

# Viral Actions: Predicting Video View Counts Using Synchronous Sharing Behaviors

**David A. Shamma**  
Yahoo! Research  
Santa Clara, CA, USA  
aymans@acm.org

**Jude Yew**  
University of Michigan  
Ann Arbor, MI, USA  
jyew@umich.edu

**Lyndon Kennedy**  
Yahoo! Labs  
Santa Clara, CA, USA  
lyndonk@yahoo-inc.com

**Elizabeth F. Churchill**  
Yahoo! Research  
Santa Clara, CA, USA  
churchill@acm.org

## Abstract

In this article, we present a method for predicting the view count of a YouTube video using a small feature set collected from a synchronous sharing tool. We hypothesize that videos which have a high YouTube view count will exhibit a unique sharing pattern when shared in synchronous environments. Using a one-day sample of 2,188 dyadic sessions from the Yahoo! Zync synchronous sharing tool, we demonstrate how to predict the video's view count on YouTube, specifically if a video has over 10 million views. The prediction model is 95.8% accurate and done with a relatively small training set; only 15% of the videos had more than one session viewing; in effect, the classifier had a precision of 76.4% and a recall of 81%. We describe a prediction model that relies on using implicit social shared viewing behavior such as how many times a video was paused, rewind, or fast-forwarded as well as the duration of the session. Finally, we present some new directions for future virality research and for the design of future social media tools.

Increasingly, more and more media is watched and shared online. Websites such as YouTube, enables people to not just share videos but socially interact with each other as well. More recently, newer social media systems allow users to synchronously interact with each other and share videos simultaneously. These real-time interactions leave behind large amounts of contextual usage data that, we believe, are reflective of the deeper and more connected social interaction that accompanies synchronous content sharing. In this paper, we present a method of utilizing usage data from synchronously sharing videos to make predictions about the popularity of a particular video. In particular, we use play/pause behavior and chat volume pulled from a real-time video sharing environment, Zync (a plug-in for the Yahoo! Instant messaging (IM) client that allows participants to view and interact with a video simultaneously during a chat session). We argue that the usage data from synchronous video sharing tools provides robust data on which to detect how users are consuming and experiencing a video. By extension, we can predict a video's popularity based on how it has been shared in a handful of sessions. To do this, we

trained a Naive Bayes classifier, informed by synchronous sharing features, to predict whether a video is able to garner 10 million views on its hosting site. Our goal is to eventually predict a video's viral potential based on how its being shared.

The ability to socially share videos online has enabled select videos to gain a viewership of thousands in a very short period of time. Often, but not always, these videos take on a viral nature and gain tens of millions of views, while other videos only receive a fraction of the attention and viewing. These popular, viral videos also benefit from *rich get richer* dynamic where the more popular they become, the more views they are likely to attract. Viral videos attract not only a disproportionate amount of attention, they also consume greater amounts of resource and bandwidth as well. Thus, it would be helpful to be able to predict and identify which videos are most likely to go viral for recommendation, monetization, as well as, systems performance.

Historically, the methods used to classify the genre of web videos have so far focused on the machine recognition of content within the videos. While this content-based approach has enabled classification of a wide variety and large quantities of online video, it also presents several problems. On its own, it can be technically complex and prior work have mainly focused on classification tasks where the genres are well-defined and commonly recognized (Roach and Mason 2002). More recently the classification of media content has started to take into account the contextual information, such as 5-star ratings (Herlocker, Konstan, and Riedl 2000) and the link structures (Mahajan and Slaney 2010) of shared content. We approach the problem differently. Rather than paying attention to the content of the video, its metadata, or its position in a network, we focus mainly on identifying particular interaction patterns from synchronous video sharing. Simply put, we assume that videos which are likely to go viral will be interacted with differently than other videos. How a video is used, interacted with, and commented is often indicative of the nature of its content.

## Related Work

Media is often experienced socially and a large part of that experience involves frequent commentary, backchannel conversations and starts/stops to review key moments in the media content. The ability to share videos in real-time while in

Video Id	Session Time	Loads	Play	Pause	Chat Lines	Scrubs	YouTube Views
11	10.32m	1	0	0	0	0	1,740,179
12	11.03m	1	0	0	56	4	25,315
13	11.52m	1	0	0	60	1	21,574,284
14	13.08m	4	4	3	21	7	22,951

Table 1: The aggregate feature vectors by video. Initial playback is automatic and counted in the *Loads* count. In this illustration, video #14 was viewed in 4 sub-sessions and its vector represents the average activity across those sub-sessions. The remaining videos were only viewed in one session.

an Instant Messaging (IM) session is an attempt to replicate the social experience of watching videos in a virtual environment. Zync is a plug-in for the Y!Messenger that allows IM participants to share and watch videos together in real-time. These video sharing sessions leave behind *digital traces* in the form of server logs and metadata (Wesler et al. 2008). These digital traces, we argue, can be used to predict how popular a video is going to be, in terms of the number of views it will garner.

