

Insight Data of YouTube from a Partner's View

Xu Cheng^{†‡}, Mehrdad Fatourehchi[†],
Xiaoqiang Ma[‡], Cong Zhang[‡], Lei Zhang[‡], Jiangchuan Liu[‡]

[†]BroadbandTV Corp., British Columbia, Canada
{xcheng,mfatourehchi}@bbtv.com

[‡]School of Computing Science, Simon Fraser University, British Columbia, Canada
{xma10,cong,zlza70,jcliu}@cs.sfu.ca

ABSTRACT

YouTube is arguably the most popular online videos sharing site nowadays. To further augment its service with better revenue, it has started working with content owners (known as *YouTube partners*) whose copyrighted videos and channels have pulled massive audience. By uploading high-quality premium videos, the partners have essentially changed the user-generated content feature of YouTube and further increased YouTube's popularity. Understanding the latest YouTube access pattern is thus crucial to both YouTube and its partners, as well as to other providers of relevant services. In this paper, we for the first time analyze a large-scale YouTube dataset from a partner's view. We make effective use of *Insight*, a new analytics service of YouTube that offers inside statistics for partners about their content accesses and audience behaviours. From the raw Insight data that are confined to simple scalars and charts, we reveal the inherent relationship among the various metrics that affect the popularity of the videos. Our findings facilitate YouTube partners to adapt their content deployment and user engagement strategies, having great potentials for them to collaborate with YouTube to generate more views and subsequently increasing their revenues.

Categories and Subject Descriptors

H.5.1 [Multimedia Information System]: Video; C.4 [Performance of Systems]: Measurement techniques

General Terms

Measurement

Keywords

Youtube Partner, Viewing Behaviour, Referral Analysis

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
NOSSDAV'14, March 19–21, 2014, Singapore, Singapore.
Copyright 2014 ACM 978-1-4503-2706-0/14/03 ...\$15.00
<http://dx.doi.org/10.1145/2578260.2578274>.

1. INTRODUCTION

Nowadays, watching online video has become one of the most important activities in people's everyday life. A survey revealed that Americans spend 3.5 hours per week [1]. The significant viewership and traffic have brought great revenues to such video sharing websites as YouTube, by displaying advertisements to monetize videos, and this has been the main source of YouTube's revenue. Thousands of channels are making six figures a year [2]. As a typical user-generated content service, YouTube had substantially changed the traditional online video services. Besides, such companies and organizations as Electronic Arts, ESPN, and Warner Brothers are also providing their premium video contents on YouTube now.

To accommodate these content owners with copyrighted videos and popular channels, YouTube has introduced a YouTube Partner Program [3], which has largely improved the quality of YouTube videos, and has further increased the YouTube's revenue. YouTube now has more than a million partners and pays out millions of dollars a year for them [2]. Over the past few years, more and more small businesses and individuals have also partnered with YouTube to benefit from the monetization of their videos.

In the literature, video sharing services, especially YouTube, have been studied extensively [4, 5, 6, 7]. A number of distinguished features of YouTube from the traditional VoD system have been identified. For example, video lengths are short, user access pattern has the unique characteristic, and social networks exist in YouTube. There are also some works specifically studied the impact of related videos and referrers on video views [8, 9]. To the best of our knowledge, the Insight data were not available in the existing research studies, since they are private to YouTube partners. Therefore, revealing more unique features of YouTube to extend the existing studies has the great contribution to researchers.

In this paper, we conduct a measurement study on analyzing two large-scale Insight datasets of YouTube partners and reveal a number of the unique features. Specifically, we investigate the characterized viewing surges, visiting behaviour, as well as analyze the impact of user subscription on video views. We also reveal the breakdown of referral sources, and specifically investigate the impact of video suggestion and external website sources on the video views.

The remainder of the paper is organized as follows. We present the dataset description in Section 2. Section 3 and Section 4 analyze the characteristics of two Insight datasets, respectively. Finally, we conclude the paper in Section 5.

