Fifteen Minutes of Fame: The Power of Blogs in the Lifecycle of Viral Political Information

Karine Nahon, University of Washington
Jeff Hemsley, University of Washington
Shawn Walker, University of Washington
Muzammil Hussain, University of Washington
Fifteen Minutes of Fame: The Power of Blogs in the Lifecycle of Viral Political Information

Karine Nahon, University of Washington
Jeff Hemsley, University of Washington
Shawn Walker, University of Washington
Muzammil Hussain, University of Washington

Abstract

This empirical study addresses dynamics of viral information in the blogosphere, presenting a new methodology which enables the capture of dynamism and the time-factor of information diffusion in networks. Data was gathered on nearly 10,000 blogs and 13,000 blog posts, linking to 65 of the top U.S. presidential election videos that became viral on the Internet between March 2007 and June 2009. The article argues that the blogosphere is not monolithic and illuminates the role of four important blog types: elite, top-political, top-general and tail blogs. It creates a map of the ‘life cycle’ of blogs posting links to viral information. It shows that elite and top-general blogs ignite the virality process, which means that they get the chance to frame messages and influence agenda setting while top-political and tail blogs act as followers in the process.

KEYWORDS: virality, blogs, blogosphere, power, diffusion of information, political information, election, elite, networks, agenda setting, economometrics

Author Notes: This research could not have been done without the help of so many people on the retroV team. We are particularly grateful to Lance Bennett, Sheryl Day, Chris Wells, Akshay Bhagwatwar, Aleesha P. Hegde. We would like also to thank David Karpf for allowing us to use his index, and Hil Lyons whom we consulted with on our statistical methods. Finally, we are thankful to Google for bestowing this project the Google Research Award.
Introduction

The topic of information virality in networks is drawing increasing attention from scholars and practitioners from different fields, who seek to understand factors that influence the process of virality (Bardzell, Bardzell, and Pace 2008; Helm 2000; Palka, Pousttchi, and Wiedemann 2009). We define virality here as the process which gives any information item (picture, video, text, or any other audio–visual–textual artifact) the maximum exposure, relative to the potential audience, over a short duration, distributed by many nodes. Despite this growth of interest, the body of literature remains slim in theory development, empirical investigation, and an in-depth understanding, and is drawn mainly from five different fields: communication, political science, marketing, information science, and sociology (Boynton 2009; Helm 2000; Wallsten 2007; Barzilai-Nahon and Hemsley 2011). Theories of information diffusion pose two diametrically opposite approaches to viral information diffusion: one that suggests that virality is a process governed by reliance on powerful gatekeeping nodes or elites (Adamic and Glance 2005), and another arguing that it is a much more dynamic bottom-up process where gatekeepers may play an important, but not a crucial role (Herring et al. 2005). Work on political information diffusion and virality in the blogosphere is even scarcer (Barzilai-Nahon and Hemsley 2011; Boynton 2009; Chadwick 2011; Shah and Marchionini 2009; Wallsten 2010). Additionally, most literature on blogs has focused mainly on “top-blogs” as representatives of the blogosphere, ignoring the role of the vast long tail of non-authoritative blogs which we refer to here as tail-blogs (Karpf 2008b; Adamic and Glance 2005; Hargittai, Gallo, and Kane 2008).

This study addresses the dynamics of viral information in the blogosphere, and is aimed at understanding how blogs play a role in the virality process. The goal of this article is to address the debate presented in the literature about whether virality is caused by a top-down or a bottom-up process. Specifically, this study focuses on blogs and whether the process is driven by top-blogs or tail-blogs. To test for these dynamic phenomena and create a map of the “life cycle” of blogs posting links to viral information, we import econometric methodology from economics into the communication and information science literature. More specifically, we examine the linking practices of blogs to the most significant viral videos of the 2008 U.S. Presidential Election. To do so, we gathered data on all blogs \((n=9,765)\) and their posts \((n=13,173)\) linking to 65 of the top U.S. Presidential Election videos that became viral in the Internet during the period between March 2007 and June 2009. The primary motivation of this article lies in addressing what we see as shortcomings of the existing literature which focuses only on a small group of blogs (elite blogs) and assumes that these represent the

Published by Berkeley Electronic Press, 2011
full spectrum of the blogosphere and its behavior. These studies ignore the role of tail-blogs and other types of non-elite blogs in information diffusion. The second motivation for this article is derived from the fact that the majority of the literature uses static linking models and tends to focus on an examination of single points in time, which we believe misses the dynamic nature of the virality process. We address both of these gaps in this article.

