

Going Viral – The Dynamics of Attention

G. R. Boynton
Department of Political Science
University of Iowa
bob-boynton@uiowa.edu

Abstract

A predominant story about YouTube is 'going viral'. 'Going viral' is about dynamics; changes in views of videos through time. The paper begins with a question about how to move the phrase from vernacular to a concept that political scientists might use. By looking at three possible interpretations of the phrase I show that the 2008 campaign videos of McCain and Obama were unlikely to be characterized as 'going viral.' However, viewing the campaign videos does have a very regular dynamics that can be conceptualized straightforwardly and represented in a simple dynamics equation. By examining each of the approximately 800 videos of the campaign the regularity in the dynamics of viewing these videos is demonstrated. After setting out the general pattern I look at interaction between videos in the campaign that seems to preclude going viral. The final point is that an examination of the impact of exogenous factors needs to start with the general pattern to estimate that impact otherwise the standard/expected views is confounded in the treatment of the exogenous factor.

The predominant stories about the web are “going viral” and “the long tail.” Going viral is a story about the dynamics of attention. And the dynamics of attention to the videos the McCain and Obama campaigns posted on YouTube is the focus of this report. How the number of views changed through time is the question being addressed. The long tail is addressed in another report (Boynton, November 1, 2008)

The web was designed by Tim Berners-Lee in 1990 and went “public” in 1991. However, video on the web did not come into its own until 2005, which was too late for the 2004 presidential election. Neither Bush nor Kerry did much with video on the web in the 2004 election. It was not until bandwidth reached a level that could carry video and YouTube was able to do the aggregating and distributing that video on the web took off. Thus, 2008 was the first presidential election that video on the web was a possible campaign strategy. By the end of 2008 more people said they got most of their news for national and international affairs from the web, 40 percent, than from newspapers, 35 percent. The web trailed far behind television, which 70 percent said was their chief source of news, but for people younger than 30 the web was said to be the primary source of news as frequently as television (Pew, December 23, 2008).

The presidential candidates, especially the Obama campaign, recognized the importance of video on the web, and set up channels on YouTube where their videos were freely available. Between July 1 and the election the Obama campaign posted almost 800 videos on YouTube and the McCain campaign posted just over 100 there. And viewers came: the Obama videos were viewed more than 42 million times; the fewer McCain videos were viewed more than 20 million times. "The Ghosts of Campaigns Past and Campaigns Yet to Come" gives a full account of the two campaigns on YouTube (Boynton, December 14, 2008)

That is a lot of political behavior political scientists have not been able to analyze before. Citizens pay attention to the campaigns, but in the age of television it was difficult to obtain evidence about that attention. By the 1990s it became possible to track political TV ads, but information about people watching was not available. Samples could be asked about their attention to the campaigns, but that does not produce very reliable numbers. For example, the Pew research reported 70% of the respondents said they got most of their news from television, 40% said they got most of their news from the web, and 35% answered newspapers. Unfortunately, that adds up to considerably more than 100%. The YouTube campaigns were only one element of the election campaigns, and thus were only one element of the attention people were paying to the campaigns. However, the numbers collected by YouTube make it possible to do analyzes we have not been able to do in the past. This report focuses on the dynamics of viewing online campaign videos.

Going Viral

“Going viral” is vernacular. Exactly what would count as going viral and what would not count has not been carefully specified. Three possibilities will be considered.

Going Viral 1

An “obvious” use of the vernacular “going viral” is describing videos that are viewed many times. “Many” remains vague, however. Would the number of views of the campaign videos count as many or not?

In October, 2008 YouTube videos were viewed by 344 million unique viewers (Gannes, December 1, 2008). The campaign videos were viewed almost 21 million times in October. However, these were not unique viewers. Unique viewers for the campaign videos would be many fewer than 21 million since the campaigns posted more than 300 videos on YouTube that month, and one would anticipate that many viewers returned again and again to watch the videos as they were posted.

With the exception of views of Obama's Chicago speech election night the most views of any campaign video were just over 2 million. The two videos that have been viewed most often on YouTube were viewed 111 million and 108 million times by the end of December, 2008 which certainly dwarfs the views of campaign videos (Gannes, December 29, 2008) .

"Celeb" was one of the most frequently viewed videos of the campaign, but when it had been viewed 2 million times in the middle of September the Paris Hilton spoof of the McCain video had been viewed 7.4 million times (Boynton, September 18, 2008).

The campaign video of Obama's speech in Chicago election night was viewed 4 million times. But "I Got a Crush on Obama by Obama Girl" was viewed 12 million times on the Barely Political YouTube channel. And "Yes We Can," the video produced by Will.i.am, was viewed more than 20 million times.

Crane and Sornette examined daily views for nearly 5 million videos and classified them into four categories. While the classification is based on dynamics rather than number of views they reported the average total views for the class with the most views was 33,693 and for the smaller class as 16,524 views (Crane and Sornette, 2008).

With mean views of this magnitude one might say that any video viewed more than, say, 100,000 times was going viral. All that would be required is for it to seem like quite a large number of views to people who follow YouTube carefully, and that is probably what is behind much of the talk about going viral. You recognize it when you see it -- one might say when it comes to the numbers (Albrecht, September 24, 2008).

Going Viral 2

A second version of “going viral” is as process. It is a “play” on the epidemiological version of going viral. The point is that there is no single source from which all persons are “infected.” Instead infection is spread through contact: It starts when a few individuals are infected. Each comes in contact with others who are then infected. Each of the infected at time 2 come into contact with additional individuals who are infected. And this continues until the process has run its course. That is the simple version of the biological process of epidemic.

How is that relevant to the web in 2008? Here is a paragraph that describes exactly the same process and also the structural change that would transform the process.

It seems like every day there's at least five new Twitter apps popping up. The news of a shiny new Twitter toy gets dispersed through Twitter streams everywhere, hits critical mass, news dries up a bit, and then, if you're like me, you forget about it completely.

So where do you turn when you need to reference multiple Twitter apps, see what others think, or view app ratings? You could ask your Twitter friends, do a quick search, or wander over to the Twitter Fan wiki, but there should be a better way, right? Given that Twitter doesn't offer a Twitter app store (and why the hell not?), Twitdom has jumped into the fray to be the answer to your twreams (twitter dreams) with their Twitter applications database. (Van Grove, 2009)

Yes, it is too cute by half, but that is the way some tech bloggers write. And it is about twitter apps rather than videos on YouTube. But it illustrates the thought process that results in drawing on the language of epidemic in thinking/writing about how the web is being used. In this case -- and in the case of finding a video to watch on YouTube -- the process is search. New videos are posted to YouTube every day. How do you find the videos you want to watch? You could ask your web friends, and then tell other web friends. You could search, but you are not quite sure what you are searching for so it is back to friends. Or there could be a common source, in this case a database, that would bring together what you are looking for. Then you only have to go to one place, and that would make the search very easy.

In 2005, 2006 and 2007 the videos posted to YouTube were largely "user generated content." The large music, movie, television, and news organizations steered clear of YouTube. And when they found that someone had placed a property of theirs on YouTube they demanded that it be removed. They had as little to do with YouTube as possible. That began to change in 2008, but for the first three years of its existence established corporate reputation could not be used as a search strategy. So the hundreds of millions of viewers relied on their friends for discovery and spreading the word. It was almost the only discovery procedure available.

The structure of spreading the word is: 1) making it your own by adding a copy to your own blog or website, which has been made easier in the last year as embedding has become popular, 2) blogging, i.e., writing publicly about the video and ordinarily including a url for the video, and 3) private communication that is also made easier by web connections. And Google is attempting to make money via exactly this same structure of going viral.

Google is starting to share more details about its high priority of making more money off YouTube's popularity, introducing an advertising product on Tuesday called buzz targeting.

The ad product uses an algorithm to find videos that are about to "go viral," when word of mouth (or IM, or blog, or e-mail) promotes a Web site to a phase in which it spreads like wildfire. In this case, ads are overlaid on the bottom fifth of viral videos supplied by YouTube partners who share ad revenue with the search giant (Shankland, May 13, 2008).

Check the instant messaging and blogging and email to note videos that are about to "take off" and sell ads.

Almost everyone would agree that Will.i.am's video "Yes we can" went viral. The version posted to YouTube February 2, 2008 (<http://www.youtube.com/watch?v=jjXyqcX-mYY>) was viewed, as of January 4, 2009, 15 million times. For News and Politics videos on YouTube it was the second most viewed of all time, the second most favorited of all time, and the fifth most discussed of all time. If you search, January 4, 2009, only on YouTube you locate 733 videos that can be identified by Will.i.am and the title. There would be many more copies if you searched the entire web. Kevin Wallsten (Wallsten, 2008) tracked the views, blog posts, and mentions in the traditional media, and concluded that blog posting, personal communication, was the driving force in viewing the pro-Obama video.

This is a persuasive story, and it is certainly possible to find other videos that seem to have the same character. President Bush's press conference on December 14 in Iraq at which a journalist tossed shoes at Bush took off in a "viral" fashion. Visible Measures reported

As of this morning (December 19, 2008), Bush vs. Shoes had generated more than 1,150 placements, 21.4+ million views, and 150,000+ comments (Visible Measures, December 19, 2008)

The first video was posted on December 15, and within four days it could be found at 1,150 places on the web and had been viewed 21.4 million times.

As persuasive as this story is it does not seem an apt characterization of the views of videos placed on YouTube by the campaigns. There are too few blogs posted for this to seem apt.

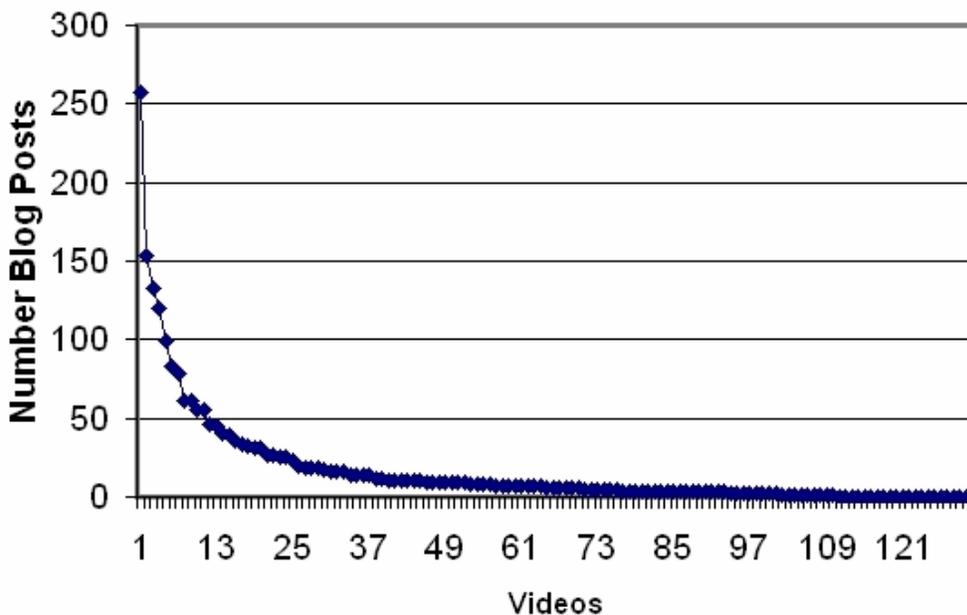


Figure 1: Blog Posts about Videos

I did a Technorati search for blog posts for each of the videos posted by McCain and Obama the first week in September and the first week in October. There were 39 videos in September and 93 in October for a total of 132.

