube and Youku videos. Our work differs itself by focusing on relationships between online video websites and OSNs. And we use data from three totally different sources to support this study.

III. DATA COLLECTION METHODS

In this section, we describe the data collection methods from three different data sources.

A. Data Collection from Youku and Tudou

Youku\(^1\) and Tudou\(^2\) are the top two most popular online video websites in China. To collect videos in Youku, previous study\(^{[12]}\) uses the simple Breadth-First-Search (BFS) method. However it is known that incomplete BFS is likely to densely cover only some specific region of the Youku video network, which can introduce uncorrectable statistic bias\(^{[13]}\), [14]. Hence, results obtained from data collected by BFS is unbelievable. Fortunately, we find that Youku assigns an eight digits numeric ID to each videos, so a uniform sample of Youku videos can be obtained by generating uniformly random numbers, and by polling Youku to known about their existences. Based on this uniform video sampling method, 1.5 millions of videos in Youku are collected. Similar data collection process is also conducted in Tudou, in which we also uniformly collected 1.5 millions videos. For each video in Youku and Tudou, its profile information is also retrieved, which is shown in Table I.

B. Data Collection from Renren and Sina Weibo

Renren\(^3\) is one of the largest OSNs in China with more than 150 million users. Every user can watch or share videos in Renren. This sharing can be external (from online video websites to Renren) or internal (inside Renren from a user to another user). We collect all the Renren accounts related to Xi’an Jiaotong University (they or their friends are in the campus), which form a network of 1,661,236 nodes and 32,050,611 edges. From these users, we further collect 0.5M videos. For each video, we record its original URL link address, the number of views and shares inside Renren.

We also collect data from Sina Weibo\(^4\), one of the largest microblogs in China, which has more than 200 million users. Sina Weibo assigns a ten digits numeric ID to each user, therefore we can collect tweets from 14,623 uniform randomly sampled users and about 3.7M tweets are collected. Each tweet is classified to a retweet (posted by retweeting an existing tweet) or an original tweet (posted by self writing). About 64.8% of the tweets are retweets in our dataset.

C. Traffic Collection from the Edge

The traffic data used in this paper is based on the actual network traffic over the backbones of CERNET (China Education and Research Network) Northwest Regional Center and the campus network of Xi’an Jiaotong University. The traffic data is collected at an egress router with a bandwidth of 1.5 Gbps by using TCPDUMP for about ten days in March 2011. Since application level characteristics are our primary interests in this paper, our analysis focuses on the HTTP traffic data. We

![Table I](http://www.renren.com)

| Profile                        | Youku | Tudou |
|-------------------------------|-------|-------|
| category                      | ✓     | ✓     |
| length (minutes)              | ✓     | ×     |
| size (bytes)                  | ✓     | ×     |
| rating                        | ✓     | ✓     |
| uploaded date                 | ✓     | ✓     |
| #views                        | ✓     | ✓     |
| #comments                     | ✓     | ✓     |
| #favorites                    | ✓     | ✓     |
| #likes                        | ✓     | ✓     |
| #dislikes                     | ✓     | ✓     |
| #external shares              | ✓     | ✓     |
| top 20 external shared links   | ✓     | ✓     |

1.http://www.youku.com
2.http://www.tudou.com. Tudou was acquired by Youku in March 2012.
3.http://www.renren.com
4.http://www.weibo.com
group packets into different bidirectional HTTP flows, where a flow is defined as a bidirectional, ordered sequence of packets generated by a pair of (packet source IP, packet source port) and (packet destination IP, packet destination port). For each HTTP bidirectional flow, we record its starting time, finishing time, total bytes, total packets, HTTP request (e.g. the method, URL, Host), and HTTP response (e.g. status code, content type, content size). 40M HTTP flows are collected, which include 13K distinct sources.

IV. DATA ANALYSIS

In this section, we analyze the relationships between online video websites and OSNs from three different perspective of views.

A. External Video Sharing in Youku and Tudou

First, we provide overview statistics of the video external shares in Youku and Tudou. Fig. 2 depicts the empirical cumulative distributions (CDF) of the number of external shares for videos in Youku and Tudou. We find that the majority of videos are not widely shared externally. For example, 80% of Youku videos have less than 22 shares, and 80% of Tudou videos have less than 18 shares. About 0.3% of the videos in Youku and Tudou have more than 10,000 external shares, and these minority videos are widely shared outside of Youku and Tudou.

