Riding the "Hits" Wave: Informational Cascades in Viewership of Online Videos

Clarice Sim W. Wayne Fu

Wee Kim Wee School of Communication and Information

Nanyang Technological University

Singapore

Abstract

The exploding spread of online peer-posted videos such as YouTube is phenomenal. The sharing and viewing of such web content are hypothesized to be influenced by the process of informational cascades, which commonly characterizes the choices of consumption by sequential patrons among (content) products of quality uncertainty. We scheme an empirical model to test the informational cascade effect on the accumulation of "views" or "hits" by videos. An entire set of 1,304 videos, uploaded to *Internet Archive*'s moving images section, during a 4-week period, are tracked for their individual daily hit captures over 33 days since launch. The views distribution is vastly skewed across the videos observed. The estimation reveals that the larger cumulative viewership a video has gained at a given time, the more views it will attract next. Further, this cascading effect for videos with thumbnails (i.e., a series of pictorial snapshots of the video) available aside is significantly lessened vis-à-vis that for those without. This finding espouses the postulation that the rise of information about quality and thus informational cascades responds to the role of quality uncertainty. Surprisingly, the presence of thumbnails marginally intensifies the cascading effect.

Riding the "Hits" wave: Informational Cascade in Viewership of Online Videos

The impact of online videos has been phenomenal to say the least. Videos on Youtube have shot unknowns to fame[1] and generated buzz for viral marketers[2], and this is only "the beginning of an Internet video revolution" (La Monica, 2006). Indeed, the popularity of online videos is already immense, with YouTube alone serving more than 100 million videos a day ("Youtube hits 100m videos", 2006). By another report, 57% of adults in the US reported watching or downloading online videos, while 19% of them engaged in this activity every day (Madden, 2007).

Yet, although online videos have gained tremendous prominence among Internet users in recent years, studies on the characteristics of Youtube videos and usage patterns have been mostly conducted by computer science scholars (Gill, Arlitt, Li, & Mahanti, 2007; Xu, Dale, & Liu, 2007), leaving the socio-communicative aspect behind the dynamics of viewership unexplored. How do online videos garner more views, and how do the view counts evolve over time? What consumer or audience behavioral theories can explain this evolution? This study grapples with these questions and empirically tests them using a sample of videos from an online video website.

One possible influence on the sharing and viewing of online videos could be through the process of informational cascades. The presence of informational cascades is characterized by convergent behavior, where sequential patrons defer to the choices made by preceding others. The information from preceding others flows to other consumers through both quality and quantity cues (De Vany & Lee, 2001). In the case of online videos, quality information is transmitted through word-of-mouth communication (WOM). WOM is encouraged on video sites by allowing the embedding of videos on users' personal blogs or

---

[1] Joe Bereta and Luke Barats are an example. Their fame on Youtube landed the duo a six figure comedy contract by NBC (Adalian, 2006).

[2] One prominent example of a successful viral marketing campaign was that of lonelygirl15, who fooled audiences into thinking that she was a regular teenager, when in fact she was a paid actress promoting a movie (Heffernan & Zeller, 2006).

by emailing the video link. The sharing of video links with others is a major force behind the online video market. In a report by Pew Internet & American Life Project, 10% of online video viewers posted video links on their personal websites, 57% of them reported sharing the links of videos with others, and 75% of them said that they have had received video links from others (Madden, 2007). Additionally, quantity information is also communicated to users on video sites through the number of historical views and comments that accompany each video.

The intrinsic nature of online videos as an experience good with little or no pre-release advertising may lead to a higher reliance on information cues from past patrons. The quality of an online video is impossible to evaluate beforehand, and without advertising capabilities, the popularity of online videos are extensively dependent on the information cues provided by previous audiences. These characteristics foster and promote informational cascades which occur best under uncertainty conditions (Bikhchandani, Hirshleifer, & Welch, 1992). The possibility of informational cascades at work is further signaled from the stark variation in popularity among videos. Gill et al. (2007) found that almost 70% of Youtube videos were only viewed once, while the top 20% of videos managed to gain more than half of all views.

Informational cascades and herding behavior

In decision making, people not only rely on their private signals, but also use the behaviors of preceding others. Such imitative behavior occurs because of the belief that the behaviors of others are the result of information that is important but unknown to the decision maker. Several scholars have demonstrated how deferring to a decision that is consistent with preceding decisions, despite one's own private signal, can result in an informational cascade or in herding behavior, where each subsequent agent eventually defers to the preceding decision (Banerjee, 1992; Bikhchandani, Hirshleifer, & Welch, 1992). De Vany and Lee

(2001) further highlighted the importance of quality information, which is feedback communicated through WOM, in the formation and preservation of informational cascades. These informational cascade models have obtained strong support from laboratory experiments (Anderson & Holt, 1997; Celen & Kariv, 2004; Hung & Plott, 2001).

One key mediating factor in the formation of informational cascades is the level of uncertainty (Bikhchandani, Hirshleifer, & Welch, 1992; Payne et al, 1993). People with less private information or less confidence in the accuracy of one's information, are more likely to rely on cues from previous consumers.

