

 Open access • Proceedings Article • DOI:10.1145/1996413.1996416

## **Boosting video popularity through recommendation systems** — [Source link](#)

[Renjie Zhou](#), [Samamon Khemmarat](#), [Lixin Gao](#), [Huiqiang Wang](#)

**Institutions:** [Harbin Engineering University](#), [University of Massachusetts Amherst](#)

**Published on:** 12 Jun 2011 - [Databases and Social Networks](#)

**Topics:** [Popularity](#) and [Recommender system](#)

Related papers:

- [The impact of YouTube recommendation system on video views](#)
- [I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system](#)
- [VideoTopic: Modeling User Interests for Content-Based Video Recommendation](#)
- [Online Video Recommendation in Sharing Community](#)
- [Overview of Content-Based Click-Through Rate Prediction Challenge for Video Recommendation](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/boosting-video-popularity-through-recommendation-systems-b6jeayf4fa>

# Boosting Video Popularity Through Recommendation Systems

Renjie Zhou<sup>†</sup>, Samamon Khemmarat<sup>‡</sup>, Lixin Gao<sup>‡</sup>, Huiqiang Wang<sup>†</sup>

<sup>†</sup> College of Computer Science and Technology  
Harbin Engineering University, Harbin, China  
{renjie\_zhou,wanghuiqiang}@hrbeu.edu.cn

<sup>‡</sup> Department of Electrical and Computer Engineering  
University of Massachusetts, Amherst, USA  
{khemmarat,lgao}@ecs.umass.edu

## ABSTRACT

While search engines are the major sources of content discovery on online content providers and e-commerce sites, their capability is limited since textual descriptions cannot fully describe the semantic of content such as videos. Recommendation systems are now widely used in online content providers and e-commerce sites and play an important role in discovering content. In this paper, we describe how one can boost the popularity of a video through the recommendation system in YouTube. We present a model that captures the view propagation between videos through the recommendation linkage and quantifies the influence that a video has on the popularity of another video. Furthermore, we identify that the similarity in titles and tags is an important factor in forming the recommendation linkage between videos. This suggests that one can manipulate the metadata of a video to boost its popularity.

## Categories and Subject Descriptors

H.4 [Information Systems]: Applications

## General Terms

Algorithms, Measurement, Human Factors

## Keywords

YouTube, Recommendation System, View Propagation, Video Popularity

## 1. INTRODUCTION

Online content providers and e-commerce sites typically provide search engines to help users discover items of their interests among the increasingly huge collection of items available on these sites. However, the capability of the search engines becomes limited when the textual descriptions of content cannot fully represent the semantic of the content. This is especially true for user generated video

sharing sites, such as YouTube, Vimeo, and Dailymotion, where textual descriptions of videos are provided by video uploaders and are sometimes incomplete and inconsistent. Recommendation systems are now widely used in online content providers and e-commerce sites. These recommendation systems take advantage of the hints from users' interactions with the sites to provide recommendation without solely depending on textual descriptions or content analysis. For example, the item-based collaborative filtering algorithm infers a relationship between items from the number of times the items are bought/viewed together by the same user, and thus it can provide recommendations even without textual information of the content. As a result, recommendation systems play a significant role in discovering content. For example, it has been shown in [11] that approximately 30% of the views on YouTube are from users clicking on related video recommendation lists.

In this paper, we focus on the role of the recommendation system in boosting the popularity of user generated videos on YouTube. For user generated video sharing sites, it is common that a recommendation list containing related videos is provided along with each video. While the first video view of a user is often through search engines, a homepage, or external links from other websites and emails, subsequent video views might come from the user clicking on a video on the related video lists. Considering the role of the recommendation system in "propagating" views, we investigate how the presence of the recommendation system impacts video popularity, measured by the number of video views. In particular, we explore how a view of one video influences the view of other videos through the related video linkage induced by the recommendation system and identify the underlying causes of strong influence between videos.