The mean number of posts per video was 16.7. The range was 257 posts to 0 posts. The figure shows the distribution of blog posts for the videos. Only 30 videos had 17 or more blog posts. The other 102 videos received fewer than 17 blog posts. It is hard to imagine that 17 blog posts can generate very many views, and certainly they are not going to generate going viral. It is even more difficult to believe that 0 blog posts, which 22 of the videos received, were producing views of the campaign videos. You might argue that 257 blog posts would be enough communication to produce going viral. But even that seems unlikely when compared to communication received by "Yes we can" and Bush vs Shoes.

There is one other point relevant even though I cannot go into the same detail here. I will examine the data more systematically in a future paper. YouTube records the views that are sent to the video from other sources. So you might learn about a video at one of the major political blogs. If you look at their record for the top five sources you find two things. One, the top five "send" a tiny fraction of the total views. Two, a campaign website or websites is almost always among the top five. So the campaign is "sending" views which is just the point I want to make. You can go to the website to find a video just as you can go to YouTube to find a video.

If not blog posts what is generating views? The answer seems very simple. There is a directly relevant "database" -- the channels, and to a lesser degree, the websites of the campaigns. If you want to check out McCain videos or Obama videos all you had to do was go to the channels of the campaigns. You could subscribe to the channels if you were really interested or you could easily search for campaign videos and find them. The structure does not require the logic of epidemic.

Going Viral 3

The third version of going viral is not vernacular; it is a functional form.

If you follow the description of the process the number infected early would be small, but it would grow very quickly until all individuals available to be infected had been infected -- the process had run its course. That takes the shape of a sigmoid curve. Figure 2 illustrates this structure of change through time, and it can effectively be modelled by a first order quadratic equation. People talking about YouTube do not ordinarily have sigmoid curves in mind when they refer to "going viral," but it is what political scientists and

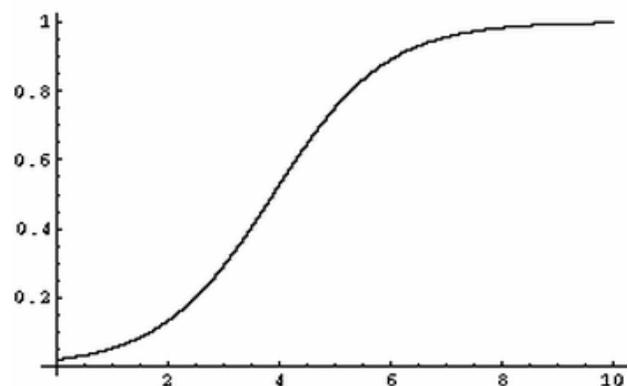


Figure 2: Sigmoid Curve

others are taught as a simple version of the epidemic process (Arneson, 2006).

However, a sigmoid curve is not what you find when plotting the views of campaign videos over time. Instead you get curves that look like these two figures (Figure 3).

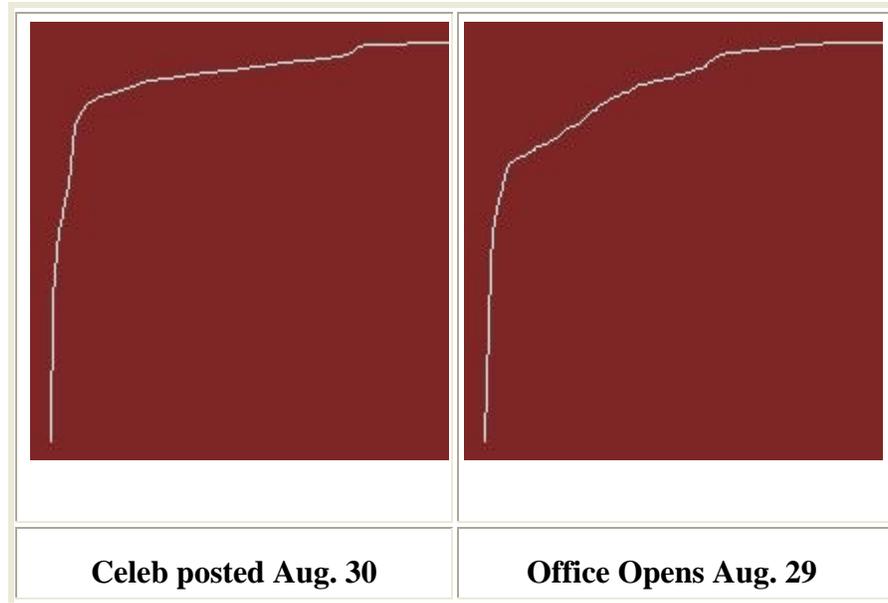


Figure 3: Curves of Campaign Videos

One is the McCain video "Celeb" that was posted August 30. By the end of the campaign it had been viewed 2.2 million times. The other was an Obama video "Detroit, MI Campaign for Change Headquarters Opens." It was posted on August 29 and by the end of the campaign it had been viewed 6400 times. I chose the two because the number of views was dramatically different. But the two curves -- the dynamics -- look similar and neither looks anything like a sigmoid curve.

One might be able to find videos on YouTube that change in time in a process that can be plotted as a sigmoid curve. But the videos of the campaign did not. I also used a service, Trendrr, to track one of the most viewed videos of the Bush vs. Shoe episode. This is only one of 1,150 videos of the event on the web. There may be others that take a sigmoid shape when plotted over time, but this one does not.

My conclusion is that the functional form that most easily captures the dynamics of going viral does not appear very often for political videos on YouTube. A different functional form is the predominant form. And it is to that functional form that I now turn.

What should one make of "going viral" talk? I certainly do not want to deny the possibility that some specification of "going viral" may be an accurate characterization of some videos on YouTube. However, I believe the structure of political campaigns will generally produce a different dynamics. We political scientists will not be able to participate in the going viral talk except in rare cases.

The Dominant Pattern

Two recent studies of the dynamics of viewing online videos are relevant to this examination of the dynamics of the videos of the 2008 presidential election campaign.

The first was by Tubemogul (Tubemogul, June 19, 2008). It is a firm that does video analytics and uploads clients' videos to a variety of sites on the web. The dynamics of viewing videos is important to them so they can tell their customers when to post and what to expect.

They used 10,916 videos and followed the daily views of each for 90 days. They analyzed both total views and average views over time, and found a pattern that is very similar to the pattern for the 2008 campaign videos on YouTube. By adding the number of views or taking an average they make the assumption that the videos are independent. That is undoubtedly a reasonable assumption for most videos on the web, but I will show that there are interdependencies between the campaign videos that cannot be captured given their procedure. Instead of analyzing a total or an average I will do a comparative analysis of the dynamics of individual videos.

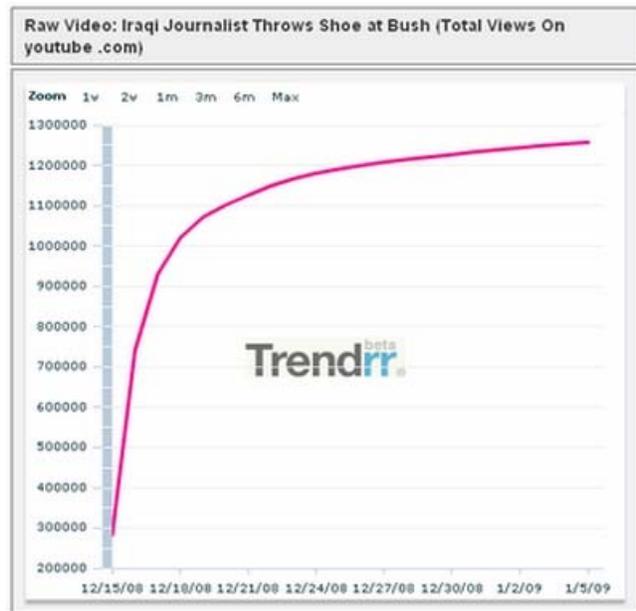


Figure 4: Tubemogul Analysis

There is a second way my research differs from theirs. They are interested in how fast viewing decreases; that is what their clients are concerned about. I believe political scientists are primarily interested in how many views a video receives. This is just a matter of "flipping" over the curve -- from views daily to views added to the total daily (see the methodological appendix for further explication of this point). But it gives a clearer picture of the political importance of the videos.

The second is the study by Crane and Sornette (2008) "Robust dynamic classes revealed by measuring the response function of a social system." The study was based on 5 million time series of views of videos posted to YouTube over an eight month period. They do not say how many days are included in each time series; the figure that gives distributions by day lists 50 days.

Their analysis of their model of the process yields a two by two table of dynamics. They found that 90 percent of the videos "either do not experience much activity or can be statistically described as a random process." They do not specify "much activity" so it is hard to compare that to campaign videos that range from a few thousand views to 2 million. But they locate 90 percent of the videos they studied in the top left cell: endogenous-sub-critical. Finding 90 percent in cell A is difficult to square with the Tubemogul study unless most of the 90 percent were excluded by Tubemogul when they did not consider videos that had a total number of views less than

1,000 in their study. There are no campaign videos for which views are related to time as in cell A. So campaign videos fall into the 10 percent that they find distributed into the other three cells.