Informational cascades and herding behavior on online platforms

In recent years, the impact of informational cascades or herding behavior on digital platforms has gained some scholarly attention. Due to the prominence of online recommender systems which simply and effectively provide new agents with information about the behaviors of preceding others, informational cascades are expectedly a major force behind decision making online.

Recommender systems are integral features on a diverse assortment of websites, including e-commerce sites like Amazon, digital auction site like eBay, and sites providing media content (e.g., music on Yahoo! Music and videos on Youtube). Quantity information about the behaviors of preceding others is provided by historical view counts or downloads, while quality information is transmitted through user reviews, comments and ratings. Quality information is further provided when users communicate about a product or provide a link on their personal websites, or on forums, and also through offline WOM.

The impact of informational cascades has been demonstrated in a multitude of online platforms including online auctions on eBay (Dholkia & Soltyinski, 2001), software downloads on AOL (Hanson & Putler, 1996), online news websites (Knobloch, Sharma, Hansen, and Alter, 2005) and e-commerce websites like Amazon (Huang & Chen, 2006). The

first study modeled real data obtained from eBay to give evidence of the presence of informational cascades, while the other studies utilized laboratory and field experiments to demonstrate how participants defer to decisions from previous others.

This study empirically models the viewing distribution of videos in an online repository. The view count data of 1,304 videos are enumerated daily, for a period of 33 days. In particular, we test for the presence of informational cascades, and model the impact of video quality uncertainty on the extent of an informational cascade.

Hypotheses

Based on informational cascade models (Banerjee, 1992; Bikhchandani et al, 1992; De Vany & Lee, 2001), the number of historical views as provided by recommender systems can be expected to affect agents' online media selection in two ways. Firstly, historical views can be used as a measure of the popularity of a video. From an information-based perspective, historical views provide quantity information about the behaviors of preceding users, which agents tend to defer to, when making decisions.. Secondly, in the absence of other marketing efforts, videos rely extensively on WOM to get more views[3]. The extent of WOM is likely to be positively related to the number of historical views, which in turn increases the additional views gained in the next time period. Hence, we hypothesize that:

H1: *Ceteris paribus*, the number of historical views of a video before a fixed time period attenuates the additional views gained in the next time period.

The video description provided by the author may also have an impact on the number of views gained in a time period. Past research has shown that with all other factors fixed, auction items which were accompanied by pictures were sold at higher prices on eBay as the inclusion of pictures reduced product quality uncertainty and made such listings more

[3] Due to the sheer number of videos online, WOM works by firstly, increasing the awareness of the existence of the video which indirectly also attenuates the number of views. Positive WOM further incites interest about the video to increase the number of video views.

attractive (Kalyanam & McIntyre, 2001). On some digital video platforms, posting-authors can provide more information by increasing the amount of textual information in the video description, or by providing thumbnails which act like snapshots of the video. When agents search or browse for videos, richer information provided in the videos' descriptions allow potential viewers to make more informed choices about the videos' content, giving such videos an edge over other videos.

H2: *Ceteris paribus*, the amount of textual information provided in the video description increases the number of views gained in a time period.

H3: *Ceteris paribus*, the inclusion of thumbnails in the video description increases the number of views gained in a time period.

We also expect an interaction effect between the reliance on historical views and the amount of available information about the video. Bikhchandani et al. (1992) theorized that informational cascades are more likely to occur with higher uncertainty levels of product quality, as decision makers are less likely to trust the accuracy of their private signals, and more likely to use the behavior of preceding others to make their choices. This was also demonstrated empirically by Dholkia and Soltyinski (2001) who found that herding behavior was more pronounced with products whose quality were hard to ascertain. With regards to online videos, uncertainty lies in the amount of information provided by a posting author in the video description. The presence of thumbnails and more textual information reduces content quality uncertainty and may mediate the cascading effect. Hence we hypothesize that:

H4: *Ceteris paribus*, the impact of a video's historical view count on subsequent views gained is mediated by the amount of textual information provided.

H5: *Ceteris paribus*, the impact of a video's historical view count on subsequent views gained is mediated by the presence of thumbnails.

Method and Data

The impact of informational cascades on online media content selection and its relationship with content quality uncertainty are tested using data from the Internet Archive (www.archive.org). The Internet Archive holds digital content (text, audio, videos, software and archived web pages) to attempt to preserve electronic artifacts from "disappearing into the past". The website is moderately popular and according to Alexa.com, it is ranked within the top 300 websites, worldwide, in terms of traffic flow. Compete.com, another site analytics website, reported over a million unique visitors to the Internet Archive in a month.