Table 1: Summary of Insight Data

	Channel A	Channel B	Channel C	Channel D	Channel E
Genre	game trailers	movie trailers	gameplay videos	TV trailers	music videos
Months since established	> 36	> 36	12	10	16
# of videos	4,879	1,975	1,051	914	743
# of views (1,000)	318,975	986,581	1,387	3,229	4,066
# of views traces (1,000)	116,512	65,450	3,551	1,454	3,623
# of referrers traces (1,000)	144,829	178,365	817	463	1,514

2. DATA COLLECTION

The Insight Analytics dashboard shows some simple scalars and charts. To further investigate, we implemented an authenticated crawler to collect the raw Insight data using the YouTube API¹. In this work, we investigated two of the raw Insight datasets. A brief explanation is as follows:

- in *views* datasets, each trace records the number of views and other statistics such as unique visitors, likes, comments, and subscriptions, for a video in a specific region, on a specific day;
- in *referrers* datasets, each trace shows how users have reached a video in a specific region, on a specific day. It provides the number of views from a particular source, including the related video list, search results, external websites, etc.

We crawled five popular channels with premium contents from the BroadbandTV’s network² of over twelve thousand content partners. BroadbandTV Corp. is one of the top multi-channel networks (MCNs) on YouTube, who has more than millions of videos in its networks and attracted billions of monthly views. Due to the confidential agreement, we replace the name of five channels by Channel A to Channel E, which can represent the popular YouTube entertainment channels. Channel A and Channel B are among the most popular channels on YouTube, attracting over hundreds of millions of annual views. Channel C, Channel D, and Channel E are relatively new, and therefore are less popular than the first two; however, they still attract millions of annual views. These five channels can represent the popular YouTube entertainment channels, and Table 1 summarizes the traces corresponding to these channels.

3. ANALYSIS OF YOUTUBE INSIGHT’S VIEWS DATA

3.1 Viewing Surge

There are a number of research works trying to model the growth pattern of the video views [10], many unpredictable factors exist. For example, a YouTube partner can conduct SEO (search engine optimization) on his/her videos by optimizing tags and thumbnails; also, if a power user on social media embeds or shares a video, the video might immediately gain a great amount of views. These actions can cause a viewing *surge*, and are hard to be captured by a generalized model.

Nevertheless, for YouTube partners, a viewing surge is crucial, as they can gain the knowledge about what causes a surge and thus try to create more viewing surges to increase

¹<https://developers.google.com/youtube>

²<http://broadbandtvcorp.com>

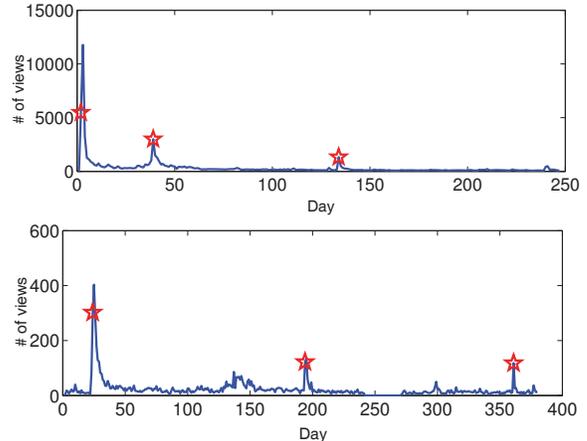


Figure 1: Two examples of detecting viewing surges

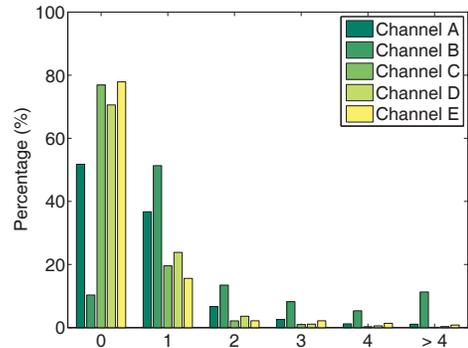


Figure 2: Breakdown of the number of surges

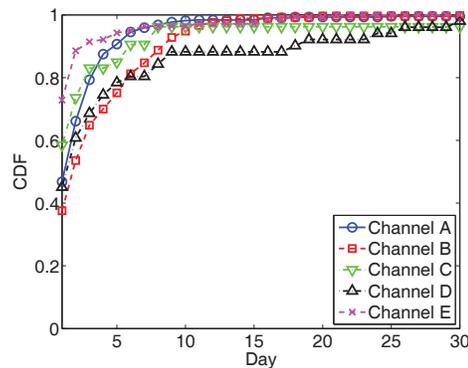


Figure 3: CDF of the location of surges

their revenues. We therefore provide a mechanism to identify the viewing surges and further characterize it. Simply put, we calculate the average of the previous views to deter-

mine the surge. In particular, if the number of views doubles from the previous 7 days and is greater than a threshold δ , we define this as a surge. Two examples of the viewing surge detection results are also shown in Figure 1.