Second, we want to study what are the factors that may cause a video to be widely shared outside of Youku and Tudou? Hence, we consider the video profiles listed in Tab. 4. The relationships between these factors and the average number of external shares are shown in Fig. 3. Most of these factors are consistent with each other either in Youku or Tudou. We will analyze them in detail.

- #views The number of views seems to have the strongest relationship with the possibility of external sharing of a video. Intuitively, a video receiving more views indicates itself is popular, hence it will be shared to external sites with a high probability. And because of the feedback effect, many external shares will also increase its views inside of a video site.

- #comments The number of comments has similar effect with the number of views, and it also has strong relationship with the external sharing. However, to comment on a video needs the user to login first, which may stop a lot of people who don’t own accounts.

- #likes and #dislikes Comparing the two factors, a video receiving more likes will be more likely to be shared than a video has less likes or many dislikes.

- Length Video length has weak correlation with video’s external sharing. Comparing short videos with long videos, short videos have higher probability to be shared. This is because users may not have enough time to watch a long video and decide to share it.

- Age Age also has weak correlation with a video’s external sharing. Videos in Tudou shows that old videos are less likely to be shared than new videos. However, this is a little different in Youku.

- Bit rate and rating Bit rate of a video is defined as Size/Length. Both the two factors can be used to characterize the quality of a video. Generally, higher quality videos has higher probability to be shared. However, this relationship is very weak.

We use Pearson correlation coefficients (PCC) to quantify these correlations in Youku and Tudou, which are shown in Table 4. PCC is the most common metric to measure the dependence between two quantities. For a pair of metrics $x$ and $y$, its corresponding values for the sampled video $i$ ($i = 1, 2, \ldots, n$) are $x_i$ and $y_i$, where $n$ is the total number of sampled videos. Then their PCC is defined as

$$
\rho_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}
$$

where $\bar{x} = \sum_{i=1}^{n} x_i$ and $\bar{y} = \sum_{i=1}^{n} y_i$ are the sample means of $x$ and $y$. The results of PCCs are consistent with our analysis above.

5All recorded IP addresses are anonymized to protect users’ privacy.
Next, the category of a video also affects the external sharing a lot. To see this, we show category’s effect in Fig. 4. We find that music and animation are the top-2 categories that have the highest percentages of videos for both Youku and Tudou. Most external links in Youku come from video categories music, games, original, comedy, TV series, and the same top-5 categories in Tudou come from categories entertainment, music, comedy, original and Hot topic. The largest averages of the number of external shares per video are both comedy for Youku and Tudou.

Finally, we study what are these external links pointing to, or what external sites are consuming these videos. Table III states the most frequently external websites identified from the external links in Youku and Tudou. Among the external websites, Renren and Qzone consume the majority of the videos, about 96% of the Youku videos and 75% of the Tudou videos. Meanwhile, Renren and Qzone consume more videos than microblogs such as Sina and Tencent Weibo.

**B. Video Consuming in Renren and Sina Weibo**

Having understood the external video sharing in online video websites, we now move to study how videos are consumed in Renren and Sina Weibo, which are the two most popular ONSs in China (as famous as Facebook and Twitter in U.S.). When a video is shared from an online video website to an ONS, the remaining behaviors related to this video is watching and sharing within the ONS. Hence, in the following, if we don’t clarify, the sharing behavior means internal sharing within an ONS, which is different from the external sharing studied in previous.

First, we provide a general picture about video views and shares in Renren. Fig. 5 shows the CCDF of the number of shares and views for a video. We find that the majority of videos are not shared or watched by a lot of people, e.g., 80% of videos are shared less than 29 times, and 80% of videos are viewed by Renren users less than 153 times, which is similar to the external video sharing in online video websites. They both have a long heavy tail. When we further investigate the relationship between views and shares, we find it has a very simple mathematic formula as shown in Fig. 5a. If denote the number of views by \( v \) and the number of shares by \( s \), then the relationship between \( v \) and \( s \) is

\[
\frac{s}{v} = 0.32v^{0.93}
\]

which is very close to a linear relationship.

Next, we do a reverse engineering study to study where the videos come from. The answer to this question can give us a rank of the popularity of online video websites. We summarize the most popular video websites in Renren in Tab. IV. In fact, using different popularity metrics will obtain different ranks. We find that videos from Youku have the largest amount, views and shares, e.g., 57% of the videos in Renren are from Youku. These videos attract 71% of all the views and 66% of all the shares. However, when averaged to each video, videos from Ku6.com has the highest average views and shares, e.g., each video in Ku6.com can attract 1102 shares and 11215 views on average. Since videos from Ku6.com only account for 4.8%, the average figures indicate that these minority videos are high quality. For videos from each video website, the average number of views per share has much smaller variance than the
average number of views per video, which indicates that the number of views is strongly related to the number of share actions.