We focus on the moving images repository, which at last count on 24 October 2007, held a collection of over 101,000 videos (www.archive.org/details/movies). In terms of popularity, the moving images repository on the Internet Archive is overshadowed by other more popular video-sharing websites[4]. However, this website was chosen among other video-sharing websites because of two qualities. Firstly, due to the sheer number of videos released daily on more popular websites, it was a practical challenge to collect data of all videos released in a period of time. It was also not possible to obtain a representative sample of videos on other video-sharing websites as complete lists of all videos were not available. Obtaining a complete survey of all videos released on the Internet Archive website in a fixed time period was feasible, as the average number of videos released daily is about 350. Secondly, the video views on more popular websites increase at tremendous speeds, making it difficult to obtain repeated enumeration of view counts. By examining the Internet Archive which is a moderately, but not hugely popular website, the changes in video views could be tracked daily with ease.

The moving images section of the Internet Archive is similar to other video-sharing websites, so the results using videos from this website may be generalized to other video-

[4] The Internet Archive video repository is missing from the top 10 rankings of video websites, released by compete.com. The rankings can be found on their corporate blog, http://blog.compete.com/2007/02/02/google-video-youtube-top -video-sites/

sharing websites. Like other video websites, users can watch videos by searching using keywords or by browsing through a complete list of all videos on the website itself. Videos can also be accessed through embedded links on external websites. Each video on the Internet Archive is also accompanied by rating scores, user reviews, and the number of historical views. When posting videos on the website, posting-authors can choose to provide textual descriptions and/or thumbnails. Thumbnails are a sequence of snapshots of the video and are presented on the website as an animated .gif file. This thumbnail feature is quite unique to the Internet Archive website as most other video-sharing websites only provide a single still frame of the video.

The data are collected from the website using a data scraping software. The data scraping tool was first programmed to check the website's RSS feed every hour. Newly-released video links were collected and data of each video were scraped from the links. After 24 hours, the software accessed the same video links again to scrape for view count data. This step was scheduled to run every 24 hours.

To test the hypotheses in this study, we looked primarily at the view count of a collection of videos, across a time period of 33 days. According to one of the website's announcements, the view count on each video page increases when the video is viewed from the Internet Archive website, through external embedded links, and also when the video is downloaded to a computer.

$Views_{vt}$ denotes the view count of video v, in time t. Data from newly released videos on five consecutive days were collected within an hour of their release such that there were no views at the first instance (i.e., $Views_{vt=0} = 0$). A total of 1304 videos were enumerated[5]. Subsequently, view count data from each video were collected again daily and t denotes the

[5] A total of 1502 videos were actually released in the five days. However 198 videos were subsequently removed by the site's moderators due to content issues.

number of days that had passed since the release of the video. The videos were tracked for a period of 33 days (t = 0 to 33), giving a panel data with 40,159 *vt* observations[6].

Additionally, data about the content quality uncertainty level were collected. On the website, thumbnails consist of multiple snapshots of the video, and give users a sneak preview of the video while textual information also provide more information about the content of the video. Videos with higher information allow potential viewers to make more informed choices about whether the videos are relevant or enjoyable, and are hypothesized to be more appealing than other videos. $Text_v$ is the number of words in video *v*'s author-provided description and $Thumb_v$ is the dummy variable that is 1 when a video *v* has pictorial thumbnails, and 0 otherwise. The average number of words was 11.1 (S.D. = 18.16), and of the 1,304 videos, 52% of videos had thumbnails. $Text_v$ and $Thumb_v$ measure the uncertainty of quality of video *v* and are time-invariant.

$?Views_{vt}$ is the number of additional views obtained in time period, *t*, that is,

$$?Views_{vt} = Views_{vt} - Views_{v(t-1)}$$

There were 37,457 valid computations of $?Views_{vt}$. Table 1 gives a summary of the mean $?Views_{vt}$ at each time interval, *t*. Figure 1 graphically presents the average $?Views_{vt}$ over time. On average, videos gained the most number of views early into its release (before day 3). Thereafter, $?Views_{vt}$ started to saturate, and steadily tended toward zero. The saturation is, however, quite a long process, and even after 28 days, the videos were still viewed an average of 10.16 times.

The view disparity among all videos can be seen from the Lorenz curve in Figure 2. The Lorenz curve is plotted for the videos' total view count as of *t* = 28. The Gini coefficient

[6] There were 1624 missing values in total. 1011 missing values (164 missing values at t = 1, 239 missing values at t = 2, and 442 missing values at t = 3 and 166 missing values at t=4) were due to technical errors with the data scraping software. The other 613 missing values were due to problems with the website server on different days of the data collection.

was 0.853, indicating substantial inequality of views among the videos[7]. If there was equality, the graph would appear as one straight diagonal line (indicated by the solid line). However, our dataset deviated from this in the convex direction (the dotted line) and the curve informs us that 88% of all views were obtained by only 20% of videos.