The article is structured as follows: First, we provide a literature review regarding studies on virality of information, putting emphasis on the marketing and communication literature. We then review the scholarship on blogs, which discusses structures and dynamics of blogs in diffusing information, and map it into three main phases. We then introduce a new methodology in this context to examine the dynamic behavior of different types of blogs in the diffusion of information. Lastly, we discuss our findings regarding the unique role of each one of the following types of blogs: elite blogs, top-general blogs, top-political blogs, and tail-blogs.

The Beginning: Viral Marketing

As the use of social networks has grown, so has the marketing research that is concerned with web advertising and information virality. The term viral marketing was first coined by Jurveston and Darper (1997) as “network-enhanced word of mouth,” which in other words refers to “a communication and distribution concept that relies on customers to transmit digital products” (Helm 2000). Montgomery (2001, 93) later explained this term with biology-related nomenclatures: “a type of marketing that infects its customers with an advertising message, which passes from one customer to the next like a rampant flu virus.”

The motivation of marketing scholars to study the topic was clear: to increase the chances of successful diffusion of a message via electronic networks. This motivation led to a body of scholarship focusing on developing mathematical models that might ignite a viral process (Kiss and Bichler 2008; van der Lans et al. 2010). Additionally, other scholars attempted to identify factors, such as motivation and social conditions, that influenced the recipient’s acceptance and active participation in further diffusing the message (Palka, Pousttchi, and Wiedemann 2009; Phelps et al. 2004).

This body of literature refers to a bottom-up approach, where factors like trust, perceived risk, and perceived cost are crucial in determining whether users will engage in the viral process (Palka, Pousttchi, and Wiedemann 2009; Phelps et al. 2004). However, most of the marketing literature has focused on a top-down approach. That is, it tries to identify focal points or influencers, who could serve
as initial sets of customers, with the intent of achieving maximum dissemination of messages for a marketing campaign. This came to a peak through the work of Malcolm Gladwell in the "Tipping Point" (2002). Gladwell, and the majority of the marketing literature, suggests identifying gatekeepers or elites who could ignite the process of virality (Kiss and Bichler 2008). This is similar to the work of Katz and Lazarsfeld (1955), who suggested identifying opinion leaders as a pathway to reaching the masses with a message. Some voices have criticized this approach and suggested that these “hubs,” or highly connected people, aren’t crucial to the virality process. For example, Duncan Watts suggests that situational factors determine whether or not someone will pass information on, rather than any particular quality of the person. This implies that non-authoritative people (represented as tail-nodes in networks) are just as likely as “hubs” to be the ones who drive virality (Watts and Dodds 2007).

Recently, the topic of information virality has started to receive some attention in information science, political communication, and political science (Boynton 2009; Wallsten 2010; Barzilai-Nahon and Hemsley 2011), focusing mainly on political participation via social networks in cases where information becomes viral. However, social scientists have been more interested in understanding the political impact of viral information, rather than analyzing the process itself or its technicalities, as their peers in marketing have done (Robertson, Vatrapu, and Medina 2010; Ricke 2010; Klotz 2010; Gulati and Williams 2010). This article focuses on blogs and viral videos. Since viral diffusion of information is a behavioral phenomenon that cannot be properly understood without the contextual environment in which it operates, we focus next on the literature that analyzes blogs within the context of information politics.

**Blogs and Political Information Diffusion**

Starting in the mid-1990s, Internet studies scholarship has begun to address questions both of the impact of politics on information flows in the Internet, and of the impact of the Internet on politics (Hughes 1997; Schneider 1996). Since the mid-2000s, political communication scholarship on new media has come to recognize the fundamental importance of new ecologies of information and communication. Early enquiries focused primarily on new media technologies, particularly political blogs and interactivity on candidate websites, as alternative pathways for political participation (Stromer-Galley 2000; Williams et al. 2005; Xenos and Foot 2005). As these fields of research have grown, paralleling the
rapid developments in the landscape of digital politics, scholarship on political blogging has undergone at least three important progressions.