Now we move to Sina Weibo, the largest microblog in China. In Sina Weibo, about 7.4% of the tweets contain a video and the same number is 2.4% in the original tweets, 11.2% in the retweets. About 85.8% of the video tweets are retweets. These figures illustrate that people are more reluctant to retweet a video tweet than uploading a video by himself.

Weibo users can post tweets via various ways. Loginning into an account from a browser and posting tweets is the most common way. Except that, about 27.2% of the users post tweets via mobile devices. And the top 10 most popular mobile OSs or devices used by Sina Weibo users are iPhone(13.92%), Android(5.62%), Nokia(4.71%), iPad(1.72%), Java(1.30%), Motorola(0.19%), HTC(0.15%), BlackBerry(0.14%), SonyEricsson(0.10%), and WindowsMobile(0.08%). There is a difference of posting behavior for video tweets and non-video tweets. For non-video tweets, the proportion of postings via mobile devices is 27.4%. But for video tweets, the number is only 0.94%. This indicates that mobile devices are still not convenient for posting video tweets.

Fig. 6 shows the top 8 most popular video sites in Sina Weibo. Youku and Tudou are the top 2 video sites which generate more than 70% of the videos in Sina Microblog. But the most retweeted and commented videos are come from Sina itself, which takes account only 9.5% of the videos. This indicates that videos from Sina are more attractive than others.

C. A Peep from Campus HTTP Traffic Data

Finally, we study the interactions between online video websites and OSNs from a much lower level—the campus HTTP traffic. The Internet traffic information should also reflect some properties of these interactions. To see this, we first show the fraction of different traffic from a campus in Table V and Table VI. We can see that 32% of the traffic is made of video, and 39.5% of the video traffic comes from Youku. The videos from Sohu have the highest download speed.

For a video flow, denote its start time by \( t \), then we say that a flow is an OSN related video flow if the previous flow with the same source IP came from an OSN in time interval \( (t - \Delta, t) \). When \( \Delta = 1 \) (second), we find that 9.9% of the video flows are OSN related video flows, which account for 25.1% of the traffic generated by all videos. When \( \Delta = 2 \) (seconds), about 10.6% of the video flows are OSN related video flows, which account for 26.0% of the traffic generated by all videos. For each video website, the traffic and flow fractions are shown in Table VII. For the shortage of space, in the following analysis we set \( \Delta = 1 \) (second).

Next, we study the traffic properties inside OSNs. The traffic distributions of OSN related video flows among OSNs are shown in Table VIII. We can see that videos flows are much more popular in Renren and Qzone, than in Sina and Tencent Weibo. This finding is consistent with the results shown in Table III which are obtained from the dataset collected from online video websites Youku and Tudou. We further show the traffic distributions of OSN related video flows among the video websites for each OSN website in Table IX. We can see that the top-3 most popular video websites in each OSN are Youku, Ku6.com, and Tudou, which is also consistent with the results shown in Table V that are obtained from the dataset collected from Renren.

V. Conclusions

We comprehensively studied video sharing between online video websites and online social networks from three perspective views: online video websites, online social networks, and

---

**Table V**

| Website      | Distribution of #bytes (%) | Distribution of #flows (%) |
|--------------|-----------------------------|-----------------------------|
| Videos       | 32.0                        | 18.8                        |
| Images       | 31.5                        | 32.8                        |
| Text         | 11.2                        | 9.5                         |
| Applications | 2.2                         | 32.9                        |
| Others       | 22.8                        | 6.1                         |

**Table VI**

| Website                  | Distribution of #bytes (%) | Distribution of #flows (%) | Download speed (kB/s) |
|--------------------------|-----------------------------|-----------------------------|-----------------------|
| Youku                    | 39.5                        | 23.0                        | 50                    |
| Sina.video.com           | 16.3                        | 1.5                         | 57                    |
| Tv.sohu.com              | 8.8                         | 1.5                         | 121                   |
| Ku6.com                  | 0.58                        | 1.3                         | 6.2                   |
| Joy.cn                   | 0.04                        | 0.16                        | 6.0                   |
| Yinyuetai.com            | 0.91                        | 0.06                        | 53                    |
| Others                   | 21.1                        | 60.1                        | 27                    |

