



<b>Title</b>	<b>Incorporating the position of sharing action in predicting popular videos in online social networks</b>
<b>Author(s)</b>	<b>Long, Y; Li, VOK; Niu, G</b>
<b>Citation</b>	<b>The 15th International Conference on Web Information System Engineering, Thessaloniki, Greece, 12-14 October 2014. In Lecture Notes in Computer Science, 2014, v. 8787, p. 125-140</b>
<b>Issued Date</b>	<b>2014</b>
<b>URL</b>	<b><a href="http://hdl.handle.net/10722/217392">http://hdl.handle.net/10722/217392</a></b>
<b>Rights</b>	<b>The final publication is available at Springer via <a href="http://dx.doi.org/10.1007/978-3-319-11746-1_9">http://dx.doi.org/10.1007/978-3-319-11746-1_9</a></b>

# Incorporating the Position of Sharing Action in Predicting Popular Videos in Online Social Networks

Yi Long, Victor O.K. Li, and Guolin Niu

Department of Electrical and Electronic Engineering  
The University of Hong Kong  
Pokfulam Road, Hong Kong, China  
{yilong,vli,glniu}@eee.hku.hk

**Abstract.** Predicting popular videos in online social networks (OSNs) is important for network traffic engineering and video recommendation. In order to avoid the difficulty of acquiring all OSN users' activities, recent studies try to predict popular media contents in OSNs only based on a very small number of users, referred to as experts. However, these studies simply treat all users' diffusion actions as the same. Based on large-scale video diffusion traces collected from a popular OSN, we analyze the positions of users' video sharing actions in the propagation graph, and classify users' video sharing actions into three different types, i.e., initiator actions, spreader actions and follower actions. Surprisingly, while existing studies mainly focus on the initiators, our empirical studies suggest that the spreaders actually play a more important role in the diffusion process of popular videos. Motivated by this finding, we account for the position information of sharing actions to select initiator experts, spreader experts and follower experts, based on corresponding sharing actions. We conduct experiments on the collected dataset to evaluate the performance of these three types of experts in predicting popular videos. The evaluation results demonstrate that the spreader experts can not only make more accurate predictions than initiator experts and follower experts, but also outperform the general experts selected by existing studies.

**Keywords:** Online Social Networks, Information Diffusion, Video, Prediction

## 1 Introduction

Online social networks (OSNs), such as Facebook and Twitter, have now become an indispensable platform for users to access and share videos which are originally hosted by video sharing sites (VSSes), such as YouTube, generating enormous traffic to these VSSes [15]. As is revealed by existing studies, the distribution of video popularity is highly skewed, with the top 0.31% most popular videos accounting for nearly 80% view requests [16]. Therefore, it is important to

identify these popular videos based on their diffusion progress in the early stage, so as to optimize the network traffic engineering system, such as the caching strategy, and to improve the video recommender system in OSNs.

Although considerable studies have been conducted to predict the popularity of videos, most of these proposed methodologies are based on all users' actions in the early stage [17] [19] [22]. Considering that it is rather expensive to acquire all users' diffusion actions in popular OSNs, which usually have millions of users, recent studies try to predict media content popularities based on only a small number of users, which are referred to as "experts" [11] [14] or "influencers" [2]. However, these studies mainly focus on other types of media contents, e.g., Wikipedia edits [14] or tweets [2] [11]. Moreover, while these studies simply treat all users' diffusion actions as the same, researchers working on the information diffusion process argue that diffusion actions of users are actually different and depend on their positions in propagation graphs [20] [21] [23]. Therefore, it is meaningful to account for the position information of diffusion actions when selecting experts, and evaluate the performance of selected experts in predicting popular videos.

To fill the above research gaps, we select Renren, the most popular Facebook-like OSN in China, as our research platform, and collect substantial diffusion traces from it. Based on the collected traces, we firstly classify users' video sharing actions of an individual video into three different groups, i.e., initiator actions, spreader actions and follower actions, according to their positions in diffusion process. Then we conduct comparison studies on the impact of these different sharing actions on video diffusion. Motivated by the importance of spreader in distinguishing popular videos, we further incorporate sharing actions' position information into the selection of experts, enabling us to identify three different groups of experts, i.e., initiator experts, spreader experts and follower experts. Experiments based on real dataset have been conducted to evaluate the performance of these experts in predicting popular videos. The main contributions of this paper are summarized as follows:

- Large-scale video diffusion traces, covering the video sharing actions of 2.8 million users for more than 3 years, have been collected from a popular OSN. According to their positions in the diffusion process, users' sharing actions have been classified into three different types, i.e., initiator actions, spreader actions and follower actions.
- When attempting to identify the initiators of information diffusion, existing studies adopt two different definitions of sharing actions, namely, actions which share media contents independently [20] [24], and those that import media contents from external sites [24] [25]. Based on our collected traces, we investigate these two types of initiator actions, and propose a more comprehensive definition of initiator actions.
- Different from previous empirical studies which mainly focus on characterizing initiator actions [20] [21], we conduct comprehensive studies on all the three types of sharing actions, i.e., initiator actions, spreader actions and follower actions. Surprisingly, while initiator actions are regarded to be piv-

total in theoretical diffusion models [13], our observations show that spreader actions play a more important role in driving video popularity.