In the algorithm, we have a threshold parameter, which determines the least number of views that can be considered as a surge. For example, even if the number of views is 20 and the previous average is 5, we do not consider it as a surge, because this small number is not statistical significant. In general, the granularity of monetizing video is one thousand views, and thus we set δ as one tenth of that, i.e., 100, for Channel C, Channel D, and Channel E; since Channel A and Channel B are extremely popular, we set δ as 1000.

We apply the algorithm to each video, and first look at the distribution of the number of surges, as shown in Figure 2. Except for Channel B, no surge is detected for over half of the videos for other channels, probably because the number of views does not reach the threshold. For Channel B, over 60% of videos have at least one surge, further indicating the popularity of the channel. Moreover, about 15% of videos in Channel B have two surges. This can be explained as follows. The first surge probably occurs soon after the video is uploaded. As videos in this channel are movie trailers, the second surge probably is associated with the time that the movie is officially released.

We further examine when the surges occur, as shown in Figure 3. We observe that over 40% of surges occur on the first day for all the channels. For Channel A, Channel B, Channel C, and Channel E, almost all surges occur within the first week. For Channel D, the surges occur later than the other channels. In summary, the viewing surges mostly occur soon after the video is uploaded.

3.2 Visiting behaviour

We should note that studying the number of views cannot give us knowledge about the visitors (for example, how many times a video will be watched by the same user). Traffic logs can infer such statistics, but the existing studies are confined by time and region [6]. Fortunately, the views dataset provides these statistics.

The ratio of the number of views and unique visitors implies on average how many times a video is watched by one viewer. We calculated the all-time average ratio for each video and plotted the CDF in Figure 4. The average ratios are 1.18, 1.24, 1.47, 1.24, and 1.46, respectively. Channel B and Channel D are similarly attractive, and both are more attractive than Channel A. On the other hand, Channel C and Channel E are more attractive than the other three due to the video genre. Note that, videos in Channel A, Channel B, and Channel D are trailers of games, movies, and TV shows. Users who watch trailers intend to get information about the game, movie, and TV show, and thus they are likely to watch only once when they obtain such information; while videos in Channel C and Channel E are game playing videos and music videos, which are mostly not informative, and if the video content is attractive, users are tend to watch more than once. Therefore, besides the number of videos and number of viewers, the attractiveness of content is crucial for non-informative video channels, such as Channel C.

We further examine having multiple views within the last 7 days and 30 days, shown in Figure 5. We can see that there are many more revisits for Channel C and Channel E than the other three trailer channels, which is consistent with the