Model specification

To test H1, the data are first tested for autocorrelated growth through a log-linear rank-views regression. Deviations from a log-linear rank-revenue relationship in the concave direction indicate positive autocorrelated growth (Ijiri & Simon, 1974). The presence of autocorrelated growth tested using this regression was provided as indirect evidence of WOM and a cascading effect in the movie industry[8] (De Vany & Walls, 1996). Substituting revenue for the number of views (i.e., $Views_{vt}$), we tested for the presence of autocorrelated growth using the following rank-views equation:

$$\log Views = C_0 + ß_1 \log Rank + ß_2(\log(Rank)^2)$$

Data from $t = 28$ are used in the above regression equation as this was the latest day with the most number of observations ($n = 1{,}279$). $Rank$ denotes the ranking of the videos based on their view count, such that the video with the highest number of views is ranked first. If $ß_2$ is found to be statistically insignificant, then there is an absence of autocorrelated growth. Should $ß_2$ be statistically significant and in the positive direction, then the rank-views relationship is concaved upwards and the data exhibits negative autocorrelated growth. If $ß_2$ is statistically significant and has a negative value, then the rank-views relationship is concaved downwards and the data exhibits positive autocorrelated growth. From H1, $ß_2$ is

[7] The Gini coefficient is a measure of inequality and can take on values between 0 and 1. A value of 0 indicates equality and a larger coefficient indicates more inequality.

[8] De Vany and Walls (1996) tested and rejected other distribution dynamics that could explain autocorrelated growth, and found that the only plausible explanation was a Bose-Einstein distribution and a Bayesian model. Essentially, at each following time period, agents updated their expectations using information provided from the previous round of viewers, which resulted in the growth being positively autocorrelated (i.e., correlated with previous lagged values). We expect similar dynamics to operate on the digital video platform as well. In our case, information transmission occurs through WOM, and also through quantity information provided by the videos' historical view count.

expected to be statistically significant and in the negative direction, indicating the positive impact of WOM and a cascading effect.

To test H1-H5, a second log-linear model is adopted as $?Views_{vt}$ is expected to be multiplicative with respect to the independent variables. Furthermore, regressing the logarithmic transformation of the variables reduces the impact of outliers.

From the theory of informational cascades, the additional views gained by videos in a time period are expected to receive a positive boost from the number of historical views (H1). From H2 and H3, the presence of thumbnails and richer textual information in the video description are also expected to increase the number of views in each time period. That is,

$$?Views_{vt} = C_{0?} \cdot [Views_{v(t-1)}]^{a} \cdot Text^{a} \cdot \exp(b \cdot Thumb_{v}) \quad (1)$$

The parameter a, denotes the cascading effect. A larger magnitude of a, indicates a larger cascading effect.

From H4 and H5, the impact of $Views_{v(t-1)}$ (i.e., the magnitude of the cascading effect, a) is expected to be mediated by the content quality uncertainty level. So,

$$?Views_{vt} = C_{0} \cdot [Views_{v(t-1)}]^{a}\, Text^{a} \cdot \exp(b \cdot Thumb_{v}) \quad (2)$$

$$\text{Where } a = a + ß_{1}Text_{v} + ß_{2}Thumb_{v}$$

Apart from the positive boost received from the cascading effect, the videos also experience a diminishing or decaying effect from saturation. From figure 1, which shows the graph of $?Views_{vt}$ over time, it can be seen that the change in views decreases over time as more people watch the videos and the pool of potential audience continually diminishes.

In order to model the cascading effect more precisely in the presence of saturation effects, we included a decay factor. Hence,

$$?Views_{vt} = C_{0} \cdot ?^{t} \cdot [Views_{v(t-1)}]^{a} \cdot Text^{a} \cdot \exp(b \cdot Thumb_{v}) \quad (3)$$

$$\text{Where } a = a + ß_{1}Text_{v} + ß_{2}Thumb_{v}$$

The term $?^t$ is flexible enough to capture both saturation as well as explosive effects. However, in our dataset, ? is expected to be smaller than 1, which would indicate a saturation effect over time. The larger the magnitude of ?, the larger is the decay factor, and the faster $?Views_{vt}$ diminishes to zero.

Taking logarithm on both sides,

$$\log(?Views_{vt}) = \log C_0 + log? \cdot t + a \cdot logViews_{v(t-1)} + ß_1 \cdot Text_v \cdot \log Views_{v(t-1)}$$
$$+ ß_2 \cdot Thumb_v \cdot \log Views_{v(t-1)} + a \cdot \log Text_v + b \cdot Thumb_v \quad (4)$$

This is the final model specification. A logarithmic transformation poses a technical situation for data with values equal to zero as the value of log0 is undefined. To correct this situation, 1 is first added uniformly to all values of $?Views_{vt}$, $Views_{v(t-1)}$ and $Text_v$ before the logarithmic function is applied. Such a uniform transformation makes the logarithmic transformation feasible while not distorting the distribution of the variables.

Table 2 presents the zero-order correlation among the variables.

Results

Table 3 shows the log-log rank-views regression results. The parameter-estimate of $(\log(Rank_{vt}))^2$ was highly significant and in the negative direction ($p < .000$, $ß_2 = -.490$). This result gives strong support for positive autocorrelated growth, which is evidence of a cascading effect. H1 is supported.