- By incorporating positions of sharing actions into the selection of experts, we can identify three different groups of experts, i.e., initiator experts, spreader experts and follower experts. The experiments conducted on the collected real traces demonstrate that spreader experts outperform not only the initiator experts and follower experts, but also the general experts [11], selected without regard to the positions of their sharing actions.

The rest of this paper is organized as follows. Section 2 introduces the video diffusion process of Renren as well as the dataset collected from Renren. Section 3 defines the three types of sharing actions and investigates their roles in video diffusion. Section 4 first introduces how to select the three different types of experts, and then evaluates the popular video prediction performance of these experts. Finally, in Section 5 we summarize our studies and discuss future work.

## 2 Data collection and basic properties

First, we will give a brief introduction of our research platform, Renren, as well as our video diffusion traces collected from Renren.

### 2.1 Background of The Renren Social Network

Renren is one of the most popular OSNs in China, and has more than 190 million users. It adopts almost the same interface as well as functionalities of Facebook, and hence is called “the Twin of Facebook in China”. In the past few years, an increasing number of users in Renren tend to watch and share/import videos which are originally hosted by external VSSes [16]. In general, a video in Renren is propagated along the friendship links. Initially, if a user finds an interesting video in a VSS, he/she can import this video to Renren by posting the URL of the video in Renren. Then this video will appear in the news feeds of all his/her friends. If any of these friends also think this video is interesting after watching it, he/she may further share this video with his/her friends by simply clicking a button, which will also forward this video to the news feeds of his/her friends, thus generating a cascade in Renren.

While such video propagation process is similar to many other popular OSNs, such as Twitter [5] and Weibo [25], Renren has two unique features making it an ideal platform for studying the video diffusion process: 1) Both Renren users’ friendship links and the complete list of all their video sharing actions are publicly accessible by any registered user of Renren <sup>1</sup>; 2) Renren maintains a separate website to store the information of each individual video sharing action, including the original URL of the video in external sites and identity of the Renren user who imports this video into Renren. Therefore, in this paper, we choose Renren as our research platform. The basic statistics as well as properties of video diffusion traces we collected from Renren are presented as follows.

<sup>1</sup> The friend lists of Renren users are protected from strangers since April, 2011.

Table 1: Basic statistics of the XJTU dataset

<b>Time period</b>	March 2008 - July 2011
<b>Number of users</b>	2,808,681
<b>Number of video sharing actions</b>	209,409,778
<b>Number of distinct videos</b>	6,416,745

## 2.2 Data Collection

Due to the large number of users in Renren, it is computationally expensive and nearly impossible for us to acquire the friend list and sharing actions of all Renren users. Fortunately, just like Facebook, Renren evolves from a university-based social network and hence is divided into regional networks with institution affiliation information. Therefore, following the sampling strategy in [12], we perform affiliation-oriented crawls of the Xi'an Jiaotong University (XJTU) network, which is composed of more than 52,685 XJTU students and is a famous online social network community in Renren. To retrieve the position information of these XJTU users' video sharing action, we also identified all of their friends, obtaining more than 2 million users. For both the collected XJTU users and their friends, we further crawled all of their video sharing actions performed in the past 3 years. The statistics of the XJTU dataset are shown in Table 1. Such a substantial dataset from Renren not only captures the information diffusion process better due to its affiliation-based sampling strategy [4], but is also more instructive for optimizing network traffic engineering (e.g., the caching strategies of content delivery network) since most of these users share the same locations and access the Internet via the campus network.

In our collected diffusion traces, each video sharing action can be represented as a tuple  $(u, v, t, s)$ , which means that at time  $t$ , user  $u$  shares video  $v$ , and this shared video is imported from an external VSS by user  $s$ . After collecting video sharing actions of all XJTU users, we can further explicitly track the diffusion process of a video in XJTU community since each video in our dataset is associated with a unique external URL.