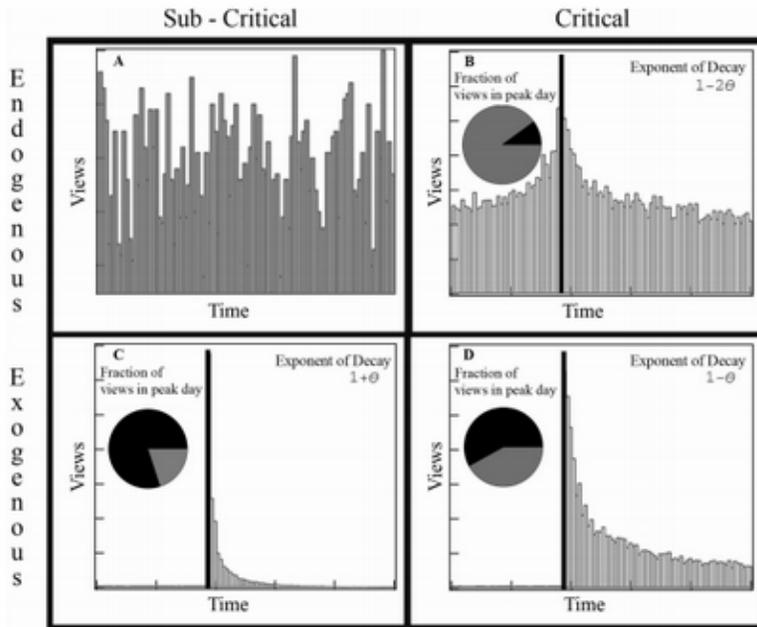


Figure 5: Crane and Sornette Analysis

The top right cell (endogenous-critical) would be an approximation of a sigmoid curve if you were looking at the cumulative total number of views instead of the views each day. They do not say how many videos are found in this class, but they do say that the average number of total views for this class is 33,693. So while the authors characterize these videos as viral they certainly are not viral by a definition that emphasizes the total number of views. And there are no campaign videos that can be described by a sigmoid

curve.

That leaves two cells: exogenous-sub-critical and exogenous-critical. Cell C, exogenous-sub-critical, is specified as videos for which more than 80 percent of the views occur on the peak day of viewing. Cell D, exogenous-critical, is videos for which between 20 percent and 80 percent of the views occur on the peak day of viewing. None of the campaign videos fall into cell C. The Tubemogul study found that on average only 25 percent of the views occurred on the peak day so there must not be very many videos in cell C. So campaign videos in the 2008 presidential election on YouTube virtually all fall into cell D.

I am going to agree with Crane and Sornette that the driver in the dynamics is, as they call it, exogenous. The campaign puts the video on YouTube, and people look at it. There is no build up. However, they ignore both interdependencies that can occur in a campaign and they ignore other exogenous shocks that influence the views of campaign videos.

The data I will analyze are the daily views for the videos the McCain and Obama campaigns placed on YouTube between July 1 and the election. At the end of each day videos posted that day were added to the datafile and the views of that day for each of the videos was added to the data collection. The data is available online at YouTube Campaign Video Stats (YouTube Campaign Video Stats, 2008). The two campaigns placed 884 videos on YouTube during the 4+ months. I also collected the daily views of the Bob Barr campaign videos, but they will not be used in this analysis. I will use only the time series posted between July 1 and October 4 allowing a minimum of 30 days for the time series to develop. That gives a data collection for analysis of 415 Obama campaign videos and 81 McCain campaign videos.

The process goes something like this. The candidates establish a channel on YouTube and start posting videos. Supporters and opponents stop by in a fairly steady stream. When they get there they are confronted by a page of 20 thumbnail images taken from the videos and video names that serve in lieu of a description. There is a video called Barack Obama at Martinsville, Virginia. Unless they are from Martinsville or Virginia they are not likely to look at that. If you are from Martinsville you may have attended the meeting, and you may go to YouTube to see again what you have already experienced. There is another video titled Barack Obama at Berlin. Many viewers would have read about Obama going to Berlin and that being a big deal. So many decide to check out the video. And they go through the list selecting or rejecting. In a few days they return. Some of the twenty on the page they had already reviewed. Unless there was one they particularly liked they only look at the new videos. Some they pass over and some they view. And a few days later . . .

To make sense of this process we need an indication of the breadth of appeal of a given video. This is an estimate of the potential audience for the video; Martinsville versus Berlin, for example. A second assumption is that the longer a video has been posted to YouTube the greater the proportion of the potential audience that has seen it. If the pattern is very regular there will be many views the first few days the video is on YouTube then each day there will be fewer views than the day before.

Notice this is the same logic as Crane and Sornette with Cell D. There is an exogenous force that acts as a constant, and there is a decay process from the first impact of the exogenous force. I have characterized the breadth of the appeal of the campaigns' videos as the exogenous force and exhausting the audience for the video as the decay process.

It is possible to use a relatively simple model for the process. I will develop it as a discrete time model since the data is, as gathered, discrete time.

$$V(t) = aV(t-1) + bU(t)$$

The total number of views for a video at time t is equal to a coefficient times views at $t-1$ plus an input that is assumed to be constant, $U(t)$.

Political scientists are not generally used to thinking about a constant driving a process of change, but that is precisely what you have in cell D in the Crane and Sornette model and what is hypothesized here.

$V(t)$ is a cumulative total. The “ a ” term in $aV(t-1)$ lets me model how quickly or slowly the potential audience is “exhausted”.

The most important difference between the Crane and Sornette study and this one is that they “lump” all of the videos together in their analysis. I will use the model to analyze each time series individually, and that permits looking at differences in the estimates of “ a ” and “ b ” that can then be further analyzed, i.e., look for specific campaign effects that one could not identify using the Crane and Sornette strategy.

The analysis of the model is straightforward (Boynton, 1980). For $0 < a < 1$ the system is stable, and approaches an asymptote. For larger values of “ a ” the progress to the equilibrium is slower, and for smaller values the progress to the asymptote is faster.

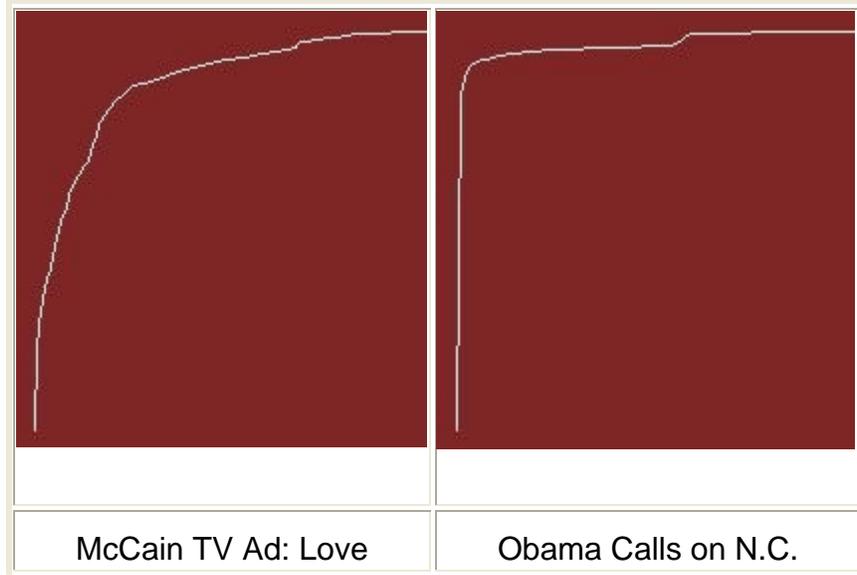


Figure 6: "a" coefficients

The "a" coefficient for the McCain TV ad "Love" is .91, and it approached the asymptote much more slowly than did the Obama video that had an "a" coefficient of .13. The visual difference is the gentle slope to the asymptote for the McCain video and the sharp corner for the Obama video. The numeric differences can be assessed by looking at the views after the first three days relative to the number of views as they approach the asymptote. Ninety-one percent of the views occurred in the first three days for the Obama campaign whereas only fifty percent of the views had occurred in the first three days for the McCain video. All but 6 of the 496 videos have an "a" that falls in this range; virtually all of the videos are stable systems.

For a > 1 the systems is not stable; it is flying off on an unending climb. There are 6 videos with this character. Three of the six are something like training videos for the campaign. "Neighbor to Neighbor How-To" is an example. It is a minute and a half instruction video about contacting voters. It tells where you can get a list of names, a map, and a script that you can use talking to the voters. It was viewed 43,000+ times. The plot of views shows a sharp turn at November 5; the views fell from 5443 on November 3 to 9 on November 6. From October 1, when it was posted, through November 4 the system was unstable; those are the data used estimating the coefficients. But almost nothing in the world is unstable forever, and this video stopped its "endless" climb when it was no longer relevant. The other five are very similar though a couple do not have quite as sharp a turn as in this example.

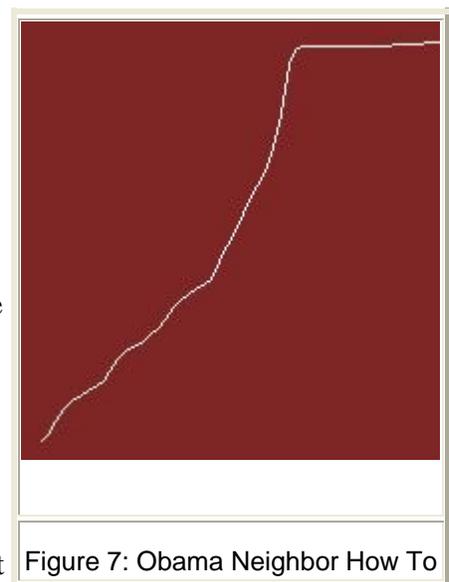
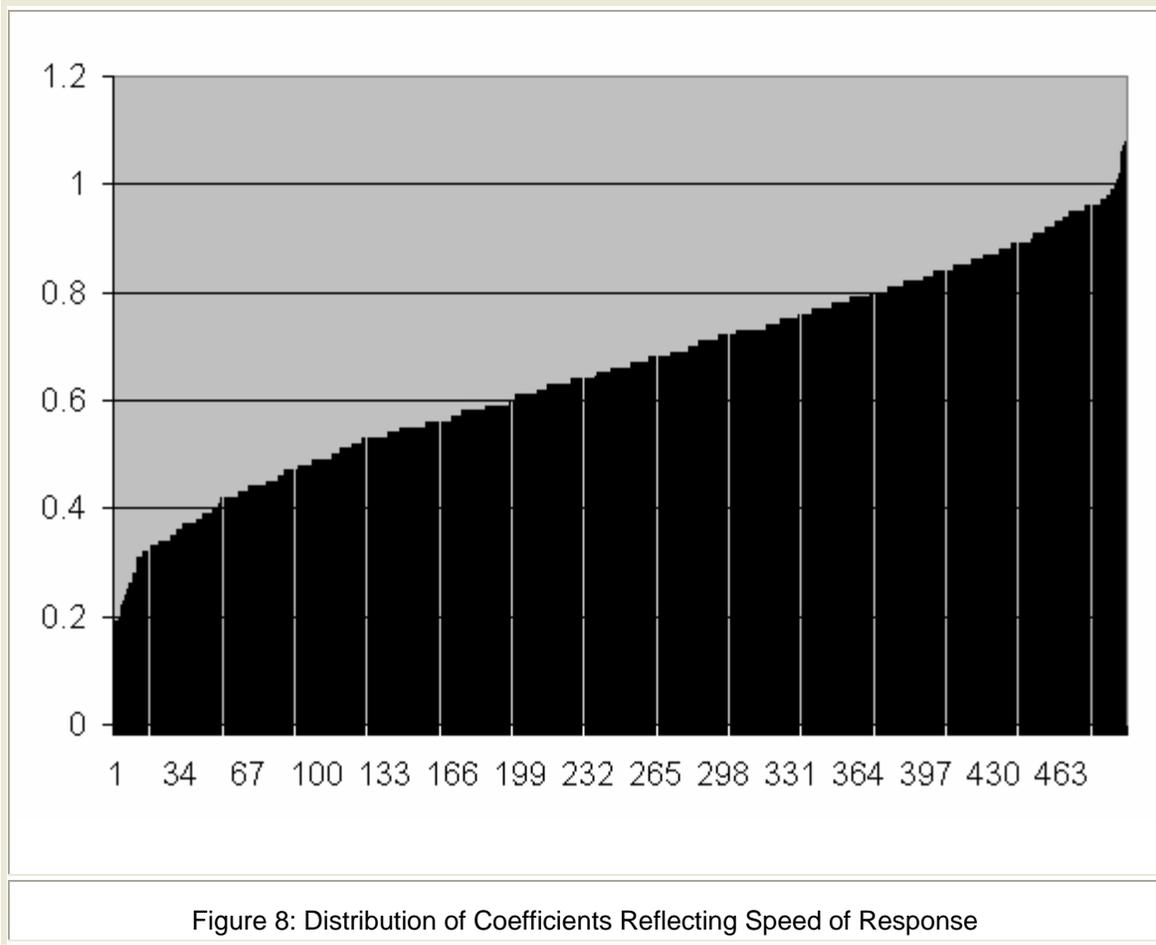