To this end, we propose a view propagation model to quantify the impact of the recommendation system. Based on the model, we study the view propagation on YouTube and show that it is possible to identify a subset of videos that are highly influential on the popularity of a video. Identifying the influence between videos is important for several reasons. First, the influence between videos provides a way to estimate the level of co-access between videos, which is useful for optimizing the architecture of online video sharing sites for better performance and efficiency in terms of data access. Second, the influence between videos provides an avenue for boosting video popularity. While search engine optimization techniques have been explored extensively, little has been done to investigate the potential of boosting video popularity through recommendation systems. We

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. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

*DBSocial '11*, June 12, Athens, Greece

Copyright 2011 ACM 978-1-4503-0650-8 ...\$10.00.

identify the underlying factors that induce the recommendation linkage between videos and factors that affect the influence between videos. With this knowledge, it is possible for a video uploader to manipulate the metadata of his/her video to exploit the recommendation system for boosting the video’s popularity. Our work is summarized as follows.

- We propose a model to analyze view propagation between videos. Based on the model, we quantify the influence between videos and demonstrate how the influence can be derived from the data statistics from YouTube. Our model can not only estimate the influence between videos, but also derive the number of views that come from the recommendation system, as well as the number of views that come from sources other than the recommendation system for each video.
- Our analysis on the view propagation on YouTube provides hints on how one could increase views via the help of the recommendation system. We quantify the level of contribution from referrer videos that are directly linked and indirectly linked to a video. Our results indicate that although direct referrer videos generally contribute most views to a video, indirect referrer videos are also significant sources of views. We find that the popularity of a referrer video (in terms of views coming from sources other than the recommendation system) is generally the most important in determining the number of views that are propagated from them.
- We identify that the similarity in titles and tags of videos is an important factor in forming the recommendation linkage between videos. From our experimental results, there is approximately 40% chance that the most similar video in terms of titles and tags is among the top referrer videos of a video. Since titles and tags of videos can be determined freely by video uploaders, strategies such as identifying the most popular videos with similar topic and using a title similar to the identified videos’ might help to attract more views.

The rest of this paper is organized as follows: In Section 2, we describe the view propagation model. Section 3 describes our dataset, and Section 4 explains our approach in computing the influence for YouTube videos. In Section 5, we analyze the factors that determine the number of views propagated between videos. Section 6 explores the roles of titles and tags of videos in creating referrer links. We discuss related work in Section 7 and conclude in Section 8.

## 2. VIEW PROPAGATION MODEL

Consider a typical system with an item-based recommendation system, in which each item is associated with a list of recommended items. When a user views an item, he/she may click and view another item in its recommendation list, and continue on. In this scenario, the first item contributes to the views of the subsequent items. Our goal is to study how one item’s popularity impacts other items’ popularity via the recommendation links. Usually, the probability that a user viewing item  $i$  will click on item  $j$  in its recommendation list, or the click through rate (CTR) from  $i$  to  $j$ , is not difficult to obtain. However, this probability is not sufficient for our goal as it only captures the relationship between an

item and the items in its recommendation list. Therefore, we propose the view propagation model, which can be used to derive the influence of one item to other items even though they are not directly linked.

Let us first define some terminology. A user who views item  $i$  may view item  $j$  in  $i$ ’s recommendation list. In this case,  $i$  is called a *referrer item* of  $j$ . A view of an item is categorized into two types. The first type is a view an item receives from its referrer item through a recommendation list, called a *propagated view*. The second type is a view an item receives from a source other than the recommendation lists, called an *injected view*.

Let  $P$  be an  $N \times N$  transition probability matrix where  $N$  is the number of items in a system, and  $P(i, j)$  is the CTR from  $i$  to  $j$ .  $P(i, j) = 0$  if  $j$  is not in  $i$ ’s recommendation list. The influence from  $i$  to  $j$ , denoted by  $F(i, j)$ , is defined as the probability that a user who views  $i$  will view  $j$  by traversing one or more recommendation lists. We note that  $P$  is different from  $F$  since  $P$  is the probability of viewing  $j$  from  $i$  directly by clicking on  $i$ ’s recommendation list.

We will now describe a view propagation process to derive  $F$ . The view propagation starts with an injected view of item  $i$ , then the view may be propagated to one of the recommended items of  $i$  or stop propagating according to the transition probability in  $P$ . This propagation continues according to the probability given by  $P$ .