## 2.3 Video popularity in Renren

Characterizing the distribution of video popularity is essential for many applications, e.g., the skewness of video popularity distribution will determine the caching performance of a video service provider [8]. The degree skewness for popularity distributions are usually described with the Pareto Principle or the 80/20 rule. In this paper, we define the popularity of a video as the number of sharing actions of this video, and calculate the popularities of the 6.4 million videos accounted by our 2.8 million users. To gain insights into the Pareto Principle of these videos, we investigate the cumulative percentage of video popularity for the  $i$  most popular videos (shown in Fig. 1). The horizontal axis sorts

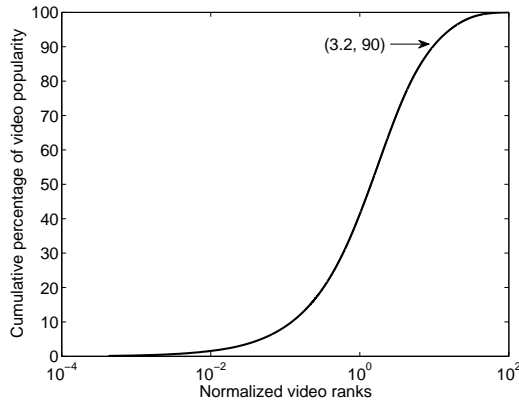


Fig. 1: Test of Pareto principle in video popularity.

the videos from the most popular to the least popular, with the rank normalized between 0 and 100. As the coordinate (3.2,90) shows, the top 3.2% most popular videos account for 90% of sharing actions, while the remaining 96.8% of the videos account for only 10% sharing actions. Such extreme skewness of video popularity in OSN is higher than in VSSes (e.g., YouTube), where the top 20% most popular videos account for 90% [3]. This is because while the videos in VSSes are mainly accessed via the search engine or recommendation system, popularity of videos in OSNs are mainly driven by the propagation along inter-personal links. Therefore, it is quite desirable to predict these most popular videos in OSN at an early stage. In our following studies, we will select the top 3% most popular videos as *top videos*, and then conduct comparison studies on users' role in the diffusion process of these *top videos* and all common videos, trying to find which group of users make those *top videos* popular.

### 3 Sharing Actions in Different Positions of Video Diffusion

Following an approach used in a study of information propagation on Twitter [2], we can track the propagation of a video  $v$  among our studied users based on the their friendship links as well as sharing actions of video  $v$ : if user  $u_1$  and user  $u_2$  are friends, and user  $u_1$  shared video  $v$  earlier than user  $u_2$ , we say user  $u_2$  is influenced by user  $u_1$  to share video  $v$ . Note that a user may have more than one friend sharing a video before him/her, and hence may be influenced by multiple users. Then we can construct the propagation graph of a video by recording a directed edge from user  $u_1$  to user  $u_2$  if user  $u_2$  is influenced by user  $u_1$  to share this video. As is discussed in previous studies [7], any propagation graph constructed as above is always a directed acyclic graph: it is directed, each node can have zero or more parents, and cycles are impossible due to the time

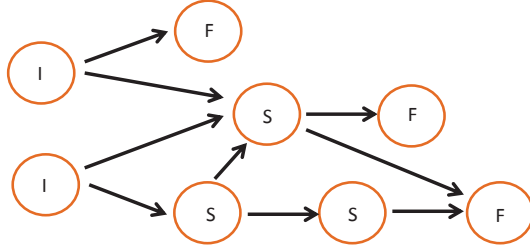


Fig. 2: Video sharing actions with different positions in a propagation graph.

constraint. Based on the propagation graph of a video, we classify sharing actions of this video into three types according to their positions in the propagation graph. Figure 2 describes a simple propagation graph of a video as an example to illustrate our definition of the three types of sharing action. We define the video sharing actions without parents (shown as I) to be **initiator actions**, video sharing actions which have both parents and children (shown as S) to be **spreader actions**, and video sharing actions without children (shown as F) to be **follower actions**.

### 3.1 Definition of Video Initiators

In fact, in addition to the criteria discussed above, various other criteria [20] [21] [23] [24] [25] have been proposed to classify users' diffusion actions into different types according to their positions in the propagation graph. Although they all investigated initiator actions which start a chain in the propagation graph, they adopted different definition of initiator actions. In particular, while studies working on the diffusion process of Facebook page [21], tweets [20] and email [24] define the initiator actions to be the actions which share media contents independently, i.e., they share media contents when none of their friends have shared them before, recent work focusing on the diffusion process of videos [24] [25] believe the root nodes of video propagation graphs are sharing actions which import videos from external VSSes. Therefore, it is necessary to conduct a comparison studies on these two different definitions of video initiators. Fortunately, based on our collected traces, we can not only obtain the sharing actions which share video independent of friends following the method discussed above, but also identify those sharing actions which import videos from external VSSes by utilizing the last entry of each sharing action  $(u, v, t, s)$ , i.e., who imports this video  $v$  into Renren. Hence, for any video  $v$ , we build a set of sharing actions which share video  $v$  independently (denoted as action set  $A$ ), as well as a set of sharing actions which import video  $v$  from external VSSes (denoted as action set  $B$ ). To explore the differences and similarities of action set  $A$  and action set  $B$  for each video, we further divide all the actions in  $(A \cup B)$  into three disjoint action sets, file:///C:/Users/Neal/Desktop/WISE 2014 revised/User Position

Final V2.pdf.i.e.,  $(A \cap B)$ ,  $(A - B)$  and  $(B - A)$ . The physical meanings of these three sets are discussed as follows.