Figure 7: Obama Neighbor How To

When you sort the "a" coefficients by size you get this distribution.



The low is the .12 of the "Obama calls on North Carolina" video, and the high is just over 1.0. This spread from rapidly to slowly responding systems is something that cannot be assessed using the procedures of the Tubemogul or Crane and Sornette research. The first point is there are quite big differences in the videos with a pretty even distribution over the entire range of values. There is a lot of potential for investigation and explanation here. I will use these differences in examining campaign effects.

Campaign Effects

There is a dominant pattern. If a single pattern was all one was looking for then lumping all of the videos together for analysis would give that dominant pattern. But there is more to the story than a single pattern. Another feature of the collection of videos is what they are because they are part of a campaign rather than being independent of each other. At regular intervals the campaigns posted videos to their channels. As the campaigns progressed the postings came thick and fast (Boynton, September 27, 2008). The YouTube display is set up so you see the most recently posted videos on the first page and have to dig through more pages to get to videos posted earlier.

If campaign videos are being found by going to the channels, and viewers are more likely to view the most recent videos rather than looking through six or seven pages then viewing should decrease more quickly than for videos that are independent of each other.

There are two ways to assess this interpretation of the campaigns. One is to compare the time taken to reach a certain proportion of the views in the Tubemogul study with these data. In the Tubemogul study they found that 50 percent of the viewing happened during the first two weeks the videos were posted. Crane and Sornette do not provide comparable information so the comparison can only be to Tubemogul. I counted the number of videos in which half of the viewing fell within the first week or one half the time of the Tubemogul study. The numbers are shown in Table 1

| | McCain | Obama |
|---------------------|---------------|--------------|
| In one week or less | 87% (71) | 95% (390) |
| More than one week | 13% (9) | 5% (19) |

Table 1: Number of Videos by Time

The videos of the political campaigns reached half their views considerably more quickly than did the videos examined by Tubemogul. While there is a difference between the McCain campaigns and the Obama campaigns it is a very modest difference.

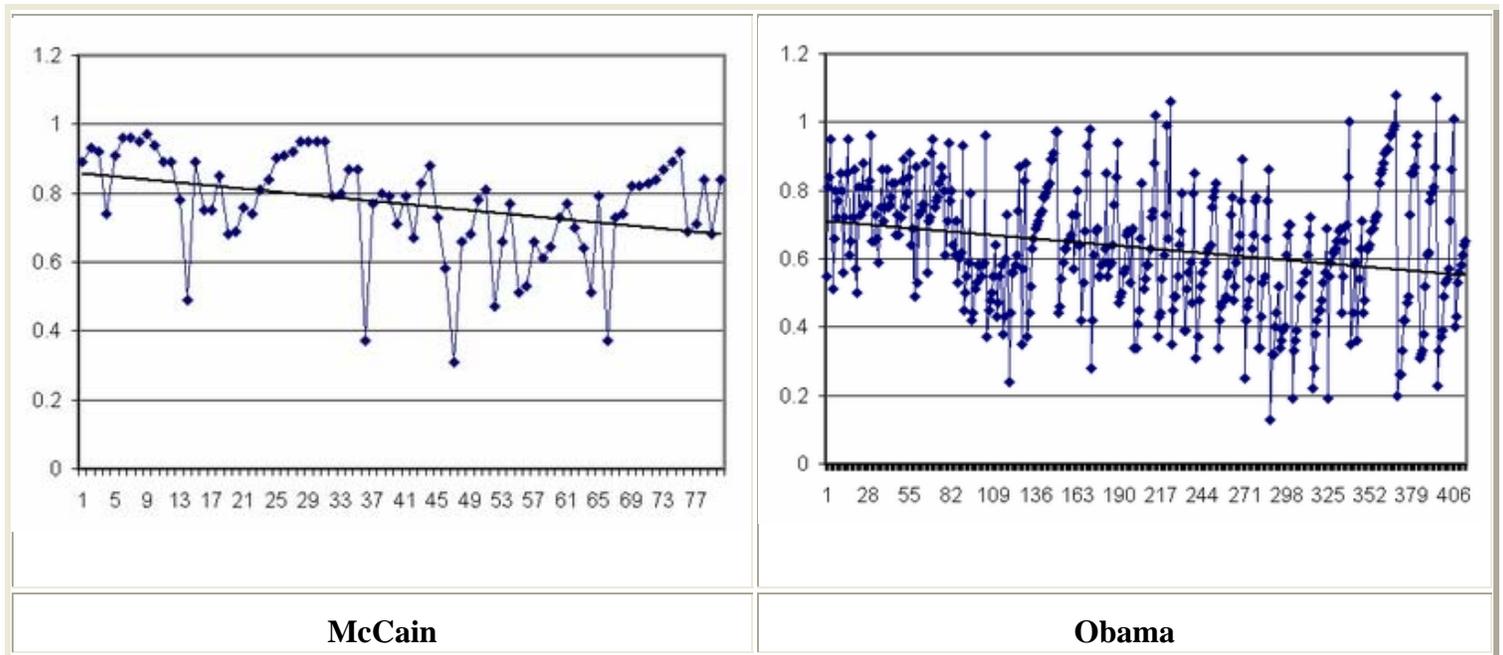


Figure 9: Speed at which Videos Approach the Asymptote

A second way to examine the importance of the compression that campaigns produce is to look at the speed at which the videos approach the asymptote. As the videos appear more and

more quickly during the campaign that should mean that people who look at the top of the page have missed several “rows” of videos that were at the top and then fell to the bottom of the page or to another page in between visits to the YouTube channel. If the campaign videos are “going away” faster then the coefficient “a” should fall during the campaign.

The videos are arrayed from left, which are the earliest, to right, which are the last posted to YouTube. That there are many more videos from the Obama campaign than the McCain campaign is clearly visualized in these figures. There is also much variability around the trend, but there is definitely a trend. The coefficients of the McCain campaign videos are a “step” higher than the Obama videos but the slope or trend is almost the same.

Another potential campaign effect is interaction between videos producing an increase of viewing earlier videos because of the posting of later videos. An easy to spot example happened at the end of July as the McCain campaign “went on a tear” posting 8 made-for-TV videos that were strongly critical of Obama in a couple of weeks.

| Date | Title | Total day 3 | Total day 10 | a | b | r-square |
|-------|----------------------------|-------------|--------------|-----|--------|----------|
| 07/16 | Wrong on Iraq | 3332 | 23656 | .94 | 1507 | .99 |
| 07/17 | Iraq documentary | 92813 | 224261 | .88 | 28768 | .99 |
| 07/18 | Troop funding | 88861 | 166826 | .89 | 18895 | .99 |
| 07/21 | Pump TV ad | 139135 | 238021 | .77 | 55407 | .95 |
| 07/22 | Obama Love | 255000 | | | | |
| 07/26 | Troops | 353210 | 439243 | .49 | 226997 | .97 |
| 07/30 | Celeb | 1339267 | 1939667 | .74 | 543678 | .97 |
| 08/01 | The one | 728772 | 1172257 | .83 | 244091 | .98 |
| 08/01 | Obama forgot Latin America | 45609 | 60721 | .66 | 21515 | .98 |
| 08/04 | calls on congress | 5882 | 8648 | .67 | 3190 | .98 |

Table 2: Interaction between Videos

The videos referenced in the top eight rows are the videos used to examine interaction. The bottom two rows are videos used to show that the interaction had stopped by that point.

One interaction that appears to be present is a mobilizing effect. The first video, "Wrong on Iraq" was viewed only 3,332 times in the first three days it was posted. Videos 2 and 3

jumped to 30 times as many views in the same number of days. "Pump" TV ad, fourth, was viewed 1.5 times the number viewing 2 and 3. "Obama Love," fifth in the series, jumped to almost 2 times the views of 4. A little explanation is needed for "Obama Love." Obama Love was two videos that were posted on YouTube at the same time. They had the same video and narration, but a different musical background. The McCain campaign sent an email to supporters inviting them to vote on which they liked best. Unfortunately for the McCain campaign, they had not checked with the company owning the copyright for the music. When the company found out their music was used they demanded that it not be used. So the videos were viewed approximately 255,000 times and then taken off of YouTube. Even though they were there only three days they were an important part of the series. After "Obama Love" they posted "Troops," which was viewed almost twice as many times in the first three days as the preceding video. "Celeb" was a huge jump in views. "The One" was not viewed as often as "Celeb" but it was viewed many more times than all of the other earlier videos.