With the described view propagation, given the injected views of items and the transition matrix  $P$ , we can compute the propagated views of items by iterating the view propagation. Let  $V_I$  be the row vector of injected views, where  $V_I(i)$  is the number of injected views of  $i$ , and  $V_P$  be the row vector of propagated views, where  $V_P(i)$  is the number of propagated views of  $i$ . Let  $V_P^{(k)}$  be the row vector of propagated views in iteration  $k$ . Then, we have  $V_P^{(1)} = V_I P$  and  $V_P^{(k)} = V_P^{(k-1)} P$ . The total propagated views are the sum of propagated views from every iteration. Thus, we have

$$V_P = \lim_{n \rightarrow \infty} \sum_{k=1}^n V_P^{(k)} = \lim_{n \rightarrow \infty} \sum_{k=1}^n V_I P^k. \quad (1)$$

If the maximum absolute row sum norm of  $P$  ( $\|P\|_\infty$ ) is less than one, the series converges [7]. This condition means that if for every item  $i$  the sum of the CTRs from  $i$  to other items is less than one, we would have

$$V_P = V_I((I - P)^{-1} - I), \quad (2)$$

where  $I$  is the identity matrix. The condition is usually true because it is not likely that a view of one item *always* leads to a view of another item. From Equation 2, the total propagated views of an item consist of a fraction of injected views from other items. The matrix  $((I - P)^{-1} - I)$  contains the ratios of injected views that propagate from one item to another item, which is the influence  $F$ . That is,  $F = ((I - P)^{-1} - I)$ .

In the following sections, using YouTube as a case study, we perform extensive analysis on injected views, influence, and propagated views between videos in order to understand the role of the recommendation system on the popularity of videos.

## 3. DATASET DESCRIPTION

YouTube video information can be crawled through YouTube data API and HTML scraping. For each video, we obtain

three pieces of information: metadata, related video list, and view statistics. First, the *metadata* contains the basic information such as title, tags, and total view count. Second, the *related video list* is a list of related videos recommended by YouTube. Third, the *view statistics* include the video’s view count series over time, its top ten view sources, and the date of the first referral and the number of views from each source (called the *refer views*). The view sources are classified into different categories. The category that we are interested in is *Related video*, for which the view statistics provide the video ID of a *referrer video*, which is a video that leads users to view the current video. We refer to the videos contained in the top ten view sources of a video as its *top referrer videos*.

Our dataset contains several snapshots of video information. Our crawling process was repeated approximately every two weeks, from November 2010 to March 2011. In each round, the crawling started from the same set of 1,502 seed videos. The information of the seed videos was collected, and their top referrer videos were extracted. Then, the top referrer videos were crawled. These steps were performed recursively in the manner of breadth-first search for six levels, resulting in approximately 120,000 videos being crawled in each round. We provide the overview of the videos in our dataset by showing the CCDF of the number of video views from one of the snapshots in Figure 1. As can be seen from the figure, the videos’ popularity ranges from one view to almost 380 million views, which allows us to study videos with a wide range of popularity.

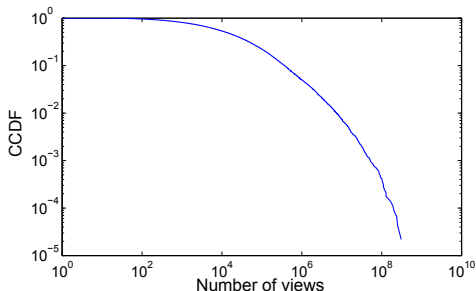


Figure 1: CCDF of video view count

From the crawled video information, the click through rate (CTR) from each top referrer video to a video can be computed. Let  $R_{ij}$  denote the number of refer views from top referrer video  $i$  to video  $j$ , and let  $V_i$  denote the total number of views of  $i$ . The CTR from  $i$  to  $j$  is given by  $R_{ij}/V_i$ . Similarly, the CTR from  $i$  to  $j$  between time  $t_1$  and  $t_2$  can be computed using the difference of the refer views and the difference of the total views in the two data snapshots taken at  $t_1$  and  $t_2$ . The CTRs are used later when we construct matrix  $P$  to analyze the view propagation.