The use of digital traces has been utilized in similar work by Crane and Sornette (Crane and Sornette 2008). In their work, they identified “signatures” in the viewership metadata of YouTube videos in order to identify “quality” content, or videos that attract attention quickly and only slowly lose their appeal over time because of their high quality. Likewise, Szabo and Huberman (Szabo and Huberman 2010) analyzed the view counts on YouTube and voting histories of stories on Digg.com in order to forecast a video’s/story’s popularity thirty days ahead. These researchers found that using the “attention” metadata paid to online content they were able to forecast the popularity of the content. However, their paper notes that they needed to track the “attention” data for YouTube videos for much longer (10 days) than stories on Digg (2 hours) in order to make their popularity forecasts. Both these projects illustrated here highlight the viability of paying attention to the social and contextual activity that surrounds the sharing of online content to make predictions about their “quality” and “popularity”.

Unlike these two related projects, our method of predicting the viewership of a shared video is based on a different form of “digital trace”. We argue that the implicit social sharing activity that occurs while sharing a video in a real-time IM conversation would result in more accurate predictions of a video’s potential viewership. Implicit social sharing activity here refers to the number of times a video was paused, rewound, or fast-forwarded as well as the duration of the IM session while sharing a video (Yew and Shamma 2011; Yew, Shamma, and Churchill 2011). We believe that implicit social sharing activity is indicative of deeper and more connected sharing constructs, and hence better fidelity data to predict how much viewership a particular video is likely to attract. How a video is interacted with and shared between users is often indicative of how popular it is likely to be in the future. For instance, videos that have great appeal and potential to be popular will mostly likely be interacted with more and generate more conversation than others. Taken in aggregate across all users, patterns and “sig-

natures” (Crane and Sornette 2008) of interactions found in the implicit social sharing data can point to how popular and even viral a video is likely to be.

Viral videos are those that have gained outsized prominence and viewership as a result of an ‘epidemic-like’ social transmission. Characteristically these videos tend to be short, humorous, and produced by amateurs. A recent example is the *Double Rainbow*<sup>1</sup> video posted by YouTube user HungryBear9562. In the self shot video, he declares his awe and wonder at the sight of a rare “double rainbow” he sighted in Yosemite National Park. The *Double Rainbow* video is a nice example of how the patterns of interpersonal transmission and social interaction that surrounds a video, helped launched it from relative obscurity to Internet prominence. In this paper, we argue that the usage data that surrounds such viral videos can be used to predict the popularity of the video. Here we capitalize on the “wisdom of the masses” (Surowiecki 2005) by identifying patterns in the metadata to make predictions about the future popularity of that content (Segaran 2007).

The main question that this study asks; Is there a correlation between watching behavior and the popular spread of a video? If so, can we start to build a predictive model of virality that is based on how people share and manipulate media? This paper aims at identifying this connection as guided by the following research question: Can we predict the view count of a video based on the patterns found in the real-time social sharing of videos?

## Study

We intend to investigate how people share and manipulate videos in order to predict how many views the source video had on its original website. In particular, we wish to explore this in a synchronous and conversational context. For this, the metadata found on sites like YouTube is insufficient. Rather, synchronous activity derived from livestream video sharing environments like JustinTV<sup>2</sup> and video-on-demand chat systems like the Yahoo! Zync<sup>3</sup> project.

## Dataset and Feature Identification

We acquired a 24-hour sample of the Zync event log for Christmas Day 2009. Zync allows two people to watch a

<sup>1</sup>Double Rainbow <http://www.youtube.com/watch?v=OQSNhk5ICTI> (Accessed 9/2010)

<sup>2</sup>JustinTV <http://justintv.com> (Accessed 2/2011)

<sup>3</sup>Yahoo! Zync <http://sandbox.yahoo.com/Zync> (Accessed 2/2011)

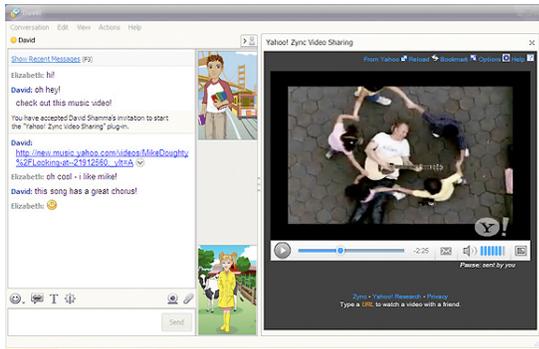


Figure 1: The Zync plugin allows two Instant Messenger users to share a video in sync while they chat. Playback begins automatically; the users share control over the video.