---

**Table VII**

| Website                  | Traffic fraction (%) | Flow fraction (%) |
|--------------------------|----------------------|-------------------|
|                          | \( \Delta = 1 \) | \( \Delta = 2 \) | \( \Delta = 1 \) | \( \Delta = 2 \) |
| Youku                    | 19.8                | 20.8              | 10.6                | 11.3                |
| Sina.video.com           | 27.6                | 28.1              | 15.6                | 16.7                |
| Tudou                    | 53.9                | 55.0              | 17.1                | 18.4                |
| Tv.sohu.com              | 21.7                | 22.9              | 32.2                | 33.4                |
| Ku6.com                  | 23.0                | 23.2              | 19.1                | 21.0                |
| Joy.cn                   | 61.3                | 61.5              | 17.3                | 19.4                |
| Yinyuetai.com            | 47.8                | 47.9              | 30.4                | 31.6                |
| Others                   | 25.9                | 26.2              | 7.3                 | 7.9                 |
network traffic of a campus network. The results of our study provide insights on the interplays between the two kinds of websites, which are summarized bellow:

a) Many factors can affect the external sharing probability of videos in online video websites. The nature of a video, e.g., its category, is also an important factor.

b) The popularity of a video itself in online video websites can greatly impact on its popularity in OSNs. Videos in Renren and Qzone usually attract more viewers than in Sina and Tencent Weibo, which indicates the different natures of the two kinds of OSNs.

c) 10% of video flows are associated with online social networks, and they account for 25% of traffic generated by all two kinds of OSNs.

### References

[1] C. Labovitz, S. IkeI-Johnson, D. McPherson, J. Oberheide, and F. Jahanian, “Internet inter-domain traffic,” in *Proceedings of ACM SIGCOMM 2010*, September 2010, pp. 75–86.

[2] comScore, “comScore releases February 2013 U.S. online video rankings,” *http://www.comscore.com/Insights/Press_Releases/2013/3/comScore_Releases_February_2013_U.S._Online_Video_Rankings* March 14, 2013.

[3] M. Cha, H. Kwak, P. Rodriguez, Y. Y. Ahn, and S. Moon, “I tube, you tube, everybody tubes: analyzing the world’s largest user generated content video system,” in *Proceedings of ACM SIGCOMM IMC 2007*, October 2007, pp. 1–14.

[4] X. Cheng, C. Dale, and J. Liu, “Statistics and social network of youtube videos,” in *Proceedings of IEEE IWOQS 2008*, June 2008, pp. 229–238.

[5] G. Chatzopoulou, C. Sheng, and M. Faloutsos, “A first step towards understanding popularity in youtube,” in *Proceedings of IEEE Infocom 2010 Network Science Workshop*, March 2010, pp. 1–6.

[6] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, “Youtube traffic characterization: A view from the edge,” in *Proceedings of ACM SIGCOMM IMC 2007*, October 2007, pp. 15–28.

[7] M. Zink, K. Suh, Y. Gu, and J. Kurose, “Characteristics of youtube network traffic at a campus network: measurements, models, and implications,” *Journal Computer Networks*, vol. 53, no. 4, pp. 501–514, March 2009.

[8] M. Saxena, U. Sharan, and S. Fahmy, “Analyzing video services in web 2.0: a global perspective,” in *IN NOSDAD 2008: Proceedings of the 18th International Workshop on Network and Operating Systems Support for Digital Audio and Video*, May 2008, pp. 39–44.

[9] V. K. Adhikari, S. Jain, and Y. Zhang, “Youtube traffic dynamics and its interplay with a tier-1 isp: An isp perspective,” in *Proceedings of ACM SIGCOMM IMC 2010*, November 2010, pp. 431–443.

[10] R. Torres, A. Finamore, J. Kim, M. Mellia, M. Munafo, and S. Rao, “Dissecting video server selection strategies in the youtube cdn,” in *Proceedings of IEEE ICDCS 2011*, June 2011, pp. XX–XX.

[11] K. Lai and D. Wang, “Measurement study of external links of youtube,” in *Proceedings of IEEE GLOBECOM 2009*, December 2009, pp. 1–6.

[12] ———, “Towards understanding the external links of video sharing networks: Measurement and analysis,” in *In NOSDAD 2010: Proceedings of the 20th International Workshop on Network and Operating Systems Support for Digital Audio and Video*, June 2010, pp. 69–74.

[13] D. Achlioptas, D. Kempe, A. Clauset, and C. Moore, “On the bias of bfs (breadth first search),” in *In ACM STOC*, May 2005, pp. 694–703.

[14] M. Kurant, A. Markopoulou, and P. Thiran, “On the bias of bfs (breath first search),” in *Proceedings of International Teletraffic Congress 2010*, September 2010.