In the first phase, the majority of scholarship pointed to the relevance of political blogospheres, directly in relation to formal political participation, usually in the context of U.S. electoral politics. The focus was primarily on the interactions between these new forms of actors (the political bloggers), their communication vehicles (their political blogs), and the changing dynamics of their relationships with more traditional political actors and activities, such as electoral candidates and election campaigning (Williamson 2009). Scholarship in this domain has investigated political blogging as a space for formal party and campaign communications (Auty 2005; Jackson 2006), election PR (Trammell 2006), and negotiating electoral deliberation (Campbell 2009; Koop and Jansen 2009). A major theoretical consequence of this thread of inquiry has been the recognition and highlighted importance of political blogospheres as constituting new outlets of the public sphere (Chadwick 2011; Wright 2009).

The second important thread of scholarship on political dynamics of blogospheres has focused on the interrelationship between bloggers and traditional mass media (for example, the appropriation of news in blogs and usage of the blogosphere by mass media as an alternative). The comparative line of studies served an evaluative impetus: scholars have been both curious and critical of these increasingly popular competitors to mass media sources for political information (Johnson and Kaye 2004; Kwon and Moon 2009), but have also examined how mass media in turn understands political bloggers (Garden 2010; Jones and Himelboim 2010). Furthermore, a major normative concern, originating in this thread of inquiry into political blogs, has been examining the lack of adherence to professional norms of balance and objectivity of political blogs (Macias, Hilyard, and Freimuth 2009; Munger 2008), and the potential balkanization of public information (Baum and Groeling 2008; Sunstein 2008).

While the previous two threads of inquiry have engaged topics of power and politics in comparison to existing hierarchies of elite political actors and mass media gatekeepers, the latest progression of scholarship on the political blogosphere does so more directly. Here, scholars have begun to investigate the unique structuring of power and influence within the blogosphere and the digital information ecology (Karpf 2008a; Drezner and Farrell 2008; Ulicny, Matheus, and Kokar 2010). Most comprehensively, this body of work investigates the structural dynamics of the political blogosphere and its impact on information diffusion, political communication, and mobilization (Park and Jankowski 2008). Scholars have identified that these structural differences have an impact on agenda setting and political participation (Woodly 2008; Liu 2010; Wallsten 2007), that they blur public and private spheres of civic activity (Keren 2010;
Youngs 2009), and that they shape political learning and deliberation (Lawrence, Sides, and Farrell 2010; Leccese 2009).

Of particular importance to our article is this third phase of literature, which has highlighted the power-law distribution of actors in networks (Adamic et al. 2000) and particularly of political blogs (Adamic and Glance 2005; Drezner and Farrell 2008). The claim is for the existence of a skewed distribution in the blogosphere, where top-blogs capture the majority of attention from mainstream media elites and readers, and receive a disproportionately large number of links compared with other blogs. In other words, nodes that are rich in ties are likely to become even richer over time. Drezner and Farrell have also claimed that the top-blogs influence political elites through media actors who read them frequently (Drezner and Farrell 2008) and who act as network gatekeepers (Barzilai-Nahon 2008; Nahon 2011).

Our investigation arrives at this key intersection, claiming that the blogosphere does indeed have an impact on political information diffusion, and positioning both elite and non-elite blogs as having different types of power to shape the flow of information, especially in the political domain. Contrary to the general consensus, which focuses on elite actors within the political blogosphere, we present empirical evidence to show that the interactions between elites and the general networked-tail of blogs are more complicated. Moreover, we introduce a new methodology in information science and communication contexts for studying the dynamics of the behavior of blogs linking to content, wherein causalities between the different types of blogs and information diffusion are uncovered. This is a contribution to the existing literature that (in contrast) concentrates mainly on identifying static patterns of links between blogs and not between blogs to content.