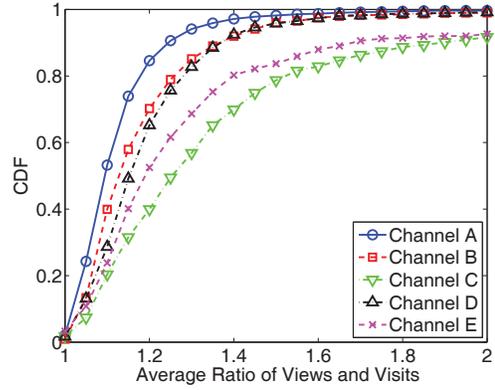


Figure 4: CDF of average ratio of views and visits

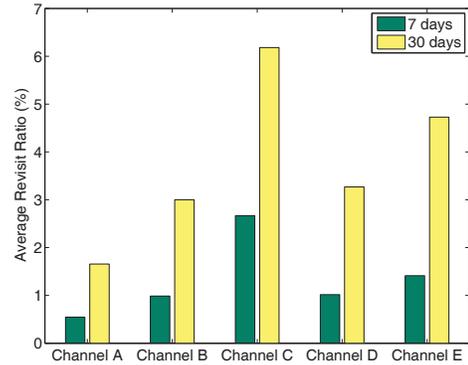


Figure 5: CDF of channels' average revisit ratio

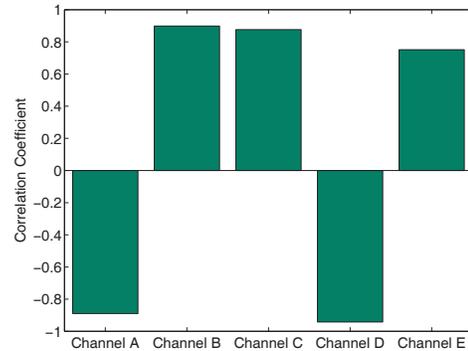


Figure 6: Correlation coefficient between subscribers and daily views

average ratio of views and visits (Figure 4). Why do users revisit a channel? A user may be interested in a particular video that he/she has watched before, and thus can search or browse from the watch history to find that video. A better and easier way is to subscribe to the channel. Next, we will investigate the impact of subscriptions.

3.3 Subscription

Users can subscribe to a YouTube channel, so that when they log on to the YouTube, some videos from the subscriptions are listed on the frontpage. Subscribers can be considered as the loyal audience of the channel, and thus are crucial to the success of YouTube partners. It is known that users

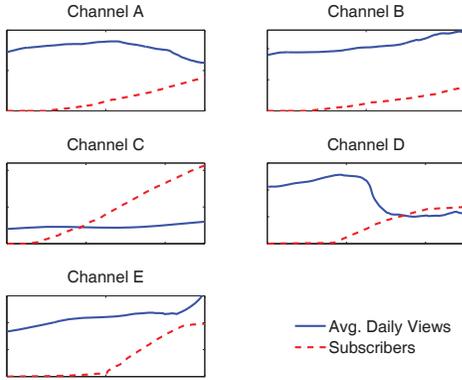


Figure 7: Daily views and subscribers over time

are likely to watch other videos following the current one [11]. There exists a section on YouTube called “more from subscriptions”. Therefore, while playing the video from the subscription list, other videos from the same channel have chance to be watched. This suggests that YouTube partners should upload more videos of similar interest, so that it will attract users to browse more videos of the channel.

As of Q2, 2012, these five channels have 206 K, 750 K, 23 K, 7 K, and 12 K subscribers, respectively. We calculate the CC (correlation coefficient³) between the number of total views and the number of subscribers, and the result is as high as 0.998, which clearly indicates that the popularity of the channels is highly correlated with the number of subscribers.

We further examine the impact of subscribers for each channel over the time. Specifically, we calculate the CCs between the number of net subscribers (the number of subscriptions minus the number of unsubscriptions) on each day and the average daily views on and after that day, and the results are shown in Figure 6. Channel B, Channel C, and Channel E display a very high positive correlation, indicating that the subscribers of these three channels have great impact on the video views. Interestingly, Channel A and Channel D display very high negative correlation.

To better understand this negative correlation, we plot the number of daily views and subscribers over time for these five channels in Figure 7. From the figure we can see that, the number of subscribers of all these five channels are increasing. The daily views of Channel B, Channel C, and Channel E are increasing, while Channel A and Channel D are decreasing at some point. This explains why the values of CC for Channel A and Channel D are negative. Regarding Channel D, an event (cannot be disclosed due to confidentiality) led to a great decrease of the views. This also confirms that a generalized model trying to capture the growth pattern of video views is difficult to be accurate, as there are various unexpected events affecting the video views.

³Correlation coefficient is calculated as $\frac{1}{n-1} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{s_X} \right) \left(\frac{Y_i - \bar{Y}}{s_Y} \right)$, where X and Y are two random variables, \bar{X} and s_X are the mean and standard deviation of X , respectively. The correlation coefficient is between -1 and 1 , whereas 1 indicates the perfect positive linear relationship and -1 indicates the perfect negative linear relationship. A value close to 0 indicates the decrease of relationship.