I believe a plausible argument can be made for a mobilizing interaction in this series. One, there is a consistently increasing number of views with each new video; until you get to "The One" an exponential curve fits reasonably well. Two, an argument that it was the content that led to the increase seems unlikely because a video about the price of gas sits in the middle of a series of videos criticizing Obama's support for the troops in Iraq. "Wrong on Iraq" was viewed 3,332 times, "Iraq Documentary" 92,813 times, "Troop Funding" 88,861 times, "Pump" 139,135 times, and "Troops" 353,210. If it was support for troops versus gas prices then the four videos concerned with support for the troops should not range in views from 3,332 to 353,210. And it is hard to believe that the people viewing Pump were likely to think it 41 times more important than "Wrong on Iraq;" it seems more likely that they just had not been drawn in before. It is a plausibility argument, but it seems pretty plausible

However, there are interactions in this series that have more supporting evidence. The dominant pattern of a decreasing number of views each day is pictured on the left of this figure. If there is a deviation from that pattern it will show up like the right hand side of the figure. There was a single day when the views increased rather than decreased.

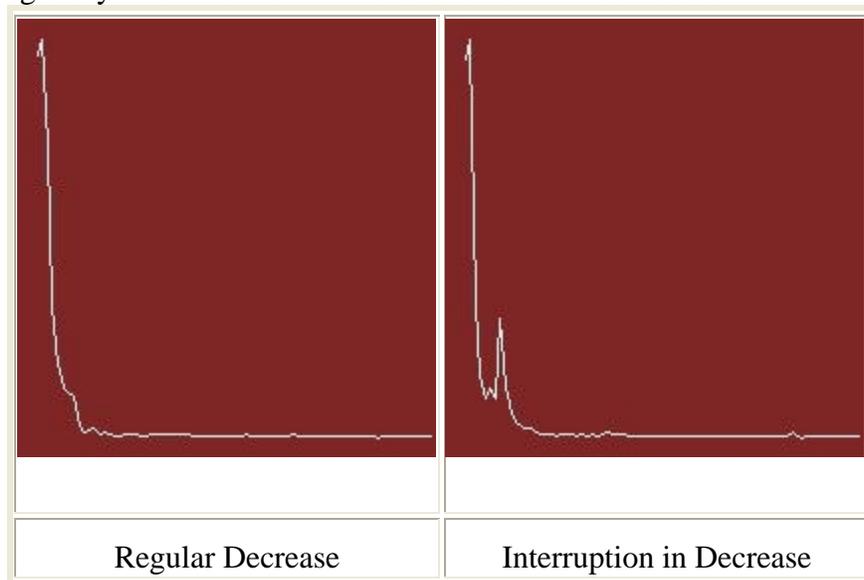


Figure 11: Decreasing Number of Views

reached. However, as just shown there are YouTube campaign effects that also have an impact on the viewing. And the YouTube campaigns do not exist in a vacuum; they were not completely closed systems. The first point is to demonstrate that exogenous effects are there and can be recognized.

When the Democrats met in their nominating convention the views of McCain campaign videos got a “bump.” The debates also produced bumps in views of videos. And many videos saw an increase in views as the election approached. But I would first like to look at two episodes that so clearly reflect both the stability of a constant driving the views and an exogenous factor that “interrupts” the time series that it will be completely clear that this is what is going on.

In the middle of July 2008 Obama campaigned in Indiana with Evan Bayh, senator from that state (UPI.COM, 2008). Bayh made a statement supporting Obama that the campaign placed on YouTube. There was speculation at the time that Obama might be considering Bayh as his vice presidential running mate. And when Obama returned to the state three weeks later the speculation became intense (Hulig News, 2008). Suddenly people needed to find out about Evan Bayh and his statement on YouTube was one place they turned. The result was a sudden, very sharp jump in views that cannot be missed in the figure. Clearly this is an exception to marginally decreasing daily. And it is not difficult to spot what brought it about. Obama “came to town” for a second time, lots of people began searching for the Senator from Indiana, and they found him.

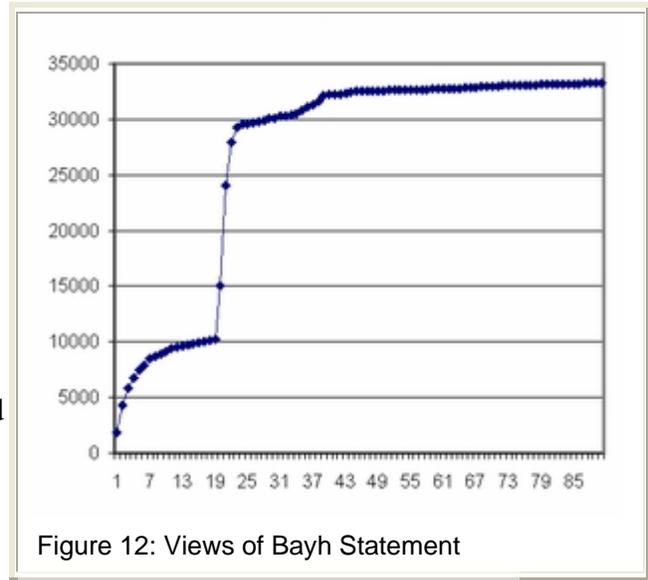


Figure 12: Views of Bayh Statement

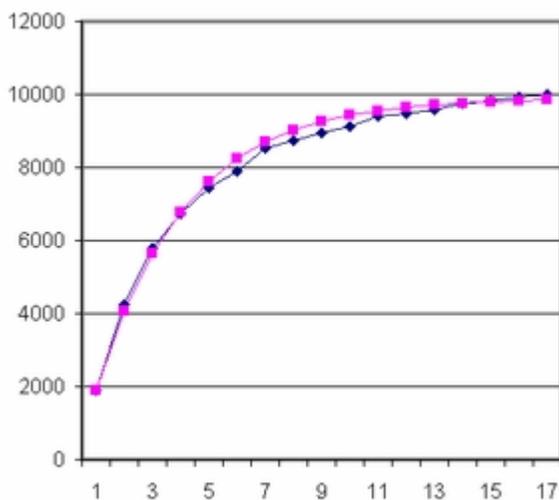


Figure 13: Actual and Expected Values

It is clear that the system describing the first few days of viewing the Bayh video could not describe the entire process; the system was altered by an exogenous event. It is possible to assess the impact of that shock on the eventual total number of views. Do the short term system and the post “shock” system begin to converge? If not what is the size of the gap?

The system description of the first seventeen days is $V(t) = .73V(t-1) + 2674$. The actual distribution and the distribution predicted by the system equation are shown in this figure, and the fit is extremely good. The r-square for this relationship is .99. Given the extremely good match between actual and expected it is plausible to project the system

ahead to the same number of days for which there is actual data, which is 90 days. There was no convergence. At the end of 17 days this system was projecting 9,800 views and by 90 days it reached 9,900 views. That makes the gap very considerable. By 90 days the actual views had risen to 33,243. So the difference made by the speculation about Bayh's possible candidacy was 23,000 views.

The second example is a post election example; old campaign videos never die they just fade away, as the saying goes. On October 15 the Obama campaign posted a video featuring Valerie Jarrett -- "Valerie Jarrett's Thoughts on Barack Obama." She had been a friend of the Obamas for decades. She was also traveling with Obama on many of his trips around the country with carefully chosen responsibilities in the campaign. The video, pictured by the left figure, looked very standard: $V(t) = .78V(t-1) + 6682$, and r-square was .98. Those are standard numbers and a standard path apparently approaching an asymptote.

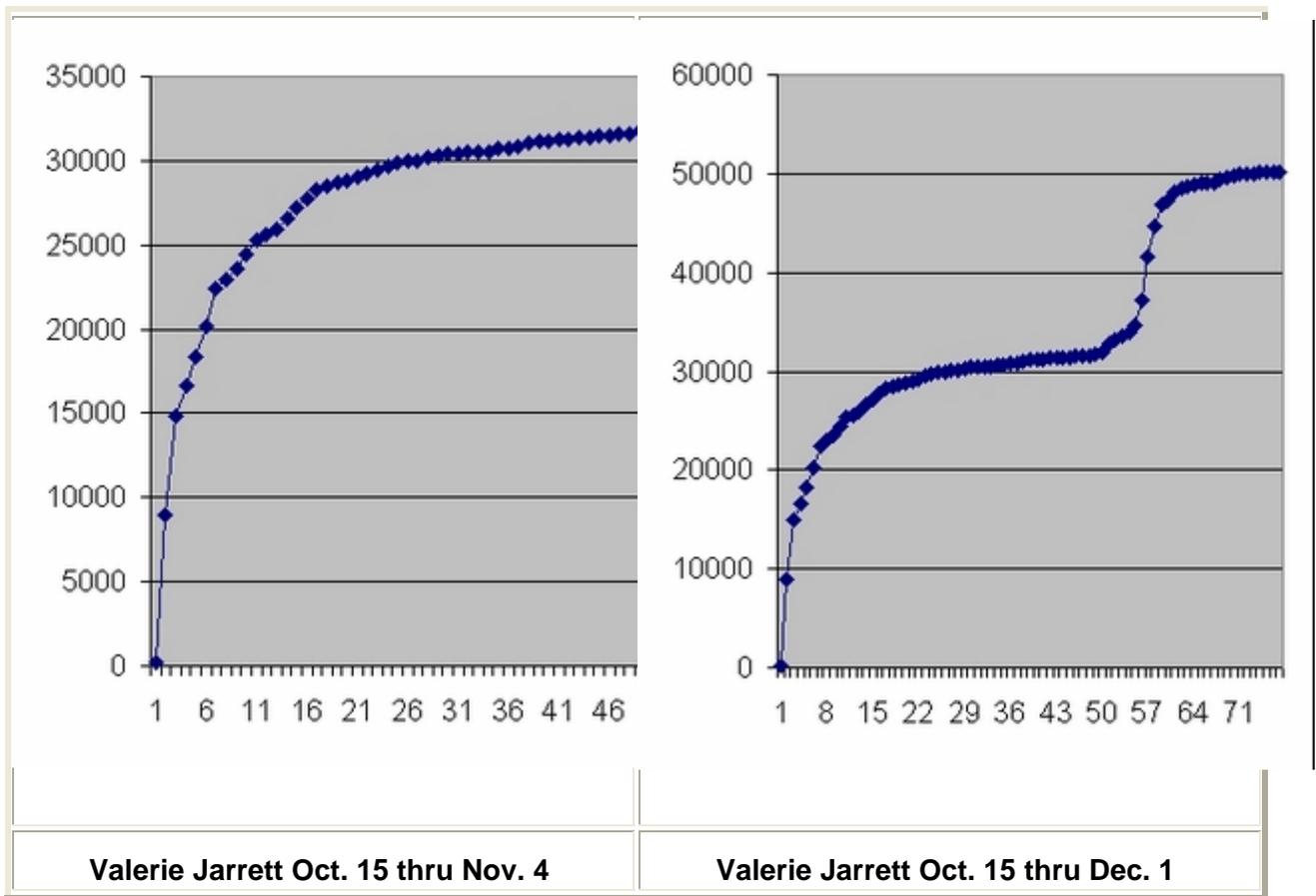


Figure 14: Valerie Jarrett Videos

Then the election, and Ms. Jarrett moved out of the shadows. The New York Times had a story about her on November 5. Time magazine followed on November 8, she was on The News Hour with James Lehrer on the 13th, The Washington Post had a story on the 14th, and The Chicago Sun Times had a story on the 15th. She had become co-chair of the Obama transition team and senior advisor to the president. And everyone was searching the web to learn about Valerie Jarrett. One of the results of the searching was the campaign video. Through November 4

it had been viewed 31,895 times. By December 1 it had been viewed 50,210 times. Twenty thousand additional views to learn about the “new” person with an important position in the Washington scene.