## 4. OBTAINING INFLUENCE FOR VIDEOS

In this section, we explain how we obtain the influence between YouTube videos from YouTube data. From the propagation model, to compute the influence and propagated views from one video to another video, we need to know the injected views of each video (corresponding to vector  $V_I$ ) and the CTR between every pair of video (corresponding to matrix  $P$ ). However, from raw YouTube data, we only

have the total views of each video and the CTRs between videos and their top referrer videos. We first address how the dynamic of recommendation lists on YouTube affects our model. Then, we describe how to estimate the injected views from the data, followed by how to obtain influence and propagated views between each pair of videos.

### 4.1 Dynamic of YouTube Recommendation Lists

The YouTube recommendation system keeps updating its recommendation based on new information. As a result, the transition matrix  $P$  is changing over time. Consider a short period  $(t, t+1)$ , where the time granularity is small enough so that the transition matrix can be considered stable. Let  $P^{(t)}$  be the transition matrix for the period,  $\Delta V_P^{(t)}$  be the vector of propagated views gained during the period, and  $\Delta V_I^{(t)}$  be the vector of injected views during the period. From Equation 2, the vector of propagated views gained during the period is given by  $\Delta V_P^{(t)} = \Delta V_I^{(t)}(I - P^{(t)})^{-1} - \Delta V_I^{(t)}$ . Thus, the total propagated views can be computed by

$$V_P = \sum_{t=1}^m \Delta V_P^{(t)} = \sum_{t=1}^m \Delta V_I^{(t)}((I - P^{(t)})^{-1} - I), \quad (3)$$

where  $m$  is the total number of periods. However, our dataset does not have full information of injected views and the transition matrix. In addition, it is not practical to obtain the data for every short period. In the following, we show how we simplify the computation using approximation from available YouTube data.

### 4.2 Estimation of Injected Views

From our dataset, we have the total views of each video. Here we show how the injected views can be estimated given the total views. Let  $V$  be a row vector of the total views of videos, which is the summation of the total injected views and the total propagated views. From Equation 2, we can derive  $V_I = V(I - P) = V - VP$ . Notice that in this equation the propagated views is computed as  $VP$ . However, computing the propagated views this way does not allow one to identify the original source of each view (as opposed to  $V_I F$ ). Considering the change on YouTube, we then have  $V_I = V - \sum_{t=1}^m \Delta V^{(t)} P^{(t)}$ .

Let us now consider how to obtain  $P^{(t)}$  from YouTube data. From the view statistics of video  $j$ , we can compute the CTR  $P^{(t)}(i, j)$  from each top referrer video  $i$  to video  $j$  as described in Section 3. The CTRs between videos and their top referrer videos constitute part of the matrix  $P^{(t)}$ , which is referred to as  $P_{top}^{(t)}$ .

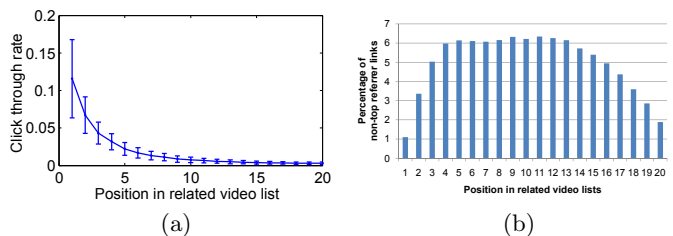


Figure 2: (a) CTR for each position. (b) Position of a video in related video lists of its non-top referrer videos.

To obtain the remaining entries of  $P^{(t)}$ , we need to identify the set of non-top referrer videos for each video and their CTRs to the video. The set of non-top referrer videos of video  $j$  can be obtained by identifying the videos that include video  $j$  in their related video lists. However, the exact CTR of a non-top referrer is not available from the data. We instead estimate  $P^{(t)}(i, j)$  for non-top referrers based on the position of  $j$  in  $i$ 's related video list, using the average CTR of the position. The average CTR and the standard deviation for each position is shown in Figure 2a. As can be seen from the figure, the CTR decreases quickly as the position becomes lower, and the variation of the CTR is quite low except for the first two positions. Since only approximately 4% of the videos are in the first two positions of related video lists of non-top referrer videos, as shown in Figure 2b, we believe that using the average CTR as  $P(i, j)$  is reasonable. We denote the partial matrix of  $P^{(t)}$  containing CTRs of non-top referrer videos as  $P_{nontop}^{(t)}$ .