- Set  $(A \cap B)$  includes each sharing action which imports a video from VSSes into Renren when none of the friends have shared this video before. Such actions usually happen in the following way: a user finds an interesting video in VSSes, and it has not been shared by any of his/her Renren friends, then he/she imports this video to Renren.
- Set  $(A - B)$  includes each sharing action of a video which has been imported by stranger into Renren when none of the friends of a user have shared this video before. Note that besides forwarding the video shared by each user’s friends to him/her, Renren also recommends hundreds of site-wide popular videos which have been imported by any Renren users to all Renren users every day.
- Set  $(B - A)$  includes each sharing action which imports a video from VSSes even though one or more of the friends have already shared this video before. Note that such actions are still important as Renren users may miss the news feeds from their friends due to the “visibility and divided attention” [10].

We investigate the average fraction of the above three sets in set  $(A \cup B)$  for all videos, and *top videos*, respectively (shown in Fig. 3). The obvious difference between all videos (shown in Fig. 3a) and *top videos* (shown in Fig. 3b) is that while  $(A \cap B)$  dominates in all videos,  $(A - B)$  becomes the majority in *top videos*. This shows that by broadcasting a video to users who are not socially connected, the recommender system of Renren may enhance the diffusion of the video efficiently.

On the other hand, both Fig. 3a and Fig. 3b demonstrate that the difference of sets  $A$  and  $B$ , i.e.,  $(A - B)$  and  $(B - A)$ , is non-trivial. Hence, simply defining the initiator action to be actions in either set  $A$  or  $B$  would underestimate the existence of initiators of a video. In our following studies, we define the initiator actions to be  $A \cup B$ . This comprehensive definition is reasonable because no matter a user shares a video independently, or directly from VSSes, under both situations, he/she is activated by external influence [18], rather than by the diffusion over the friendship of the social network.

### 3.2 Characteristics of Sharing Actions with Different Position

Based on the positions of sharing actions in the corresponding propagation graphs as well as the definitions discussed earlier, we then classify our collected sharing actions into three types: initiator actions, spreader actions and follower actions. While initiator actions are believed to play a key role in the diffusion process [13] and hence have been intensely investigated in existing empirical studies [20] [21], hardly any work has been done to characterize spreader actions and follower actions in real traces. In the following, we will investigate the distribution of the elapsed time as well as the proportion of the size for all the three types of sharing actions. In particular, in order to understand which



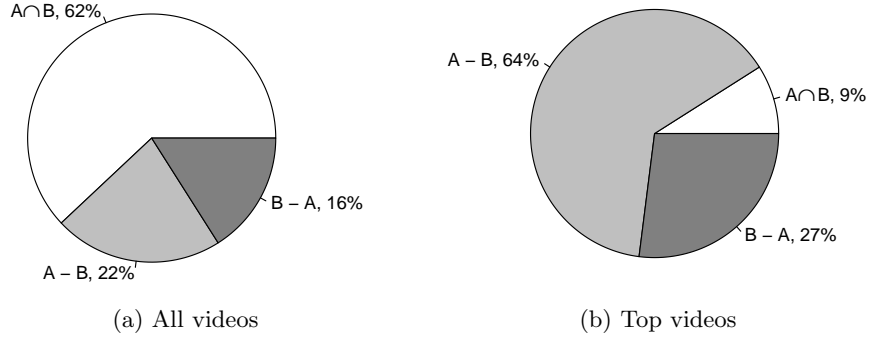


Fig. 3: Breakdown of initiator actions for all videos and for top videos.

group of actions make these popular videos outstanding, we will compare the observation obtained from all videos and *top videos*.

**The elapsed time for different sharing actions** The elapsed time of sharing action  $(u, v, t, s)$  is defined to be the difference of the time  $t$  and the *birth time* of video  $v$ . Here, the *birth time* of video  $v$  is the time that video  $v$  is firstly shared by any user in our studied community. According to the definition of “influence” as well as the three types of sharing actions, we would expect that elapsed time of initiator actions, on average, will be smaller than the elapsed time of spreader actions, and the elapsed time of spreader action will be, on average, smaller than the elapsed time of follower. To verify this intuition, we will investigate the elapsed time distribution of initiator actions, spreader actions and follower actions for all videos and *top videos* respectively (as shown in Fig. 4). While the elapsed times of initiator actions of all videos (as shown in Fig. 4a) and *top videos* (as shown in Fig. 4b) are, on average, smaller than spreader actions and follower actions, which accords our intuition, there is no distinct difference between the CDF of elapsed time for all videos and *top videos*. In contrast, for *top videos*, the elapsed time of spreader actions is, on average, smaller than follower actions. Therefore, “spreader” propagation is an important factor to make a video popular.