The point is two examples in which both the constant and the interrupt are apparent in the movement of viewing through time. In both cases the exogenous factor made quite a large difference in the views of the campaign videos.

The one “exogenous” factor to which all campaign videos were subject is the election itself. As the election approached there was increased interest and it is easy to imagine people turning to YouTube in greater numbers than before producing a “bump” in viewing. To examine this possibility systematically I looked at the views for each video for two three day periods -- ten days before the election, October 22, 23 and 24, and the three days November 2, 3, and 4. The views for the two three day periods were summed, and I computed the ratio of the immediate election period to the period ten days earlier.

As in other analyses, only videos posted before October 5 are considered. By the election they might be characterized as “spent” systems. Virtually all were characterized by the dominant pattern, and, thus, were approaching an asymptote by the time of the election. So the numbers were generally very small. Only about five percent of the Obama videos summed to more than 1,000 views for the November 2-4 period. On average the McCain campaign videos were viewed much more frequently than the Obama videos and this is reflected in twenty percent of them summing to 1,000 views November 2-4. So many “bumps” were very small compared to the total number of views received by the videos. For most videos there was nothing quite like the impact of the exogenous influence seen in the Bayh or Jarrett videos.

The set is 413 videos of the Obama campaign and 82 videos of the McCain campaign.

| Campaign | No Bump | Bump | Total |
|-----------------|----------------|-------------|--------------|
| McCain | 28% (23) | 72% (59) | 82 |
| Obama | 37.5% (155) | 62.5% (258) | 413 |
| Total | 36% (178) | 64% (317) | 495 |

Table 4: Viewing “Bumps”

Sixty-four percent of the videos were viewed more often the last three days of the campaign than they had been viewed ten days earlier. There is modest difference between the two campaigns; more McCain videos saw an increase in views than did Obama videos. However, the difference is only ten percent. It is clear that the approaching election day produced a bump that should be interpreted as exogenous to the basic structure of viewing campaign videos on YouTube.

What was not there was a temporal trend. When the ratio is plotted by date of first post to YouTube is no discernible trend of any sort. That suggests that in the final days the search used by viewers was not simply looking at what was posted that day. An obvious search procedure would be to search for names of the relevant political figures. I searched the titles of the videos by name and counted videos that increased and decreased in the last three days.

| Person | Views Down | Views Up | Total |
|-----------------|------------|-----------|-------|
| Barack Obama | 25% (34) | 75% (102) | 136 |
| Michelle Obama | 63% (12) | 37% (97) | 19 |
| Joe Biden | 67% (24) | 33% (12) | 36 |
| Jill Biden | 75% (3) | 25% (1) | 4 |
| Bill Clinton | 0 | 100% (3) | 3 |
| Hillary Clinton | 60% (3) | 40% (2) | 5 |
| David Plouffe | 40% (2) | 60% (3) | 5 |
| John McCain | 8% (1) | 92% (11) | 12 |
| Sarah Palin | 50% (1) | 50% (1) | 2 |
| Total | 36% (80) | 64% (142) | 222 |

Table 5: Viewing Increases and Decreases by Person

The titles of forty-five percent of the videos contained the name of one of these persons. It is obvious that the strategy of the McCain campaign was different, with respect to naming, from the Obama campaign. Over the entire set the number of videos that went up and down matches the distribution for the entire 495 videos. Within that overall count there is some variation. The “winners” are Bill Clinton, John McCain and Barack Obama -- in that order. Searching for the candidates by name results in a modest improvement over the ratio for all the videos. But searching by name does not seem to have been a dominant strategy for searching or the increases for the other persons would have been greater.

Conclusions

No going viral. Going viral is a standard idea about how communication happens on the web. It has a structure that many political scientists could find compatible -- there is something that is changing and producing changes in views. The report begins by reviewing three possible constructions of “going viral” and argues that none is a good account of what happened in the YouTube campaign for the presidency in 2008.

I have suggested a different construction. It is the video that is driving the views. Of course, that is incomplete. You need an election and people who are interested in following the campaign. You need a network that can carry the files. You need a technology to aggregate and distribute the videos. You need a population that is comfortable with viewing video online. You need an organization such as channels to simplify the discovery process. In the time frame of an

election these are all constants though we usually take them for granted and look for something that is changing to explain the dynamics of viewing.

There is a dominant pattern to the dynamics of viewing campaign videos on YouTube. It is exceedingly regular, as represented by the goodness of fit of the model (see appendix for further explication of this point). That does not mean they are all alike. There is variation in the breadth of appeal of videos and in the speed with which the potential audience finds them. But it means this is the starting point for analysis. Looking at the impact of an exogenous factor starts with the dominant model. Then one can assess the impact of the exogenous factor as the examples of Bayh and Jarrett illustrate.

The most important point here is 62 million views between July 1 and the election. That is a lot of political behavior we can analyze in many different and interesting ways. It is also a portent of things to come. The communication technology and economic organization that are the basis for this political action is changing very rapidly. Four years from now the web will be as different from today as today is from four years ago. We can be confident, however, that politicians will find the ability to communicate with their constituencies more effectively than in the past something they will not give up. This analysis of the dynamics of viewing lays the base for many types of analysis. It suggests that one line of thinking is not productive except in extraordinary situations. And it becomes the starting point for many additional explorations of this political behavior that is newly available to us. Next we must ask: What if you had a choice? What do citizens choose when given a choice in ways that has not been possible before?

Methodological Appendix: Procedures

The procedures I use in analyzing the dynamics of viewing campaign videos are somewhat unusual so they need some explication and justification, which I will do here.

It is what you want to learn

The basic theory posits that viewing will take the form of a smoothly decreasing number of views each day. However, I am not so much interested in the daily views as I am in the total views. How many times was a McCain video called Taxman viewed? The McCain campaign posted it on YouTube on August 15. How many times was it viewed, and what was the dynamics of getting to that total? By election day the video had been viewed 114,379 times. If you plot both daily cumulative total and daily views you can see that the two views of the dynamics reflect each other.

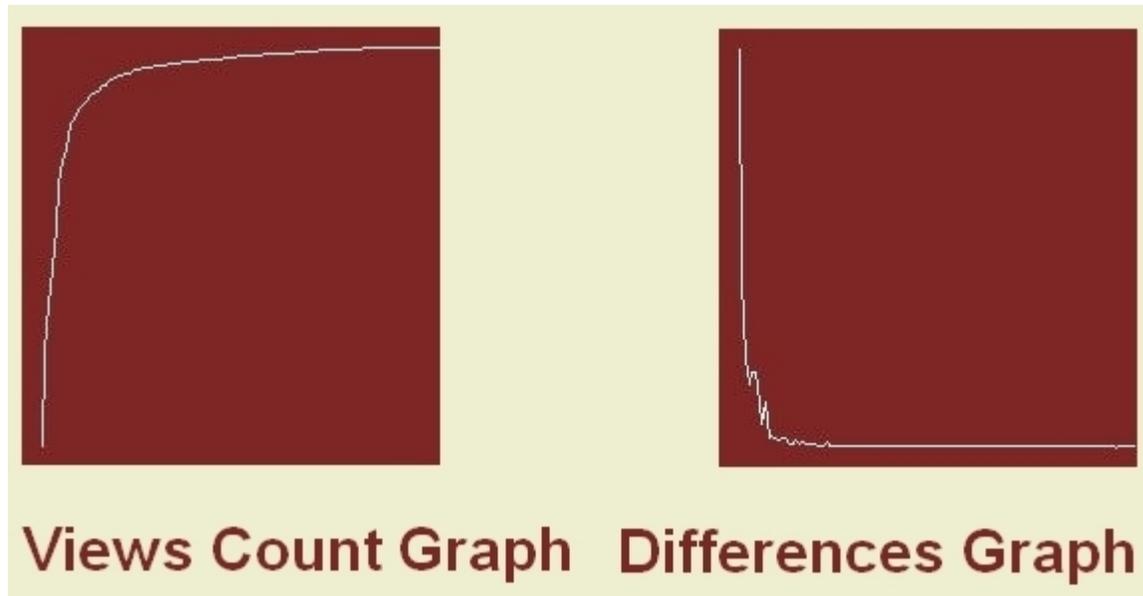


Figure 15: Daily Cumulative Totals and Number of Views per Day

The daily cumulative total is plotted in the left figure and the number of views each day is plotted on the right. All you have to do is “turn them over” to get from one to the other. Since my interest is in total views the model of the theory is about total views -- the left figure.

However, the figure on the right shows jagged lines at the bottom left and the comparable spot on the total views seems smooth. What should be made of that? In this case a third view of the dynamics suggests a perspective on the difference.

The equation is $V(t) = aV(t-1) + bU(t)$. The coefficients a and b can be estimated with regression. Figure 16 shows the residuals from that regression.

This plots the data points on the regression line. It is clear that the residuals are exceedingly modest. Based on the regression $a = .80$, $b = 22042$ and r -square is $.99$.

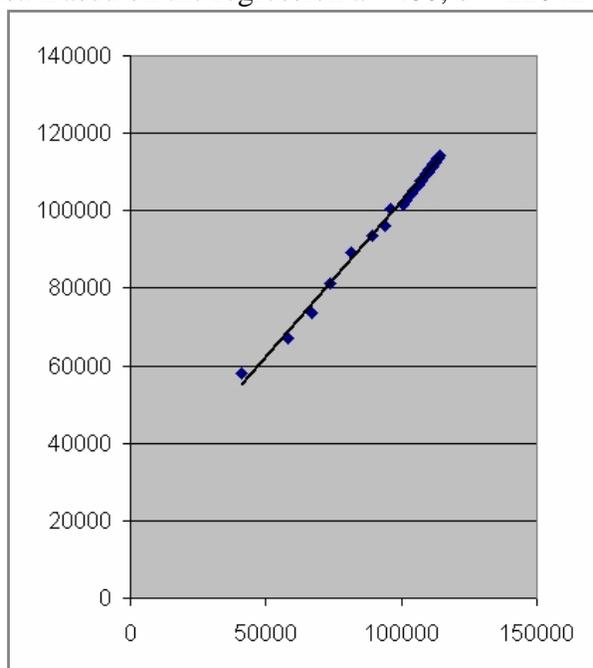


Figure 16: $V(t) = aV(t-1) + bU(t)$

The cumulative total cannot decrease

The standard political science interpretation of r-square is “explained variance.” While I use r-square I do not treat it as explained variance. Instead I understand it as a measure of goodness of fit. How much or how little deviation from the expected is found.