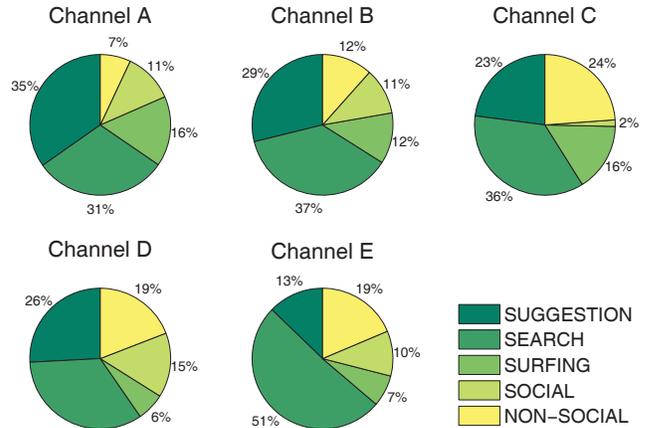


Figure 8: Breakdown of the referral source

4. ANALYSIS OF YOUTUBE INSIGHT’S REFERRERS DATA

YouTube users have various means to reach YouTube videos. The last webpage where the viewers came to land on a YouTube watch page is called *referral sources*. Understanding referrals is essential for YouTube partners to adapt their user engagement strategy. The *referrers* dataset provides this valuable information. We can classify the referral sources into five categories:

- **SUGGESTION.** The referral came from YouTube’s related video links;
- **VIDEO SEARCH.** The referral came from YouTube or Google search results;
- **YOUTUBE SURFING.** The referral came from any YouTube pages except related video links and search results, including annotation links, YouTube channel pages, subscriber links, paid and unpaid YouTube promotion, and other pages on YouTube;
- **SOCIAL REFERRAL.** The referral source was a link on an external web page, or the video was embedded on an external web page;
- **NON-SOCIAL DIRECT.** YouTube Analytics did not identify a referral source, indicating that the viewer navigated directly to the video, e.g., by copying and pasting the URL.

4.1 Referral Sources

Figure 8 shows the breakdown of the above five categories for each channel. It is clear that the breakdown percentages are channel-dependent. For example, one-third of the users reach Channel A videos from suggested videos, and one-third reach from search results; Channel B and Channel D are similar to Channel A; very few users reach Channel C videos from external sources; half of the users reach Channel E videos from search results.

This observation confirms and extends the findings in the previous works [8, 9] that the search results and related videos are the top sources of views. In addition, the large portion of referrals from SUGGESTION and YOUTUBE SURFING confirms our hypothesis of subscribers will watch more

videos as discussed in Section 3. This breakdown provides YouTube partners helpful knowledge about the traffic source and audience behaviour.

- With respect to **YOUTUBE SURFING**, user watching behaviours mainly depends on YouTube services, such as playlists and trending videos. Moreover, related videos (**SUGGESTION**) are also affected by the ranking algorithm of YouTube. Therefore, YouTube partners are suggested to abide by the guidelines specified in YouTube’s Creator Playbook⁴.
- From **SOCIAL REFERRAL**, YouTube partners can gain knowledge about external websites that brought more views. Thus, partners can further establish relationships with those websites, or create a new relationship with the websites that do not bring in significant external traffic.
- From **VIDEO SEARCH**, correlation between search keywords and video popularity is worth studying. YouTube partners can improve the search rank of their videos by tagging those videos with popular yet relevant keywords. This interesting topic is out of the scope of this paper, and we will investigate it in the future studies.
- No further information is currently available for **NON-SOCIAL DIRECT**, and thus we cannot investigate this category further.

In the following, we specifically investigate the impact of **SUGGESTION** and **SOCIAL REFERRAL** on video views.

4.2 YouTube Suggestion

As **SUGGESTION** is one of the most important referrals, we examine how related videos contribute to video views. In the **referrers** dataset, the video IDs of the referral related videos are recorded. Therefore, we calculate the percentage of referral related videos that are from the same channel, that is, the referral related videos belong to YouTube partners themselves. It is known that a YouTube partner can have multiple channels and can organize them as a network. Therefore, we further calculate the percentage of referral related videos that are not from the same channel but from the same content network, that is, from the other channels in our data.