**The fraction of different sharing actions** Considering that the above studies on the distribution of elapsed time within each type of sharing actions cannot reveal the size of each type of sharing actions, here we further investigate the fraction of each type of sharing actions for all videos and *top videos*, respectively (as shown in Fig. 5). We observe that, while all videos are mainly driven by initiator actions, *top videos* are driven by spreader actions.

In summary, both the studies on elapsed time and on the fraction of sharing actions suggest that spreader actions play an essential role in making popular videos outstanding.

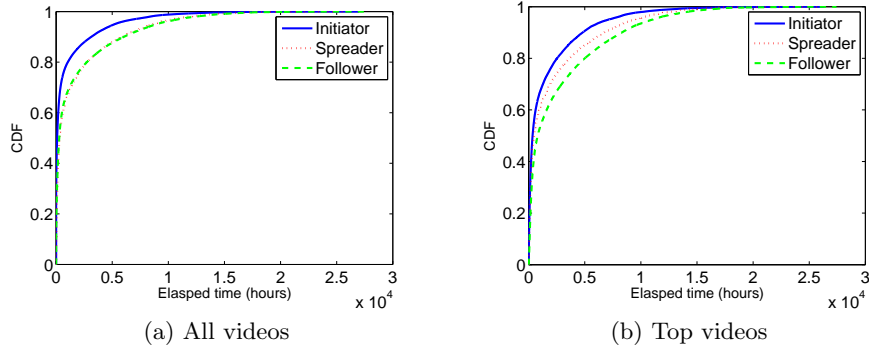


Fig. 4: CDF of elapsed time for initiator actions, spreader actions and follower actions

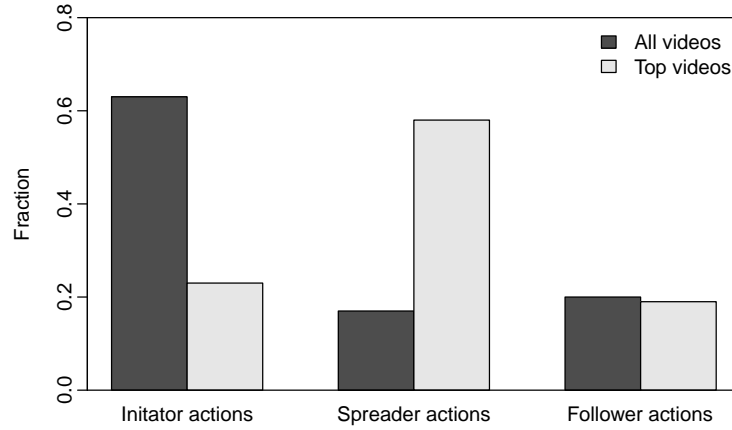


Fig. 5: Fraction of different sharing actions

## 4 Predicting Popular Videos Based on Specified Experts

Although the above empirical studies demonstrate that spreader actions are important for the diffusion process of popular videos, it is difficult to make a rigorous comparison on the importance of the three different types of actions due to their different sizes (as shown in Fig. 5). In this section, we will try to mitigate the impact of the size imbalance by selecting an equal and small number of “experts” based on the three types of sharing actions, respectively, to compare the performance of these three types of actions in predicting popular videos.

In fact, it is highly desirable to predict the information diffusion process by only relying on the profile or decisions of a very small number of experts because

of the huge population in current OSNs. Various studies have been conducted to identify experts [11] [14], also known as “influencers” [2]. In the following, we will adopt the CI model proposed in [11] to distill experts from common users since it is based on users’ past sharing actions rather than profiles. While [11] utilized all users’ sharing actions without regard to their positions in information diffusion, we will first classify users’ sharing actions into three different groups, i.e., initiator actions, spreader actions and follower actions, and then select three different groups of experts accordingly, i.e., initiator experts, spreader experts and follower experts.

#### 4.1 Expert Selection Model

Basically, under the CI model [11], the experts are supposed to be able to find and then share/import *popular* videos at an *early* stage. Therefore, the following two criteria are proposed to distinguish these experts from common users.

- 1) Precision: Here we use the precision term to measure the ability of a user to identify and share popular videos. Supposed we want to identify the top  $r$  percent most popular videos, referred to as *golden videos* here, the precision of a user based on a given set of his/her sharing actions is defined as

$$\text{precision} = \frac{\text{number of sharing actions of golden videos in given action set}}{\text{number of all sharing actions in given action set}} \quad (1)$$

- 2) Promptness: As is discussed above, besides the ability to make the right recommendation, i.e., find and share the popular videos, experts are also expected to make the decision promptly. Otherwise, common users could trick our selection criteria by sharing a lot of popular but outdated videos. Hence, we will set a promptness threshold, and only consider sharing actions with elapsed time smaller than the promptness threshold.