Having obtained  $P_{top}^{(t)}$  and  $P_{nontop}^{(t)}$ , a good estimation of  $P^{(t)}$  can be obtained from  $P_{top}^{(t)} + P_{nontop}^{(t)}$ , which gives  $V_I \approx V - \sum_{t=1}^m \Delta V^{(t)}(P_{top}^{(t)} + P_{nontop}^{(t)})$ . When  $P^{(t)}$  is not available for every period,  $V_I$  can be estimated as

$$V_I \approx V - V(P_{top} + P_{nontop}), \quad (4)$$

where  $P_{top}$  contains the CTRs of the current top referrer videos, and  $P_{nontop}$  contains the CTRs of non-top referrer videos from a certain period, which in our case is the latest period. Following are our arguments.

For the term  $\sum_{t=1}^m \Delta V^{(t)} P_{top}^{(t)}$ , we argue that  $P_{top}$  is a good estimation of  $P_{top}^{(t)}$  for every  $t$  since the top referrer videos are stable. The stability of the set of top referrer videos indicates that the CTRs of top referrer videos are relatively stable. Otherwise, the set of top referrer videos would be changing. This is supported by a measurement result from our dataset which shows that 96% of the video/top referrer video pairs extracted from two data snapshots 45 days apart remain the same.  $P_{top}(i, j)$  can be obtained from the view statistics as explained in Section 3.

The term  $\sum_{t=1}^m \Delta V^{(t)} P_{nontop}^{(t)}$  is the computation of the total propagated views coming from non-top referrer videos. We argue that  $\sum_{t=1}^m \Delta V^{(t)} P_{nontop}^{(t)}$  can be estimated by  $V P_{nontop}$ . First, we show that the percentage of propagated views contributed by non-top referrer videos into the system are quite stable. For each period, the number of views propagated out from each video  $i$  is  $\Delta V^{(t)}(i) \sum_j (P_{top}^{(t)}(i, j) + P_{nontop}^{(t)}(i, j))$ . It is reasonable to assume that the sum of outgoing CTRs for each video  $i$ ,  $\sum_j (P_{top}^{(t)}(i, j) + P_{nontop}^{(t)}(i, j))$ , is stable over time. Therefore, the total views propagated out from video  $i$  in the long term can be approximated by  $V(i) \sum_j (P_{top}(i, j) + P_{nontop}(i, j))$ . As discussed previously, the top referrers and their CTRs are relatively constant. Thus, the percentage of propagated views contributed by all non-top referrer videos into the system are also quite stable over time. To verify this, we measure the percentage of views from non-top referrers from the last snapshot and for the whole period from our data, which gives 4.3% and 3.9%, respectively. Next, we verify that  $V P_{nontop}$  is also a good estimation of  $\sum_{t=1}^m \Delta V^{(t)} P_{nontop}^{(t)}$  for each individual video. We measure the difference between percentage of propagated views from non-top referrer videos for each

video of the last snapshot and the whole period. The average difference is 1.2% with the standard error of 1.6%. From these observations, we believe that  $\sum_{t=1}^m \Delta V^{(t)} P_{nontop}^{(t)}$  can be well estimated by  $V P_{nontop}$ . With this approximation, we can now compute the injected views as shown in Equation 4.

### 4.3 Estimation of Influence Matrix and Propagated Views Between Videos

To estimate the influence matrix, let us reconsider Equation 3 with  $P_{top}^{(t)}$  and  $P_{nontop}^{(t)}$ :  $V_P = V_I F = \sum_{t=1}^m \Delta V_I^{(t)} [P_{top}^{(t)} + P_{nontop}^{(t)} + (P_{top}^{(t)} + P_{nontop}^{(t)})^2 + \dots]$ . In the following paragraphs, we provide our arguments that  $F$  can be well estimated by  $P_{top} + P_{top}^2$ .