Table 4 presents the regression results of equation 4. For clarity, the parameter notations as listed in equation 4 are included in the table. Overall, the regression model significantly explained 43.2% of the variation of $?Views_{vt}$ ($F = 4746.185$, $p = .000$).

Saturation effects

As expected, the videos experienced a saturation effect, and $?Views_{vt}$ decreases over time ($p = .000$). The decay parameter $?$ is obtained by taking the antilog of the coefficient of t

in the regression model ($? = e^{-0.05}$). In real terms, the decay parameter, ?, is 95.1%. This large magnitude indicates a substantial saturation impact, with videos gaining substantially less views as they get older.

H1: Impact of $Views_{v(t-1)}$ on $?Views_{vt}$

$logViews_{v(t-1)}$ is statistically significant ($p < .000$), indicating the significant impact of $Views_{v(t-1)}$ on $?Views_{vt}$. H1 is supported. In real terms, an increase of 100% in $Views_{v(t-1)}$ increases $?Views_{vt}$ by 41.7%. The elasticity between the additional views gained and the number of views in the previous time period is quite large, indicating that the cascading effect is quite pronounced, with more popular videos gaining a moderately large boost from its historical view count. This result supports the theory of informational cascades and sheds light on one of the key dynamics behind the selection of online videos.

H2 and H3: Standalone effect of $Text_v$ and $Thumb_v$

The regression results showed that $Thumb_v$ had a negative and significant effect on $?Views_{vt}$ ($p < .000$). This result is contrary to H3, and H3 is rejected. However, although the parameter estimate of $Thumb_v$ is statistically discernible from zero, the real impact of thumbnails is quite small. In real terms, the presence of a thumbnail only decreased the number of additional views by 5.0%

$Text_v$ did not have a significant effect on $?Views_{vt}$. H2 is also rejected.

H4 and H5: Interaction effects between $Text_v$, $Thumb_v$, and $Views_{v(t-1)}$

We found no significant effect of $Text_v \cdot \log Views_{v(t-1)}$ and H5 is rejected.

As hypothesized in H6, $Thumb_v \cdot \log Views_{v(t-1)}$ was found to be statistically significant ($p = .000$) and the inclusion of thumbnails significantly mediated the cascading effect. In real terms, videos without thumbnails experienced a cascading effect which was 12.5% larger than videos with thumbnails. H6 is supported.

Thumbnails reduced the uncertainty about video quality, and enabled viewers to make more informed choices about the attractiveness of the video. This reduced the reliance on information supplied by the behaviors of preceding others. Hence, there was a reduction in the influence of cascades for videos which had thumbnails.

Discussion

Online videos are already hugely popular, and the consumption of videos via the Internet is expected to be an increasing trend in the coming years (Lee, 2006). This study comes at a timely moment, and is one of the first to shed some light on the view distribution and evolution of online videos. Results confirmed that the view distribution of online videos is highly disparate, with only a few popular videos garnering a large percentage of all views. This inequality in popularity is much greater than other media industries like movies (De Vany &Walls, 1996), home video (Elberse & Oberholzer-Gee, 2007) and books (Anderson, 2006).

One interesting trait we found about online videos, was its rapid saturation rate. Although videos continually attracted a handful of views even months later, the bulk of views were obtained very early on (within 2 -3 days of the video's release). These results indicate a very short window of opportunity for which videos can garner popularity, and unless a video manages to make headway during this time period (like getting on a featured list, or being spotted by an editor or an opinion leader), audiences very quickly lose interest.

By drawing on theories of consumer behavior, this study also provided theoretical explanations for the disparity in views among online videos and the dynamics behind how audiences select their media content. We found strong empirical support for the presence of informational cascades in the selection of online videos. In real terms, the cascading effect

had a moderately large impact on increasing view numbers; videos which had views counts of 100% more, gained a 41.6% boost in additional views during the next time period.

We also found partial support for the role of uncertainty in the mediation of the cascading effect. Uncertainty was theorized to increase the reliance on the behaviors of preceding others as it fostered imitative behavior (Bikhchandani et al., 1992). In this study, we found that the inclusion of thumbnails by the posting-author reduced content uncertainty level, and significantly mediated the cascading effect. However, textual information in the video description had no impact on the cascading effect. The thumbnails as provided on the Internet Archive website were in the form of an animated .gif, which appears as a sequence of moving pictures. Moving images have been found to induce orienting responses and call for attentional resources (Reeves & Nass, 1996), which may have diverted attention away from the textual information.

Very surprisingly, we also found that the standalone effect of including thumbnails actually decreased the number of additional views. However, this impact was minute, and not very significant in real terms.

The results from this study are not limited to merely online videos. Similar recommender systems exist on an assortment of websites providing different media content, and the WOM dynamics behind the selection of different media content are likely to be similar, making the results from this study important to the study of other online media content.