Figure 9 shows the results. We can see that over half of the referral related videos are from the same channel, indicating the importance of the related videos for YouTube partners. We also find that the percentage for **Channel C** is much higher than the other channels, reaching over 85%, suggesting that related video source is more important for this channel. There is also a small percentage of referral related videos from different channels that are within the same content network, confirming that other channels in the same content network can contribute to video views.

We further calculate the ratio between the referral related videos that are from the same content network and the total views, (the line shown in Figure 9). The ratios are between 0.1 and 0.2, implying the varying importance of video suggestion for different channels. For example, the ratio for **Channel E** is relatively low, because the most important view source is the search result (see Figure 8). Nevertheless, video

⁴<http://www.youtube.com/yt/playbook/>

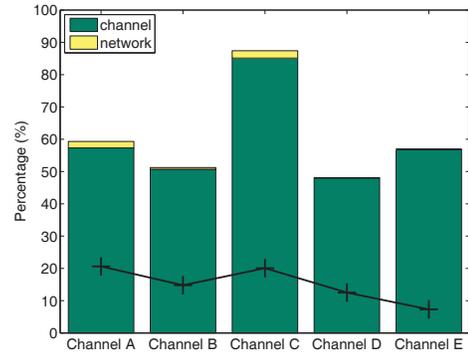


Figure 9: Percentage of referral related videos from the same channel/network

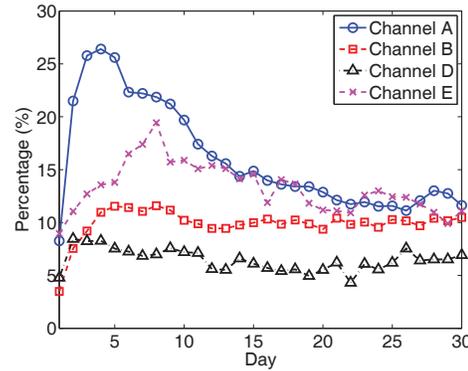


Figure 10: Average percentage of views from **SOCIAL REFERRAL** in the first 30 days

suggestion is an important view source, and thus YouTube partners are encouraged to upload more relevant videos.

4.3 External Website Referral

Although **SOCIAL REFERRAL** is not the top source for views, it does not indicate that external website referral can be ignored. Similar to the impact of subscriptions, there is a great chance that users will watch more videos from the related video list after finishing watching one.

Table 2 lists the top-five external website sources for each channel. We do not disclose the specific names of those websites except notable social networking service such as Facebook, Twitter, and Reddit, or general descriptions. From the table, we observe the channel-dependency characteristic again. No external website dominates the external referrals for **Channel A**, and the small percentage indicates that there are many sources. In **Channel E**, over 60% of the referrals are from Facebook, over 20 times more. Facebook also dominates for **Channel B**, yet the percentage is not as high as **Channel E**. Facebook is the second for both **Channel C** and **Channel D**, while the first ones have a high percentage. In summary, this result gives YouTube partners a clearer strategy for promoting their videos, e.g., by establishing a relationship with those websites.

To understand the impact of external website referrals during the lifetime of a video, we specifically looked at the percentage of views from **SOCIAL REFERRAL** on each day for the first 30 days after the video was uploaded, and calculate the average. The results for each channel are shown

Table 2: Summary of Top External Websites Referrers

	Channel A	Channel B	Channel C	Channel D	Channel E
1st	9.0% downloading site	16.2% Facebook	31.9% gaming wiki	41.2% Reddit	62.4% Facebook
2nd	4.4% Facebook	2.2% not available	7.6% Facebook	9.9% Facebook	2.4% music streaming
3rd	2.6% forum	1.5% downloading site	5.3% gaming blog	4.7% Twitter	2.0% music blog
4th	1.7% gaming site	1.2% not available	5.1% gaming site	2.0% blog	2.0% Twitter
5th	1.5% gaming site	0.9% downloading site	3.7% Internet video site	1.7% entertainment site	1.6% music blog

in Figure 10. Note that, as only 2% of views came from external sources for Channel C (refer to Figure 8), we do not include this channel.