First, the terms with the power of three are very small since each entry in  $P_{top}^{(t)}$  and  $P_{nontop}^{(t)}$  is less than one and most of them are less than 0.2, as shown by the CTRs in Figure 2a. Therefore, they can be ignored. Next, we argue that the non-top referrer links are less significant compared to the top referrer links in terms of their contribution to the total propagated views since the percentage of the total propagated views from the non-top referrer videos is approximately 14% of the total propagated views. Furthermore, our measurement shows that the non-top referrer video links are often nonpersistent. When we compare  $P_{nontop}^{(t)}$  from two snapshots 45 days apart, only 46% of the non-top referrer links remain the same. This nonpersistence means that in the long term the strength of the non-top referrer links is relatively small compared to the top referrer links. Therefore, any term containing  $P_{nontop}^{(t)}$  is much less significant for estimating  $F$ , relative to the terms containing  $P_{top}$ .

From these arguments,  $F$  can then be approximated from  $P_{top} + P_{top}^2$ , which would be able to capture most of the significant entries in  $F$ . Further, the approximation provides a way to prune to a very dense matrix  $F$  into a sparser version that includes only significant entries. Now that we are able to compute both the injected views and the influence matrix, the propagated views between each pair of videos can then be computed, and we have the three elements for our analysis in the next section.

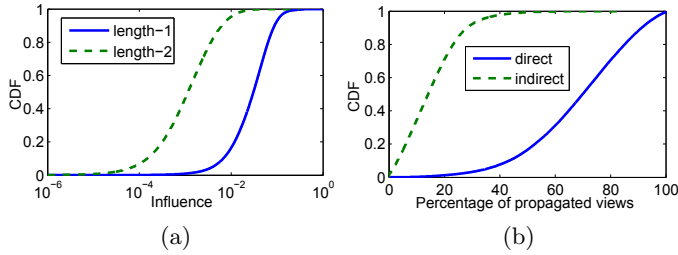
## 5. FACTORS IN VIEW PROPAGATION

In this section, the roles of injected views and influence in view propagation are investigated. Better understanding of how injected views and influence play their parts in view propagation could give hints on determining which videos would be the most beneficial referrers.

### 5.1 How distance between videos impacts view propagation

By estimating influence as  $P_{top} + P_{top}^2$ , the influence can be decomposed into the influence from paths of length 1 and the influence from paths of length 2. To understand the difference between influence from the different path lengths, we compare  $P_{top}$  and  $P_{top}^2$ , the influence from paths of length 1 and length 2, respectively. Figure 3a shows the distribution of values in  $P_{top}$  and  $P_{top}^2$ . As shown in the figure, the influence from paths of length 2 is significantly lower. Even though there may be multiple paths of length 2 between a pair of videos, the aggregated influence from all of them is still almost always smaller than the influence from a path of length 1. The average influence from a path of length 1 is

0.045, while the average influence from paths of length 2 is 0.0023, 20 times smaller than the influence from a path of length 1.



**Figure 3: (a) Influence from paths of length 1 and length 2. (b) Propagated views from direct and indirect top referrers.**

Next, we quantify the importance of the direct and indirect top referrer videos. The direct and indirect top referrer videos of a video are videos that connect to the video via a path of length 1 and 2, respectively. For each video, we compute the percentage of propagated views contributed from its direct and indirect top referrers. The distribution of the percentage of propagated views from direct and indirect top referrers is shown in Figure 3b. As shown in the figure, on average 68% of propagated views are from direct top referrers, and nearly 15% of propagated views are from indirect top referrers. While direct referrer videos contribute the majority of the propagated views, it should be noted that the contribution from indirect top referrer videos is nonnegligible, even though the previous results show that influence is greatly decreased with the distance. In combination, the direct and indirect top referrers account for more than 80% of the propagated views. This result suggests that to increase views through recommendation links, being on the related video lists of beneficial videos would be the most effective way. Also, when direct attachment is difficult, indirect attachment, that is, to be on the related lists of related videos, could be effective.