video together in an instant message session; both participants share playback and control of the video and the video stays in sync across both participants IM windows, see Figure 1. The dataset provides a list of watched videos from YouTube as well as synchronous activity from the shared control of the video. These features are: anonymous user id hashes, session start/stop events, the session duration, the number of play commands, the number of pause commands, the number of scrubs (fast forwards or rewinds), and the number of chat lines typed as a character and word count. For the chat lines, the dataset contained no actual text content, only the aggregate count of characters, words and lines. The only textual content that is collected is video URLs and emoticons. Each activity collected is a row in the dataset and is associated with the time of the event and the playback time on the video (Liu et al. 2007).

In total, the dataset contained 2,188 dyadic sessions. Across these sessions there were a sum total of 9,364 unique YouTube videos. Of these sessions, several were discarded as they were missing the segmentation (start/stop) events, was lacking a participant, etc. Of the valid sessions, we extracted the YouTube video identifiers and retrieved the related metadata associated with each video using YouTube’s Application Programming Interface<sup>4</sup>. Some videos were no longer available on YouTube due to copyright infringement or owner deletion; these videos and their respective sessions were also discarded. The final test sample contained 1,580 videos with valid YouTube metadata and valid session data.

The data collected from YouTube consisted of a video identifier, the video’s title, its published date, its description, the genre category the uploader used, the tags the video was labelled with, the video’s duration and the 5-star rating score it attained. Of these data, we only use the video’s YouTube view count. For this article, the view count will be the predictive variable for the classifier and allows us to investigate if there’s a match between YouTube view count and the synchronous social session actions. Similar prediction has been demonstrated in the past to determine the video’s content

<sup>4</sup>YouTube API <http://code.google.com/apis/youtube/overview.html> (Accessed 2/2011)

category (Yew, Shamma, and Churchill 2011); we will investigate the ability to predict the video’s popularity.

As mentioned earlier, each single event from every session is a row in the dataset. This data needs to be aggregated into a feature vector for training a classifier. To do this, every session was divided into segments (sub-sessions) where a single video was viewed. This was necessary as many sessions contained multiple videos. The sub-sessions were then grouped by their representative video, mixing all the sessions which watched the same video. Finally, each event type and the overall sub-session durations were averaged into a single feature vector. Lastly, we assign a label indicating if the YouTube view count is over 10 million views, see Table 1 for some sample feature sets. In this aggregate, we find median number of watches a video had was 2 ( $IQR = 5$ ); in fact, only 15% of the videos in the dataset received more than one session viewing. The median number of YouTube views in the dataset was 264,400 ( $\mu = 2956000$ ,  $\sigma = 11088549$ , and  $IQR = 1674540$ ). In the data, there is no correlation between the number of times a video was watched in Zync and that video’s view count on YouTube ( $p = 0.4232$ ). In fact, as illustrated in Table 1, the viral video with over 20 million views was only watched in one session. Additionally, while we are not accounting for the video’s playback time, there was no correlation between the Zync session time and the video’s playback length ( $p = 0.3378$ ).

## Prediction

As 1,580 feature vectors is a relatively small sample, we begin by transforming some of the interval, continuous features into categorical sets as described by Yew and Shamma’s (2011) previous YouTube category study. A Naive Bayes Classifier predicts if a video has over 10 million views with an accuracy of 96.6% using an 80% training set. Smaller training sample sizes offered similar performance (95.6% using a 60% training set). By comparison, an educated random guess based on the statistical distribution of videos with over 10 million and under 10 million views is 88.3% accurate. Likewise, simply guessing the video has less than 10 million views would predict at 93.7% accuracy. Table 2 shows a complete breakdown of predictions, methods, and training sizes. In effect, the classifier had a precision of 76.4% and a recall of 81%. By comparison, random prediction precision and recall scored 4% and 4.1% respectively.

It is important to note the training features result from observed, implicit synchronous sharing behaviors and not explicit social annotations (like ratings data). We hypothesized these implicit sharing features to be equally predictive as the explicit features that Yew and Shamma (2011) defined.