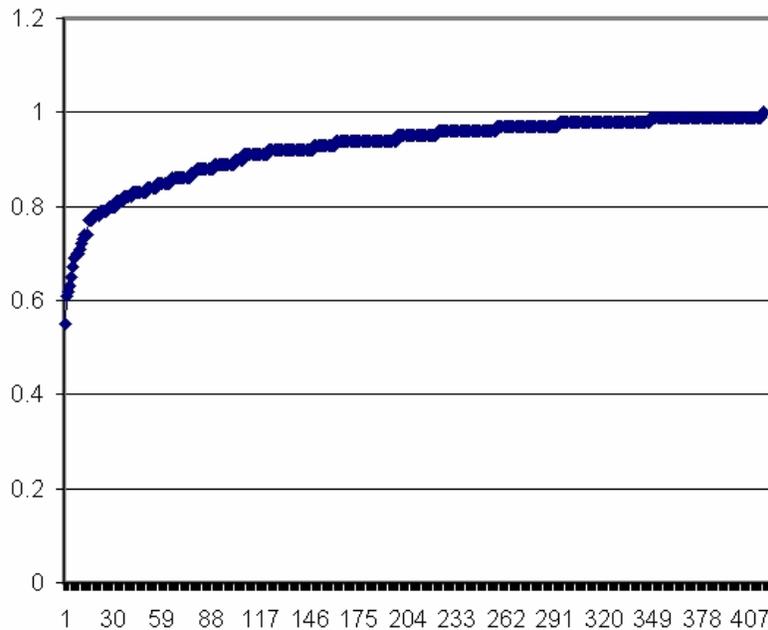


Figure 17: Plot Value of r-square for Obama's Videos

There is, however, a wrinkle when using regression with cumulative totals. A cumulative total cannot decrease; it can stay the same and it can increase. That has a considerable biasing effect on the size of r-square.

The dynamics are very regular for the Obama videos; there is very little deviation from expected. Most of the more than 400 r-squares are above .9. However, they fall as low as .55, which suggests that it would be possible to have much lower numbers if the dynamics were less regular.

In addition, r-square is not the final word when it comes to examining deviation from expected. There are a number of videos that have very high r-squares that do not fit the expected curve very well. Those deviations can be spotted more easily by looking at the plots and the residuals than by looking at the value of r-square.

Measurement error

Every day in each of the time series has 24 hours other than the first day. The procedure was to check the number of views each day at midnight. A video could have been posted to

YouTube only minutes before midnight, and the views would have been the result of only that few minutes. Other videos might have been posted early in the day giving them an entire day of viewing. Some are easy to spot. For example, the Obama campaign posted "Barack Obama on the Economy in Elko, NV" on September 10 and the views that day were 258. The next day the video was viewed 8570 times. It seems very likely that it was posted late on the tenth otherwise the disparity between day 1 and day 2 is hard to account for. The same can be true even for days when the number of views the first day are well over 200 if the next day sees a very big jump in views.

| | Day 1 thru Nov. 4 | Day 2 thru Nov. 4 |
|------|-------------------|-------------------|
| a | .34 | .70 |
| b | 8069 | 3668 |
| r-sq | .82 | .95 |

Table 6: Time Series Measurement Error

The problem with this measurement error is what it does to the estimates of the coefficients a and b. When I start the computation with the second day instead of the first you get a very big change in the coefficients. The fit, assessed with r-square, has gotten much better if you leave out the first day of the series. More important is what it does to a and b because these are the nub of the theory. The size of a indicates how quickly the system is going to the asymptote; if it is .34 that is very fast and if it is .7 that is much slower. And b is an indication of the constant that is driving the dynamics. It changes dramatically from 8069 to 3668.

I do not have a solution for this because the "obvious" measurement error may supplement dynamics that are going faster than predicted by the theory. Sorting into measurement error and rapid system change is not easily done. In this paper I have stayed with the original data; day 1 is in. When I want to look at the b coefficient more closely I will probably have to make an adjustment.

Sampling in the Time Domain

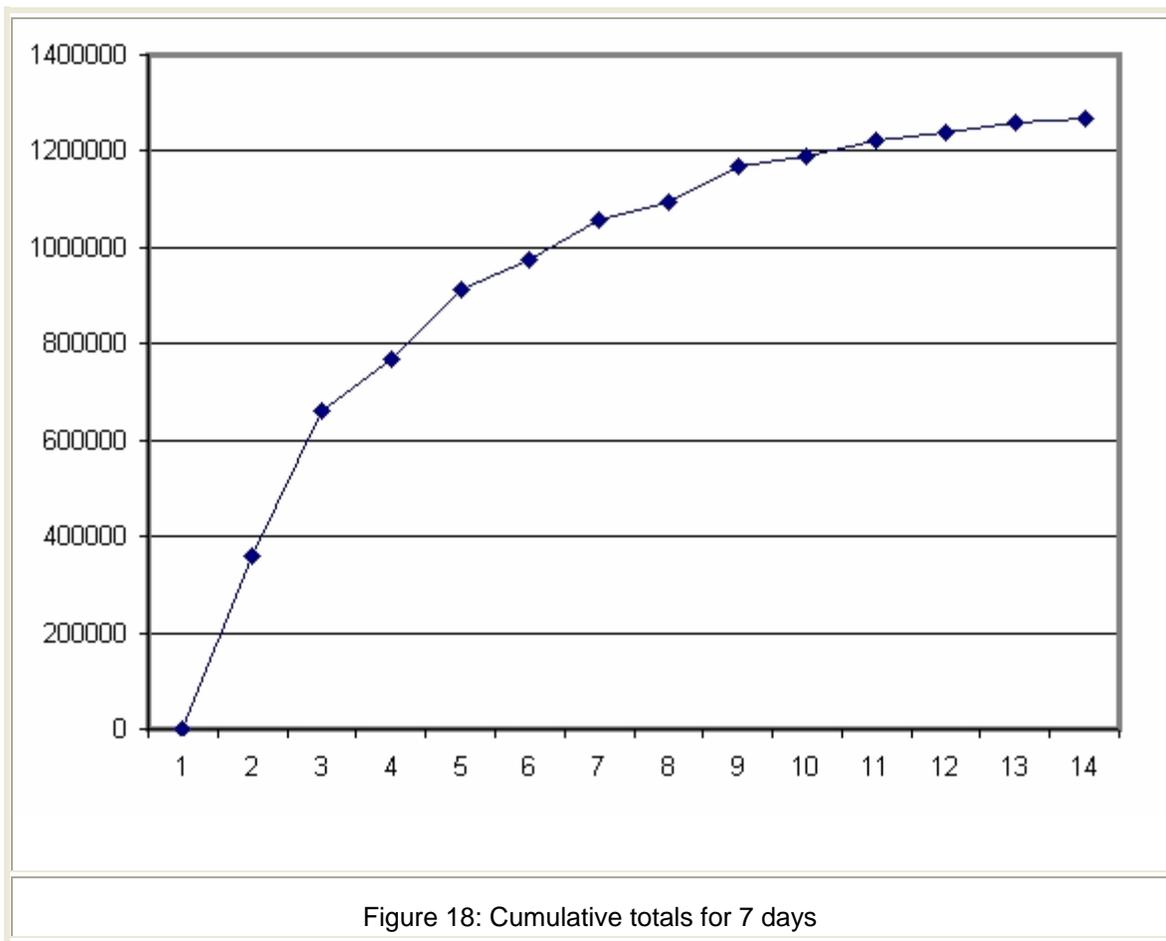
When the process being investigated is changing in time an important question to ask is how frequently to sample, that is, record information about the process. The answer is straightforward -- it depends. It depends on the variation in the process being studied. If you sample too frequently you get either redundancy or variation you are not interested in. If you sample too infrequently you miss variation you are interested in. To decide how frequently to sample you have to learn about the dynamics of the process.

In examining the dynamics of viewing campaign videos on YouTube the assumption was that once a day was sufficient. The counts were collected each night just after midnight. This assumption was not based on careful research, however. So I looked for an opportunity to do data collection more frequently than once a day to determine what variation, if any, was missed when sampling once a day.

The opportunity presented itself in the form of a very public conflict between Jon Stewart, of the Daily Show, and CNBC reporter Jim Cramer. Jon Stewart mocked Cramer's reporting on the economy. After a lot of back and forth Cramer was invited to visit the Daily Show, and he accepted. The confrontation, occurring on the evening of March 12, was touted far and wide in the online video domain. It looked like a good video to study; it seemed there would be enough interest that viewing would not decline to zero in one or two days.