In our following experiments, we will use the *precision with promptness threshold* (referred to as *precision in the following for simplicity*) to measure a user’s ability to find and share popular videos promptly. Based on this metric, we can then rank users’ precisions, ordered from high to low, and then select the top  $t$  percent users as experts. However, such a methodology only focuses on the percentage of desirable video sharing actions, i.e., sharing actions of golden videos, while ignoring the actual number of desirable sharing actions. In other words, if a user  $u_1$  shares one video promptly and one video only, and this only video happens to be a golden video, while another user  $u_2$  shares 100 videos promptly and all 100 videos are golden videos, then user  $u_1$  and user  $u_2$  will be regarded to have the same ability to find and share popular videos promptly in our selection model. Obviously, this result is not reasonable because while user  $u_1$  is likely to choose the single shared video by luck, we have more confidence to regard user  $u_2$  as an expert. Therefore, instead of ranking user according to the precision proposed above, we need to further take the uncertainty into consideration. It is known that confidence level is an intuitive metric to measure

the uncertainty of a value. In particular, considering that precision calculated in Equation (1) is binomial, we employ the adjusted Wald confidence interval [1] to measure the uncertainty of the precision. Then in practice, we will use the lower bound of the precision under the 95% confidence interval to rank users, and choose the top  $k\%$  users with the highest lower bound to be experts. Next we will conduct experiments to evaluate the performance of selected experts on predicting popular videos at an early stage.

## 4.2 Experiments on Predicting Popular videos

Note that given different sharing actions of a user, his/her precision as well as the lower bound will be different according to Equation (1). Thus, based on the three different types of users' video sharing actions discussed in Section 3, we can obtain three different sets of experts, i.e., initiator experts, spreader experts and follower experts, by selecting the top  $k\%$  users with the highest lower bounds calculated from the corresponding set of sharing actions. Based on the large-scale video diffusion traces collected from Renren, we conduct experiments to evaluate and compare the performance of the three groups of experts in predicting popular videos. Next, we introduce the dataset used in our experiments.

**Selection Dataset and Evaluation Dataset** Our selection of experts is based on all users' past video sharing actions, and we refer to these video sharing actions as *selection dataset*. Then we can evaluate the performance of selected experts in predicting popular videos based on another set of video sharing actions, referred to as *evaluation dataset*. Following previous work [11] and also considering that our studied users are most active (i.e., the number of sharing actions is largest) from March 1st to June 30th, 2011, we choose videos whose birth time is within these 4 months. In particular, the selection dataset contains sharing actions of videos with birth times between March 1st to April 30th, and the evaluation dataset contains sharing actions of videos with birth times between May 1st to June 30th.

**Parameter Settings and Evaluation Method** Given a set of experts, either initiator experts, spreader experts or follower experts, we can then, based on the evaluation dataset, count how many times each video has been shared by these experts within the promptness window, and regard the counted times as the number of votes it has received from experts. Then we further rank all the videos in the evaluation dataset by the number of votes they have received in descending order, and choose the top  $n$  videos as the predicted popular videos. Based on the ordered videos as well as the real golden videos, we then employ the precision-recall curve [6] to evaluate the precision as well as recall score of our predicted popular videos under all possible values of  $n$ . In fact, the precision-recall curve has been widely used in the performance evaluation of information retrieval algorithms, especially for applications with highly skewed datasets, such as our dataset which has much less golden videos than common videos. In general,

the optimization goal of the precision-recall curve is to be in the upper-right-hand corner, which means that in reality, a precision-recall which is closer to the upper-right-hand corner will have a better performance.

Next, we introduce the detailed experimental parameter settings. Following previous studies on predicting video [19] [22], we will try to predict the popular videos at the end of 30 days based on their sharing actions occurring within their first 7 days. This means that the promptness threshold is set to 7 days, and video popularity is set to be the popularity at 30 days. The golden videos are the top 3% because such videos account for nearly 90% sharing actions. As to the ratio of experts, we have tried the values of 0.01, 0.02, 0.05, and 0.1. It is observed that the following conclusions can hold on all these values. Here we will only present the results with the ratio of experts being set to 0.02 due to space limitations.

**Performance Evaluation for Different Experts** Recall that after identifying the experts based on all users’ video sharing actions in the selection dataset, our objective is to predict the popular videos only based on the sharing actions of selected experts since this can avoid the difficulty of acquiring all users’ actions over a long period. As a result, we cannot classify the video sharing actions of selected experts into three types of actions due to the fact that we are not supposed to be aware of all users’ sharing actions except for experts. Therefore, when we predict popular videos based on selected experts, no matter they are initiator experts, spreader experts or follower experts, we will make use of all of their video sharing actions. The performance evaluation results of the above three types experts are presented in Fig. 6, and for comparison, we also present the performance of the general experts [11], which is selected without distinguishing users’ sharing actions by their positions.