Conclusion

The web has evolved from one-to-one communication via applications like MSN, to many-to-many communication. Coined web 2.0, the new Internet era is built around social networks, communities, and user-generated content (O'reilly, 2005). With regards to online

videos, user-made videos ensure a sustained supply of videos, with more that 65,000 new videos uploaded to Youtube daily ("Youtube serves up 100 million videos", 2006). However, the extent to which a single video will be watched by a large audience depends extensively on how well the video can tap onto the benefits of WOM, and ride the cascading wave.

Results from this study highlight the importance of informational cascades in ensuring the popularity and viability of an online video. It is crucial for videos to gain a large number of views at the beginning of their run, as this would ensure that they continually remain popular. However, we did not examine the video-specific features that help ignite the initial audience base. Are some videos intrinsically more appealing than others? What content or formal features help to increase the number of views in the first instance? Future studies will be needed to reveal a more complete picture of the viewership of online videos.

One major limitation of this study is the lack of information about whether viewers watch a video through an external link, or by browsing and searching for videos on the website. Such information will serve to tease out the impact of quality information (i.e., WOM through the posting of the embedded video link on an external website) and quantity information (provided by the number of historical views) on the cascading effect. This study also did not examine the impact of ratings and reviews. Future studies may consider investigating these aspects and their impact on increasing the popularity of online videos.

References

Adalian, J. (2006, September 26). NBC clicks Youtube duo. *Variety*. Retrieved October 31, 2007, from http://www.variety.com/article/VR1117950782.html?categoryId=14 &cs=1&s=h&p=0

Anderson, C. (2006). *The long tail: Why the future of business is selling less of more*. New York: Hyperion Books.

Anderson, L. R., & Holt, C. A. (1997). Information cascades in the laboratory. *American Economic Review, 87*(5), 847-862.

Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics, 107*(3), 797-817.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy, 100*(5), 992-1026.

Celen, B., & Kariv, S. (2004). Distinguishing informational cascades from herd behavior in the laboratory. *American Economic Review, 94*(3), 484-498.

De Vany, A., & Lee, C. (2001). Quality signals in information cascades and the dynamics of the distribution of motion picture box office revenues. *Journal of Economic Dynamics & Control, 25*(2001), 593-614.

De Vany, A., & Walls, W. D. (1996). Bose-einstein dynamics and adaptive contracting in the motion picture industry. *The Economic Journal, 106*(439), 1493-1514.

Dholakia, U. M., & Soltyinski, K. (2001). Coveted or overlooked? The psychology of bidding for comparable listings in digital auctions. *Marketing Letters, 12*(3), 225-237.

Elberse A. & Oberholzer-Gee (2007). Superstars and underdogs: An examination of the long tail phenomenon in video sales. Working paper, Harvard Business School. Retrieved October 23, 2007, from http://www.people.hbs.edu/aelberse/papers/hbs_07-015.pdf

Gill, P., Arlitt, M., Li, Z., & Mahanti, A. (2007, October). *Youtube traffic characterization: A view from the edge*. Paper presented at the Internet Measurement Conference, San Diego, CA. Retrieved October 31, 2007, from www.imconf.net/imc-2007/ papers/imc78.pdf

Hanson, W. A., & Putler, D. S. (1996). Hits and misses: Herd behavior and online product popularity. *Marketing Letters, 7*(4), 297-305.

Heffernan, V. & Zeller, T. (2006, September 13). The lonelygirl that really wasn't. *New York Times.* Retrieved October 31, 2007, from http://www.nytimes.com/2006/09/13/ technology/13lonely.html

Huang, J. H., & Chen, Y. F. (2006). Herding in online product choice. *Psychology & Marketing, 23*(5), 413-428.

Hung, A. A., & Plott, C. R. (2001). Information cascades: Replication and an extension to majority rule and conformity-rewarding institutions. *American Economic Review, 91*(5), 1508-1520.

Ijiri, Y. & Simon, H. A. (1974). Interpretations of departures from the Pareto curve firm-size distributions. *Journal of Political Economy, 82*(2), 315-332.

La Monica, P. R. (206, October 6). Google to buy Youtube for $1.65 billion. *CNN News*. Retrieved October 31, 2007, from http://money.cnn.com/2006/10/09/technology/ googleyoutube_deal/

Lee, E. (2006, July 17). In video shake-out, will Eefoof go poof? *San Franscisco Chronicle*, Retrieved October 23, 2007, from http://www.sfgate.com/cgi-bin/article.cgi? file=/c/a/2006/07/17/BUGA3JUOAB1.DTL&type=tech

Kalyanam, K. & McIntyrem S, (2001). Return on reputation in online auction markets. Working paper, Santa Clara University. Retrieved October 31, 2007, from http://business.scu.edu/faculty/research/working_papers/pdf/ kalyanam_mcintyre_wp10.pdf

Knobloch, S., Sharma, N., Hansen, D. L., & Alter, S. (2005). Impact of popularity indications on readers' selective exposure to online news. *Journal of Broadcasting & Electronic Media, 49*(3), 296-313.