## Discussion

In the model and results we have addressed our research question: we can predict the view count of a video based on how it is viewed in a rich, shared, synchronous environment. In total, 100 of the 1580 had over 10 million views. The Naive Bayes classifier correctly identified 81 of these popular videos. By comparison, 6 of the 100 popular videos

Method	Training Sample	Accuracy	$F_1$ score
Guessing	All Yes	6.3%	0.119
	Random	88.3%	0.041
	All No	93.7%	NaN*
Naive Bayes	25%	89.2%	0.345
	50%	95.5%	0.594
	60%	95.6%	0.659
	70%	95.8%	0.778
	80%	96.6%	0.786

Table 2: Random and Naive Bayes Prediction Accuracies. Random guess is calculated by using the distribution bias (6.3% of guessing Yes). The  $F_1$  score illustrates the overall performance accounting for both Precision and Recall. The Naive Bayes predictions use crossfolded verification.

\*NaN means Not a Number; the  $F_1$  score results in a divide by zero due to this method’s poor performance.

were correctly identified by randomly guessing “yes” based on the dataset’s distribution. However, the high success rate of guessing “no” (as seen in Table 2) does illustrate a bias in the dataset. There are far more videos that have less than 10 million views and thus a higher prediction accuracy. It is important to note that our classifier produces a larger increase over a fair random prediction. We believe the session’s duration to be the dominate feature in the predictive model as the correlation between Zync session duration and YouTube view count had the highest correlation ( $p < 0.12$ ). While not quite significant, the average session duration in the feature vector is completely independent of the view count. Furthermore, there is no significant or near significant correlation between the session duration and the video’s playback time. Similarly, no significant correlations were observed within the other meta-data from YouTube (ratings and playback time) and the YouTube view count. As the Zync dataset of 1,580 is a rather small sample of videos, looking for some deeper alignment between session duration and YouTube view count is needed. Furthermore, we predicted *Yes* and *No* to the question “Does this video have over 10 million views?” With a larger dataset, we expect to be able to predict the actual view count.

Additionally, the aggregate grouping of the feature vectors by video came at a cost to the session grouping. In a single IM session, several videos are shared. Isolating the videos from their sessions discarded these collections which could be indicative of sets or co-occurrence of high view count videos.

## Future Work

We have demonstrated the possibility to a classifier, based on social synchronous sharing patterns, to predict if a video has a high view count. Our goal in this research is to predict if a video will go viral based on how it is shared within a conversation. The successful predictions in our classifier are based on most videos (85%) viewed once in only one session. The next step in this work is to collect various data samples over

time and investigate how a video’s implicit sharing behaviors change as it becomes viral. In effect, this is somewhat of a fishing exercise over time; we need to collect data on videos as they turn viral to train a classifier on how to predict them. We expect the temporal changes between the feature vectors (the deltas and rate of change across our video feature vectors) to enable accurate viral predictions for recommendations. Additionally, when socially filtered, unique viral patterns found in some social groups and networks could bring socially targeted recommendations and content promotion.

## Acknowledgments

The authors would like to thank Sarah Vieweg, Matt Cooper, and Bryan Pardo.

## References

- Crane, R., and Sornette, D. 2008. 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*.
- Herlocker, J.; Konstan, J.; and Riedl, J. 2000. Explaining collaborative filtering recommendations. In *ACM 2000 Conference on Computer Supported Cooperative Work*, 241–250. Association of Computing Machinery.
- Liu, Y.; Shamma, D. A.; Shafton, P.; and Yang, J. 2007. Zync: the design of sychronized video sharing. In *DUX 2007: Proceeding of the 3rd conference on Designing for User Experience*.
- Mahajan, D., and Slaney, M. 2010. Image classification using the web graph. In *Proceedings of the International Conference on Multi-Media*. ACM.
- Roach, M., and Mason, J. 2002. Recent trends in video analysis: A taxonomy of video classification problems. In *In Proceedings of the International Conference on Internet and Multimedia Systems and Applications, IASTED*, 348–354.
- Segaran, T. 2007. *Programming Collective Intelligence: Building Smart Web 2.0 Applications*. O’Reilly Media, 1 edition.
- Surowiecki, J. 2005. *The Wisdom of Crowds*. Anchor.
- Szabo, G., and Huberman, B. A. 2010. Predicting the popularity of online content. *Commun. ACM* 53(8):80–88.
- Wesler, H. T.; Smith, M.; Fisher, D.; and Gleave, E. 2008. Distilling digital traces: Computational social science approaches to studying the internet. In Fielding, N.; Lee, R. M.; and Blank, G., eds., *The Sage handbook of online research methods*. London: Sage. 116 – 140.
- Yew, J., and Shamma, D. A. 2011. Know your data: Understanding implicit usage versus explicit action in video content classification. In *IS&T/SPIE Electronic Imaging*.
- Yew, J.; Shamma, D. A.; and Churchill, E. F. 2011. Knowing funny: Genre perception and categorization in social video sharing. In *CHI 2011: Proceedings of the SIGCHI conference on Human factors in computing systems*. Vancouver, Canada: ACM. Forthcoming.