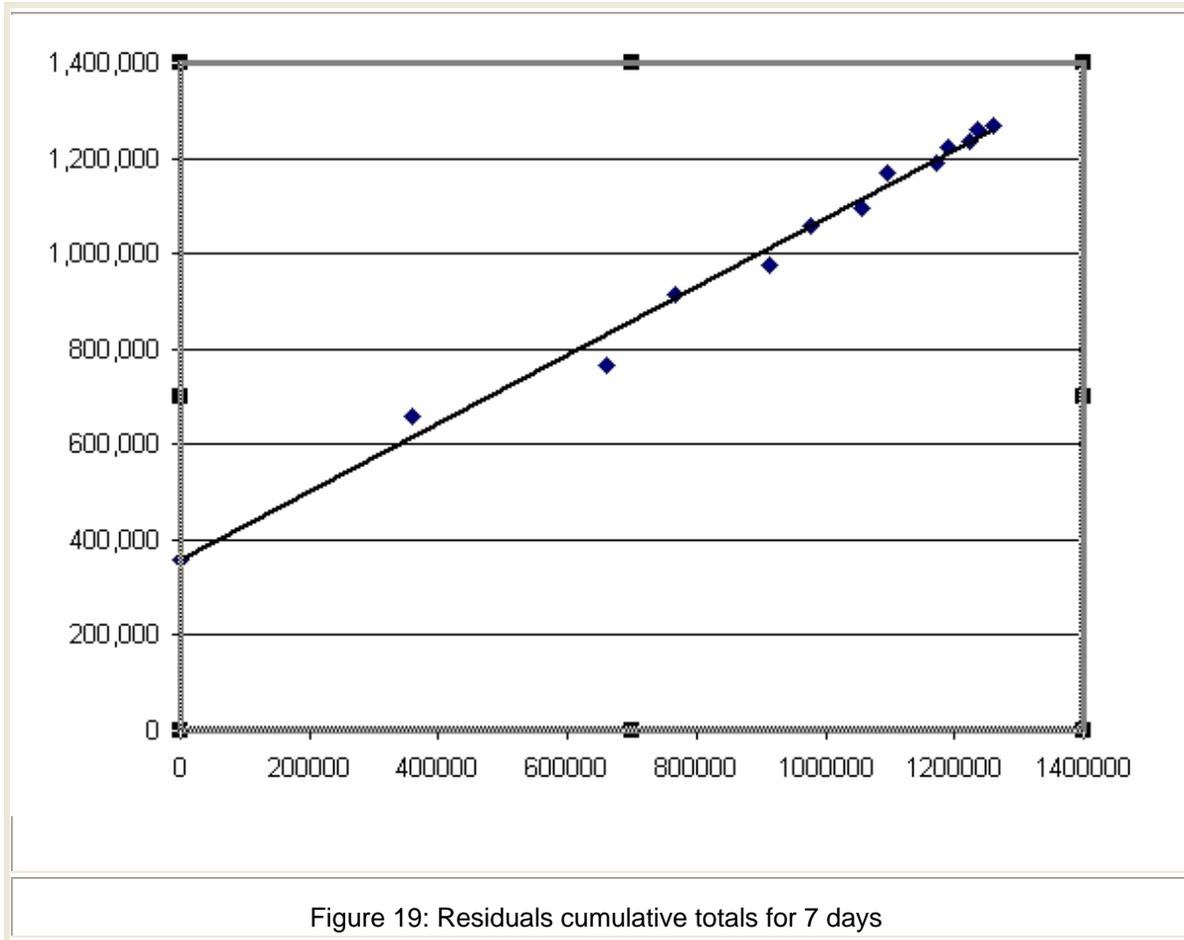
I collected the total number of views to that point each evening at 9:00 p.m. and morning at 9:00 a.m. -- with modest variation in the sampling points. The first number recorded was 9:00 a.m. March 13, which was the morning after the interview on The Daily Show.

Figure 18 displays the cumulative totals for each data point.



The first data point is 9:00 p.m. March 12 when there were no views. This increased to 1.2+ million by the end of the seventh day in what seemed to be a very regular pattern. When modelled as a first order difference equation the result was $y(t) = .7174y(t-1) + 356374$. The r square was .99

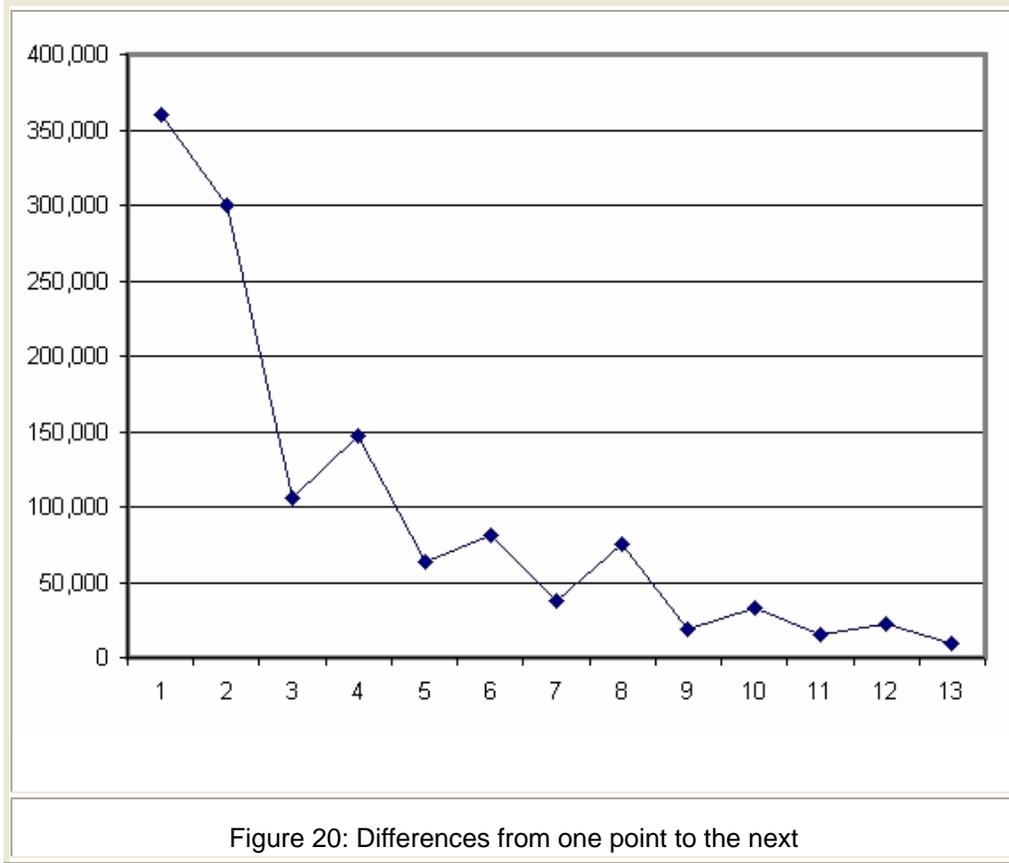
As one would expect there was very little residual variation.



But this is not the only way to examine the data. Figure 20 shows a rather different pattern. This is the difference in the number of views from one data point to the next.

The decrease in the number of views consistently falls over the time period, which is what the cumulative figure shows as well. They are the obverse of each other. However, there is a pattern in this figure that was less apparent in the display of total views over time. The even numbered data points are up and the odd numbered data points are down. This corresponds to the time of data collection. The odd numbered points were in the morning and the even numbered points were in the evening. The point is simple: given the general pattern of declining views, the video was viewed more often during the day than during the night.

When the parameters of the model were estimated with the full data set -- morning and evening -- the results were shown above. However, if you estimate the parameters with data once a day you get a rather different result. The model becomes $y(t) = .6935y(t-1) + 427496$, and the r-square is .979.



Then the situation becomes choosing between the two. The limit to which the process is going can be computed for both models. The limit for the two data points a day model is 1,261,054. The limit for the one a day model is 1,394,773. There is one more data point that was not included in the computations -- the evening of the seventh day. That value is 1,282,117. That exceeds the limit expected from the first model -- only one day later. However, it is still well under the limit expected using the one a day data. The cumulative number of views does not get to 1,390,000 until the evening of March 31, and the increase in views the next morning was 477.

If you are not interested in the day-night variation then dropping one observation each day is not a problem. You know you are missing variation, but it is not important and it would require a more complicated model to set things "straight." If you are interested in the day-night variation then a second order equation could be the appropriate way to model the process.

At least for views of Stewart versus Cramer a single data point a day seems sufficient. It seems plausible to consider the same may be true for viewing many political videos.

References

- Albrecht, C. (2008, September 24). NTV Rerun: What Constitutes an Online Hit? *NewTeeVee*. Retrieved from <http://newteevee.com/2008/09/24/ntv-rerun-what-constitutes-an-online-hit>
- Arneson, P.A. 2001 (Revised 2006). Plant Disease Epidemiology: Temporal Aspects. Retrieved from <http://www.apsnet.org/education/advancedplantpath/topics/Epidemiology/UpperLimit.htm>
- Boynton, G. R. (1980). *Mathematical Thinking about Politics: An Introduction to Discrete Time Systems*. Longman Publishing Group.
- Boynton, G. R. (2008, September 18). Going Viral in the YouTube 2008 Campaign. Retrieved from <http://cli.gs/08qSym>
- Boynton, G. R. (2008, September 27). The Pace Quickens. Retrieved from <http://youtube08election.pbwiki.com/The-Pace-Quickens>
- Boynton, G. R. (2008, November 1). Going local in a global domain. Retrieved from <http://cli.gs/DpBsT6>
- Boynton, G. R. (2008, December 14) The Ghosts of Campaigns Past and Campaigns Yet to Come. Retrieved from <http://dl.getdropbox.com/u/46529/02-text.html>
- Crane, R. and Sornette, D. (2008) .Robust dynamic classes revealed by measuring the response function of a social system. Retrieved from <http://www.pnas.org/content/105/41/15649.short>
- Gannes, L. (2008, December 1). Just Wow: YouTube had 344M Global Uniques in October. Retrieved from <http://newteevee.com/2008/12/01/just-wow-youtube-had-344m-global-uniques-in-oct>
- Gannes, L. (2008, December 29). Evolution of Dance 2 to Finally be Released in January. Retrieved from <http://newteevee.com/2008/12/29/evolution-of-dance-2-to-finally-be-released-in-january>
- Hulig News. (2008) Evan Bayh At The Heart Of Obama VP Buzz. Retrieved from <http://www.hulig.com/65954/evan-bayh-heart-obama-vp-buzz>
- The Pew Research Center for the People and the Press. (2008, December 23). Internet overtakes newspapers as news source. Retrieved from <http://people-press.org/report/479/internet-overtakes-newspapers-as-news-source>
- Shankland, S. (2008, May 13). YouTube ads for viral videos: “Buzz targeting.” Retrieved from http://news.cnet.com/8301-10784_3-9943548-7.html
- Tubemogul. (2008, June 19). Online Video's Short Shelf Life. Retrieved from <http://www.tubemogul.com/research/index.php?r=7>
- UPI.COM. (2008, July 17) Two VP potentials stump with Obama. Retrieved from http://www.upi.com/Top_News/2008/07/17/Two_VP_potentials_stump_with_Obama/UP-I-18071216299179/

Van Grove, J. (2009, January 4) Twitter App Database: Will You Use It? Retrieved from <http://mashable.com/2009/01/04/twitter-app-database>

Visible Measures. (2008, December 19). Bush vs. Shoes: Viral Video Wrap-up, <http://www.visiblemeasures.com/news-and-events/blog/bid/7698/Bush-vs-Shoes-Surpasses-10-Million-Viral-Video-Views>

Wallsten, K. (2008) “Yes We Can”: How Online Viewership, Blog Discussion and Mainstream Media Coverage Produced a Viral Video Phenomenon. Paper presented at the 2008 American Political Science Annual meeting.

YouTube Campaign Video Stats. (2008). Retrieved from <https://at-lamp.its.uiowa.edu/youtube>