Madden, M. (2007). *Online video*. Washington: Pew Internet & American Life Project. Retrieved October 20, 2007, from http://www.pewinternet.org/pdfs/ PIP_Online_Video_2007.pdf

O'Reilly, T. (2005). *What is web 2.0: Design patterns and business models for the next generation of software*. Retrieved October 1, 2007, from http://www.oreillynet.com/ pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York: Cambridge University Press.

Pew Research Center (2000). *Internet sapping broadcast news audience.* Retrieved October 23, 2007, from http://people-press.org/reports/pdf/36.pdf

Reeves, B. & Nass, C. (1996). *The media equation: How people treat computers, television and new media like real people and places*. New York: Cambridge University Press.

Xu, C., Dale, C., & Liu, J. (2007). Understanding the characteristics of internet short video sharing: YouTube as a case study. Working paper, Simon Fraser University. Retrieved October 23, 2007, from http://www.camrdale.org/Resume/youtube.pdf

Youtube hits 100m videos per day. (2006, July 14). *BBC News*. Retrieved October 23, from http://news.bbc.co.uk/2/hi/technology/5186618.stm

Youtube serves up 100 million videos a day online. (2006, July 16). *USA Today*. Retrieved October 23, 2007, from http://www.usatoday.com/tech/news/2006-07-16-youtube-views_x.htm

Zillmann, D., Knobloch, S., & Yu, H-S. (2001). Effects of photographs on the selective reading of news reports. *Media Psychology, 3*(4), 301-324.

Table 1
Summary statistics of $?Views_{vt}$ *at each time period, t*

		$?Views_{vt}$				$?Views_{vt}$				$?Views_{vt}$	
t	*n*	*M*	*SD*	*t*	*n*	*M*	*SD*	*t*	*n*	*M*	*SD*
1	1119	35.916	183.711	12	1261	18.295	69.685	23	1269	11.878	60.983
2	890	83.721	264.714	13	1262	16.294	79.700	24	1259	6.322	35.465
3	623	106.568	399.473	14	1271	17.131	113.134	25	1252	11.747	59.611
4	684	71.055	551.193	15	1272	17.658	102.993	26	1249	12.855	94.648
5	1136	31.828	179.995	16	1242	17.157	111.825	27	1255	9.748	73.670
6	1301	25.200	143.813	17	1228	6.900	37.432	28	1258	10.160	76.869
7	1300	32.275	205.102	18	1242	12.752	53.693	29	1245	12.769	74.565
8	1291	17.691	79.780	19	1244	11.306	44.916	30	1020	11.176	70.275
9	1284	12.172	49.790	20	1249	12.590	51.067	31	785	11.268	103.307
10	1279	16.589	83.620	21	1246	12.571	61.776	32	591	15.226	104.423
11	1275	8.453	53.073	22	1256	11.478	59.420	33	319	13.445	76.772

Note. $?Views_{vt}$ = the number of additional views gained in the *t*th day after the release of the video.

Table 2
Zero-order correlations of ?Views$_{vt}$ and the independent variables

	1	2	3
1. *?Views$_{vt}$*			
2. *Views$_{v(t-1)}$*	[a] .311		
3. *Text$_v$*	[a] - .042	[b]-.067	
4. *Thumb$_v$*	[a] -.063	[b]-.113	[c].191

Note. *Views$_{v(t-1)}$* = the total view count of video *v* as of day *t* – 1. *Text$_v$* = the number of words in video *v*'s description. *Thumb$_v$* = 1 if thumbnails were included in the video description, or 0 otherwise. [a]*N* = 37,506. [b]*N* = 38,870. [c]*N* = 41,783.

Table 3
The loglinear rank-view regression results (dependent variable = logViews)

Variable	*B*	*t*
logRank	2.861*	42.833
$(logRank)^2$	- .490*	-77.887
constant	5.320*	30.397
R^2	.974	
Adjusted R^2	.974	
N	1279	

*$p < .001$

Table 4

The loglinear regression results of $Views_{v(t-1)}$, $Text_v$, $Thumb_v$ and the interaction terms on ?$Views_{vt}$

Variable	Coefficient notations	*B*	*t*
t	log?	-0.051***	-72.758
$logViews_{v(t-1)}$	*a*	0.417***	117.641
$Text_v \cdot logViews_{v(t-1)}$	$ß_1$	0.00002	0.796
$Thumb_v \cdot logViews_{v(t-1)}$	$ß_2$	-0.052***	-10.064
$logText_v$	*a*	-0.003	-0.651
$Thumb_v$	*b*	-0.050*	-2.293
constant		0.520***	26.445
R^2		0.432	
Adjusted R^2		0.432	
N		37,457	