As mentioned earlier, a precision-recall which is closer to the upper-right-hand corner will have a better performance. While it is intuitive that a follower expert achieves the worst performance due to its position in the diffusion process, two unexpected and important conclusions can be drawn from the precision-recall curves of different experts’ prediction results (as shown in Fig. 6). Spreader experts outperform both the initiator experts and follower experts. This observation contradicts existing work [13] [20] [21] but accords with our earlier empirical observations in Section 3.2; 2) spreader experts even outperform the general experts proposed in existing work, which demonstrates the necessity of distinguishing sharing actions by their positions.

Motivated by the fact that spreader experts would outperform the general experts in existing work, we will further verify the long-lasting debate “Do experts make better decisions than crowds?” [9]. In our context, the “crowd” corresponds to all of the studied users, except the selected experts. We also conduct random sampling on crowds to construct smaller crowds with (2%, 5%, 10%, 33%) of the complete crowd. The performance of these crowds as well as our proposed spreader expert is shown in Fig. 7. The results show that, similar to the conclusion in [9], the spreader experts cannot outperform the complete crowd. However,

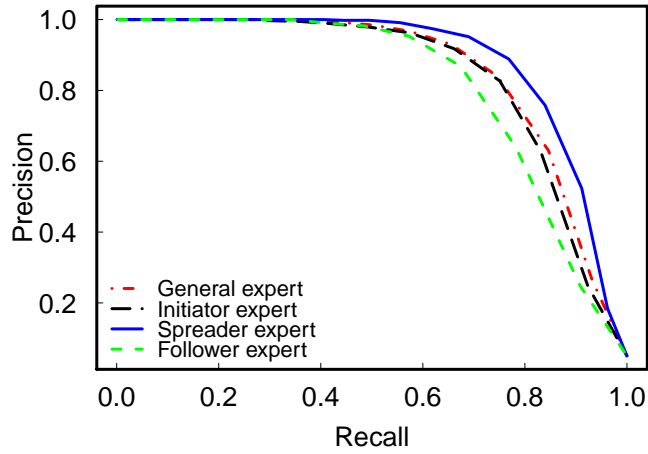


Fig. 6: Precision-recall curve of the prediction performance for different experts

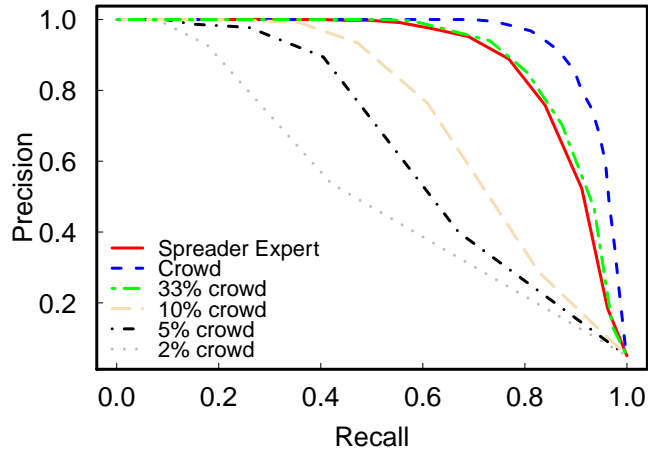


Fig. 7: Precision-recall curve for spreaders experts and crowds

our selected spreader experts can achieve equivalent accuracy of the 33% crowd while the general experts in [11] can only perform as well as the 20% crowd.

## 5 Conclusion and Future Work

This paper works on predicting popular videos only based on the video sharing actions of a few users, referred to as experts. Based on the huge diffusion traces collected from a popular OSN, we first classify users' sharing actions into three different types, i.e., initiator actions, spreader actions and follower actions, according to their positions in the propagation graph. Different from existing

work, our empirical studies show that spreader actions are essential to make popular videos outstanding. Therefore, we incorporate the position information into the existing expert selection model, and use this new model to select initiator experts, spreader experts and follower experts based on the corresponding sharing actions. We further evaluate the performance of these experts in predicting popular videos. The results demonstrate that spreader experts can not only outperform initiator experts and follower experts, but also the general experts selected by existing methods. Our future work will investigate whether users' centrality information in the social graph, such as betweenness centrality or PageRank score, is helpful to improve the selection of experts.