All four channels have notable increasing trend in the first few days, and then their views start decreasing or remain constant. Channel A, Channel B, and Channel D reach their peak at the fourth, fifth, and second day, while Channel E reaches the peak later, at the eighth day. To explain that, the first three channels contain trailer videos of various genres, so it is very likely that users are expecting such new games, movies, and TV shows, and thus the videos extensively spread in the external websites; while Channel E contains music videos from independent artists, and thus videos need more time to be noticed.

The peak values of Channel A and Channel E are greater than that of Channel B and Channel D, and are much higher than the overall percentage (note that in Figure 8, the percentage of SOCIAL REFERRAL views are 11%, 11%, 15%, and 10%, respectively), implying that external sources are important to these two channel in the early stage. Moreover, the percentage of Channel B is almost constant and is similar to the overall percentage (11%) after reaching the peak, indicating that although not driving as much percentage of traffic as Channel A and Channel E do, external sources remain impacting the video views for a long time. On the other hand, the percentage of Channel D is much lower than the overall percentage (15%), mainly because some very popular videos have more SOCIAL REFERRAL views, as the deviation of the result is quite large (between 10% and 15%).

5. CONCLUSION

YouTube partners have changed YouTube content distribution landscape by uploading and monetizing premium videos. YouTube provides Insight Analytics to these partners. In this paper, using up to three years of Insight data of five YouTube partners' channels, we have revealed a number of unique features of YouTube partners. First, the viewing surges mostly occur soon after the video is uploaded. This characteristic suggests YouTube partners to utilize this crucial duration to promote their videos, for example, by sharing on social media websites. Second, attractiveness is more important for non-informative videos than informative videos, therefore, subscription is extremely crucial for those channels that mostly contain non-informative videos.

We have also studied the breakdown of referral sources, and specifically investigated the impact of video suggestion and external website referrals on video views. We have found that video suggestion is an important view source, and have suggested YouTube partners to upload more relevant videos. Moreover, we have found that the percentage of SOCIAL REFERRAL is much higher in the first week. This characteristic implies that external sources are important in videos' early stage, and YouTube partners are encouraged to promote their videos in order to create viewing surges soon

after uploading. This study is of great value to YouTube partners for attracting more views and subsequently generating more revenue.

Acknowledgment

Part of this research is supported by a Canada NSERC Discovery Grant, an NSERC Engage grant, and an NSERC Strategic Project Grant.

6. REFERENCES

- [1] M. Georgieva, "Americans Spend Nearly 3.5 Hours Per Week Watching Online Video [Data]," <http://blog.hubspot.com/blog/tabid/6307/bid/11888/Americans-Spend-Nearly-3-5-Hours-Per-Week-Watching-Online-Video-Data.aspx>, 2011.
- [2] "YouTube Press," <https://www.youtube.com/yt/press/statistics.html>.
- [3] "YouTube Elevates Most Popular Users to Partners," <http://youtube-global.blogspot.com/2007/05/youtube-elevates-most-popular-users-to.html>, 2007.
- [4] 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 *Proc. of ACM IMC*, 2007.
- [5] X. Cheng, C. Dale, and J. Liu, "Statistics and Social Network of YouTube Videos," in *Proc. of IWQoS*, 2008.
- [6] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, "YouTube Traffic Characterization: A View From the Edge," in *Proc. of ACM IMC*, 2007.
- [7] A. Mislove, M. Marcon, K. P. Gummadi, P. Dreschel, and B. Bhattacharjee, "Measurement and Analysis of Online Social Networks," in *Proc. of ACM IMC*, 2007.
- [8] F. Figueiredo, F. Benevenuto, and J. M. Almeida, "The Tube over Time: Characterizing Popularity Growth of YouTube Videos," in *Proc. of ACM WSDM*, 2011.
- [9] R. Zhou, S. Khemmarat, and L. Gao, "The Impact of YouTube Recommendation System on Video Views," in *Proc. of ACM IMC*, 2010.
- [10] G. Gürsun, M. Crovella, and I. Matta, "Describing and Forecasting Video Access Patterns," in *Proc. IEEE INFOCOM*, 2011.
- [11] X. Cheng and J. Liu, "Exploring Interest Correlation for Peer-to-Peer Socialized Video Sharing," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, vol. 8, no. 1, pp. 5:1–5:20, January 2012.