*$p < .05$ ** $p < .01$ *** $p < .001$

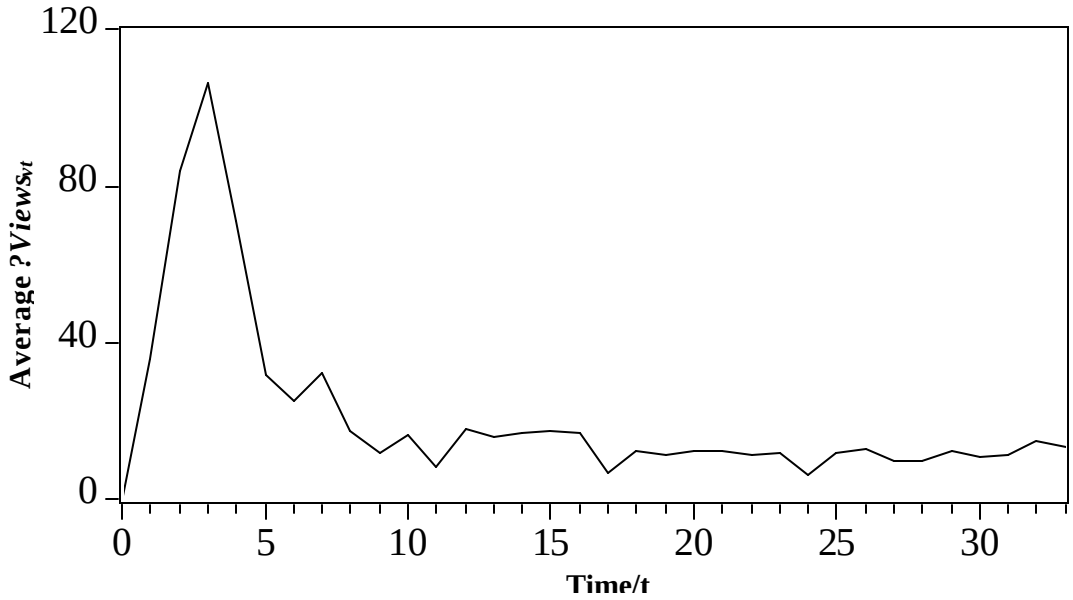

Figure 1. The average number of additional views gained (*?Views*$_{vt}$) with respect to time.

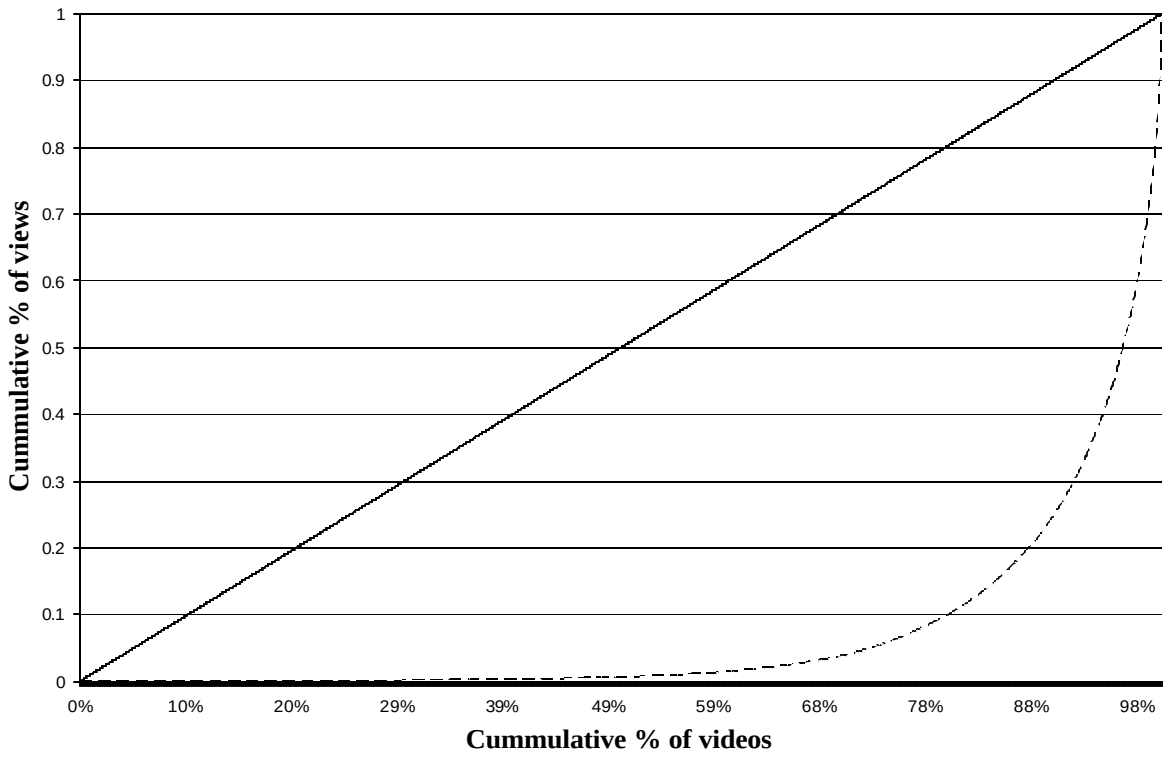

Figure 2. Lorenz Curve of the video views at t = 28. The solid line represents the expected graph if there was equality among videos in terms of views. The dotted line represents the observed distribution of our dataset.