## References

1. Agresti, A., Coull, B.A.: Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician* 52(2), 119–126 (May 1998)
2. Bakshy, E., Hofman, J.M., Mason, W.A., Watts, D.J.: Everyone's an influencer: Quantifying influence on twitter. In: *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*. pp. 65–74. WSDM '11, ACM, New York, NY, USA (2011)
3. Cha, M., Kwak, H., Rodriguez, P., Ahn, Y.Y., Moon, S.: Analyzing the video popularity characteristics of large-scale user generated content systems. *IEEE/ACM Trans. Netw.* 17(5), 1357–1370 (Oct 2009)
4. Choudhury, M.D., Lin, Y.R., Sundaram, H., Candan, K.S., Xie, L., Kelliher, A.: How does the data sampling strategy impact the discovery of information diffusion in social media? In: Cohen, W.W., Gosling, S. (eds.) *ICWSM*. The AAAI Press (2010)
5. Christodoulou, G., Georgiou, C., Pallis, G.: The role of twitter in youtube videos diffusion. In: Wang, X., Cruz, I., Delis, A., Huang, G. (eds.) *Web Information Systems Engineering - WISE 2012, Lecture Notes in Computer Science*, vol. 7651, pp. 426–439. Springer Berlin Heidelberg (2012)
6. Davis, J., Goadrich, M.: The relationship between precision-recall and roc curves. In: *Proceedings of the 23rd International Conference on Machine Learning*. pp. 233–240. ICML '06, ACM, New York, NY, USA (2006)
7. Goyal, A., Bonchi, F., Lakshmanan, L.V.S.: A data-based approach to social influence maximization. *Proc. VLDB Endow.* 5(1), 73–84 (Sep 2011)
8. Hefeeda, M., Saleh, O.: Traffic modeling and proportional partial caching for peer-to-peer systems. *IEEE/ACM Trans. Netw.* 16(6), 1447–1460 (Dec 2008)
9. Hill, G.W.: Group versus individual performance: Are  $n+1$  heads better than one? *Psychological Bulletin* 91(3), 517 (1982)
10. Hodas, N., Lerman, K.: How visibility and divided attention constrain social contagion. In: *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Conference on Social Computing (SocialCom)*. pp. 249–257 (2012)
11. Hsieh, C.C., Moghbel, C., Fang, J., Choo, J.: Experts vs the crowd: Examining popular news prediction performance on twitter. In: *Proceedings of the 22Nd International Conference on World Wide Web. WWW '13, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland* (2013)

12. Jiang, J., Wilson, C., Wang, X., Huang, P., Sha, W., Dai, Y., Zhao, B.Y.: Understanding latent interactions in online social networks. In: Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement. pp. 369–382. IMC '10, ACM, New York, NY, USA (2010)
13. Kempe, D., Kleinberg, J., Tardos, E.: Maximizing the spread of influence through a social network. In: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 137–146. KDD '03, ACM, New York, NY, USA (2003)
14. Kittur, A., Pendleton, B.A., Suh, B., Mytkowicz, T.: Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. In: Proceedings of the 25th Annual ACM Conference on Human Factors in Computing Systems (CHI 2007). ACM (Apr 2007)
15. Lai, K., Wang, D.: Understanding the external links of video sharing sites: Measurement and analysis. *Multimedia, IEEE Transactions on* 15(1), 224–235 (2013)
16. Li, H., Liu, J., Xu, K., Wen, S.: Understanding video propagation in online social networks. In: Proceedings of the 2012 IEEE 20th International Workshop on Quality of Service. pp. 21:1–21:9. IWQoS '12, IEEE Press, Piscataway, NJ, USA (2012)
17. Li, H., Ma, X., Wang, F., Liu, J., Xu, K.: On popularity prediction of videos shared in online social networks. In: Proceedings of the 22Nd ACM International Conference on Conference on Information and Knowledge Management. pp. 169–178. CIKM '13, ACM, New York, NY, USA (2013)
18. Myers, S.A., Zhu, C., Leskovec, J.: Information diffusion and external influence in networks. In: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 33–41. KDD '12, ACM, New York, NY, USA (2012)
19. Pinto, H., Almeida, J.M., Gonçalves, M.A.: Using early view patterns to predict the popularity of youtube videos. In: Proceedings of the Sixth ACM International Conference on Web Search and Data Mining. pp. 365–374. WSDM '13, ACM, New York, NY, USA (2013)
20. Rodrigues, T., Benevenuto, F., Cha, M., Gummadi, K., Almeida, V.: On word-of-mouth based discovery of the web. In: Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference. pp. 381–396. IMC '11, ACM, New York, NY, USA (2011)
21. Sun, E., Rosenn, I., Marlow, C., Lento, T.: Gesundheit! modeling contagion through facebook news feed. In: International AAAI Conference on Weblogs and Social Media (2009)
22. Szabo, G., Huberman, B.A.: Predicting the popularity of online content. *Commun. ACM* 53(8), 80–88 (Aug 2010)
23. Wang, D., Wen, Z., Tong, H., Lin, C.Y., Song, C., Barabási, A.L.: Information spreading in context. In: Proceedings of the 20th International Conference on World Wide Web. pp. 735–744. WWW '11, ACM, New York, NY, USA (2011)
24. Wang, Z., Sun, L., Chen, X., Zhu, W., Liu, J., Chen, M., Yang, S.: Propagation-based social-aware replication for social video contents. In: Proceedings of the 20th ACM international conference on Multimedia. pp. 29–38. MM '12, ACM, New York, NY, USA (2012)
25. Wang, Z., Sun, L., Zhu, W., Yang, S., Li, H., Wu, D.: Joint social and content recommendation for user-generated videos in online social network. *Multimedia, IEEE Transactions on* 15(3), 698–709 (2013)