## 5.2 Impacts of influence versus injected views

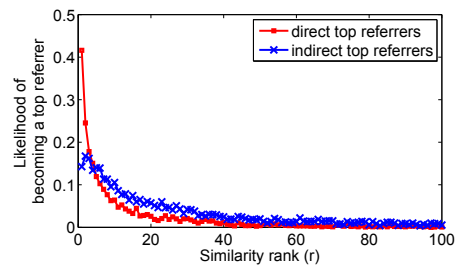
We investigate whether a video usually receives more views from referrer videos with large number of injected views or from referrer videos that have high influence on the video. This should provide some hints to select beneficial videos to attach to. For each video, we compute the ranking correlation between the injected views of its referrer videos and the propagated views from the referrers and the ranking correlation between the influence and the propagated views. Our investigation shows that the number of propagated views is better correlated with the number of injected views, with an average correlation of 0.51, while the correlation between the propagated views and the influence is 0.29. The result indicates that a video tends to receive many propagated views from referrer videos that have large number of injected views, rather than referrer videos that have high influence. Therefore, we conclude that referrer videos with large number of injected views are more preferable for attracting more propagated views. With these guidelines on which videos would be the most beneficial referrers, in the next section we investigate how a video can create recommendation linkage with these referrers.

## 6. POTENTIAL OF MANIPULATING META-DATA

Our question in this section is how one can make the beneficial videos become referrer videos. There are several underlying factors that induce the referral linkage. Identifying these factors is useful in order to leverage the recommendation system for increasing video views. In this section, we study the relationship between a video and its top referrer videos to identify some of the factors that directly or indirectly induce their relationship.

We initially observed that it is common for videos and their top referrer videos to share similar titles, tags, or uploaders. To investigate how much the similarity of titles, tags, and uploaders can define the relationship between a video and its top referrer videos, we define a metric to measure the similarity between videos based on their titles, tags, and uploaders as  $S_\alpha(i, j) = \alpha S_k(i, j) + (1 - \alpha) S_u(i, j)$ .  $S_k(i, j)$  is the cosine similarity between the TF-IDF vectors of video  $i$  and  $j$  created from keywords in the video’s titles and tags.  $S_u(i, j)$  is equal to one if  $i$  and  $j$  have the same uploader; otherwise,  $S_u(i, j)$  is zero. The  $\alpha$  is a parameter for controlling the weights of title/tag similarity and uploader similarity.

We first examine whether a video is more similar to its top referrer videos compared to other videos. We will refer to a pair of a video and one of its top referrer videos as a *referrer video pair*. Our sample video set contains 1,500 videos, resulting in 5,875 direct referrer video pairs and 10,743 indirect referrer video pairs. We create a pool of approximately 120,000 videos from our full dataset to compare the top referrers with other videos. For each video in the sample set, we rank all videos in the pool based on the defined similarity  $S_\alpha$  and obtain the ranks of its direct and indirect top referrer videos. The value of  $\alpha$  used for this measurement is 0.7. We find that 90% of the direct top referrer videos are ranked higher than 100, i.e., among the top 0.1% most similar videos. The indirect top referrer videos’ ranks are generally lower than the direct top referrer videos’ ranks, but are still high. 80% of the indirect top referrers are ranked higher than 178. These results suggest that titles, tags, and uploaders of the referrer videos are usually more similar to the video’s, compared to those of other videos.



**Figure 4: Likelihood of becoming top referrers**

Additionally, we compute the likelihood of a video with similarity rank  $n$  becoming a top referrer video. Figure 4 shows the percentage of videos with similarity rank  $n$  that becomes a top referrer video. From the figure, 41% of the videos with similarity rank 1 are direct top referrers, and 14% of them are indirect top referrers. In combination, 55% of the time, the most similar video is either a direct or an indirect top referrer. This likelihood decreases for lower ranks,

and after the 20<sup>th</sup> rank, the decrease becomes gradual. Let us consider the fact that the top referrers of a video are determined from the number of views they refer to the video. This means that the popularity of the referrer videos is one of the key factors in determining whether they are the top referrers or not. Despite this fact, we still observe the influence of title/tag similarity on the top referrer relationship; this emphasizes the importance of having high title/tag similarity with the top referrers.

From the results shown, we conclude that the similarity in titles, tags, and uploaders of videos is a significant factor that influences the top referrer relationship between videos. Our hypothesis is that for videos (e.g. new videos) with no or little co-view information, title/tag/uploader similarity is the main information used for determining recommendation lists. These bootstrapping lists encourage co-viewing between videos with similar titles and tags, resulting in videos with higher similarity being more likely to be the top referrer videos. Therefore, it is possible to manipulate titles and tags of videos to increase video popularity.

## 7. RELATED WORK

The research community has been interested in how recommendation systems influence online content popularity. In [1], Chen et al. studied the impact of recommendation on sales using book sales data from Amazon.com, demonstrating the positive correlation between the number of times the books are recommended and their sales rank. The impact of recommendation systems in e-commerce sales has also been studied through analytical models and simulation in [5] and [6], which concluded that collaborative filtering recommenders tend to decrease sales diversity. Our work in [11] is a measurement-based study on the impact of the recommendation system on YouTube video views. In a larger context, several models have been proposed to understand and predict online content popularity. In [3], Crane et al. investigated the response of YouTube video popularity to social events. Models for predicting popularity of online content including YouTube videos and Digg stories are proposed in [9], [10], [2], and [8].

In this work, we focus on studying the impact of a recommendation system on video popularity in a user-generated video site. In contrast with the previous works that investigate the impact of recommendation systems, our investigation is based on a view propagation model formulated from user browsing behavior. The model allows us to study how videos influence each other's popularity through the linkage induced by the recommendation system and how one can exploit the influence to boost the popularity of a video.

## 8. CONCLUSION

In this paper, we study the role of the recommendation system on the popularity of videos in YouTube and then explore the possibility of leveraging the recommendation system to boost video views. We propose a view propagation model which models how views are propagated between videos through video linkage induced by the recommendation system. The model allows us to quantify the influence between videos even though they are not directly linked to each other. Our analysis shows that the majority of propagated views of a video are driven by a few referrer videos that are within two hops away, i.e., direct and indirect top

referrer videos. In addition, the popularity of a referrer video plays a larger role in determining the number of propagated views in comparison with the click through rate. We further analyze important characteristics of the top referrers to infer factors that influence recommendation linkage formation. We find that similarity of videos' titles and tags has significant influence in the formation of referral relationship between videos. These results indicate the potential of manipulating videos' metadata to increase video views through the recommendation system.

## 9. ACKNOWLEDGMENTS

The authors are grateful to the anonymous reviewers for their valuable comments and suggestions. This work is partially supported by U.S. NSF grant CNS-0626618. Renjie Zhou was a visiting student at University of Massachusetts, Amherst, when this work was performed.

## 10. REFERENCES

- [1] P. Chen, S. Wu, and J. Yoon. The impact of online recommendations and consumer feedback on sales. In *Proceedings of the International Conference on Information Systems*, volume 7, pages 1–724, 2004.
- [2] X. Cheng, C. Dale, and J. Liu. Statistics and social network of youtube videos. In *International Workshop on Quality of Service (IWQoS'08)*, pages 229–238. IEEE, June 2008.
- [3] R. Crane and D. Sornette. Viral, quality, and junk videos on youtube: Separating content from noise in an information-rich environment. In *Proc. of AAAI symposium on Social Information Processing, Menlo Park, CA*, 2008.
- [4] J. Davidson, B. Liebald, J. Liu, P. Nandy, T. Van Vleet, U. Gargi, S. Gupta, Y. He, M. Lambert, B. Livingston, et al. The YouTube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 293–296. ACM, 2010.
- [5] D. Fleder and K. Hosanagar. Recommender systems and their impact on sales diversity. In *Proceedings of the 8th ACM conference on Electronic commerce*, page 199. ACM, 2007.
- [6] D. Fleder and K. Hosanagar. Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management Science*, 55(5):697–712, 2009.
- [7] R. Horn and C. Johnson. *Matrix analysis*. Cambridge university press, 2005.
- [8] K. Lerman and A. Galstyan. Analysis of social voting patterns on digg. In *Proceedings of the first workshop on Online social networks*, pages 7–12. ACM, 2008.
- [9] K. Lerman and T. Hogg. Using a model of social dynamics to predict popularity of news. In *Proceedings of the 19th international conference on World wide web*, pages 621–630. ACM, 2010.
- [10] G. Szabo and B. Huberman. Predicting the popularity of online content. *Communications of the ACM*, 53(8):80–88, 2010.
- [11] R. Zhou, S. Khemmarat, and L. Gao. The impact of YouTube recommendation system on video views. In *Proceedings of the 10th annual conference on Internet measurement*, pages 404–410. ACM, 2010.