**Research Design**

This article attempts to address a gap in the literature about the way in which political blogs link to content. Specifically, the content that we consider is videos that went viral during the 2008 U.S. Presidential Election. Our hope is to illuminate structures of behavior of political blogs in reference to content. More specifically, we address the following research questions: What are the relationships between different types of blogs and political viral information diffusion? What is the difference between elite blogs and tail-blogs in that process? Are there other types of blogs worthy of our attention as scholars? What would a life cycle that represents the chain of information diffusion in the context of blogs and content posting look like?
The literature presented above has certain limitations that we hope to overcome. First, it focuses mainly on static maps of blogs linking to blogs, surveys, or static design analysis of blogs. Second, it focuses on elite political blogs as representatives of the blogosphere, neglecting the role of other types of blogs, which may distort our understanding of the ecology of the political system. Our methodology aims to minimize these limitations. Moreover, the contributions of this article to the literature are numerous: First, we provide a *dynamic* analysis of real behavior of bloggers linking to content. Second, we identify four types of blogs that are factors in driving a process of political information diffusion, as opposed to only looking at elite blogs. The four types of blogs are (1) elite blogs, (2) top-political blogs, (3) top-general blogs, and (4) tail-blogs. The next section defines each of these types of blogs. Finally, we suggest that general blogs (and not just political blogs) should be taken into account when analyzing political information diffusion. Their role, as we will see later, is critical to the understanding of the political virality process.

**Data Collection**

Since we are looking at political blogs linking to content (specifically, videos) during the time of the 2008 U.S. Presidential Election, we necessarily require a large and complex dataset. Thus, our dataset is an amalgam of five different datasets aimed at addressing the following areas: (1) identifying the list of viral videos to be considered in the study; (2) daily-view data for the viral videos; (3) identifying the blogs linking to the viral videos; (4) obtaining traffic data for the blogs linking to the viral videos; and (5) categorizing our blog set into four types of blogs.

**Viral Videos List**

The list of viral videos was drawn from Viral Video Chart, a service that offered us a unique advantage over other services. This service used a sophisticated tracking and identification system that covered all known platforms of online video sharing (YouTube, DailyMotion, etc.) at the time of the 2008 U.S. Election. It provided daily lists of viral videos in multiple categories that it culled by scraping known areas of the web (such as social networking sites). It also analyzed and provided real-time information about these videos. At the time of the election this was the only service that provided holistic tracking of electoral and political viral videos. Other services, such as break.com, collegehumor.com,
failblog.com, and funnyordie.com, focused on specific types of content, none of which focused on the election as a primary area of focus.

We selected January 20, 2009, the inauguration day, as the date for archiving our video data, because we anticipated that the activity of viral videos would not conclude on election day (November 4, 2008). We drew the top 100 viral videos over the preceding year from each of the following three categories: political, election, and general. We ignored other categories that were not relevant to the election. We combined these lists and removed any duplicates. Next, four researchers (a faculty member and three Ph.D. students) coded the videos as being related to the election or not. We used a conservative approach in deciding whether to include a video in our sample: for a video to be included, all four had to agree that it was related to the election. The result was 120 unique videos out of 300 possible videos, all of which were on YouTube.

Viral Video Daily-View Data

Once we had the list of viral videos, we searched for a provider that could give us data about the daily views of each of the 120 videos. The daily-view data was critical since it serves as our dependent variable in our model. TubeMogul was the only service\(^1\) with historical, as opposed to real-time, video data, since it had access to the YouTube daily-view data (through the YouTube API). The service works on a crowd-sourced data retrieval business model, where it provides full viewing and video engagement statistics for videos requested by its service subscribers. We were able to gather complete daily-view data for 65 of our 120 videos that spanned from March 2007 to June 2009, and partial data for the remainder. Since the service works on a demand basis, the 65 videos with full data were identified by crowd-sourcing as the important viral videos.

Figure 1 is an example of what a viral video looks like, in terms of daily views. It plots the video “Yes We Can Obama Song” by Will.I.Am from when it was released, February 2008, until May 2009. As frequently happens in our collection of videos, the number of views spikes within a day or so of release, and generally declines more slowly over the next few days or weeks. In this example, the video went viral twice.

\(^1\) Other services like TubeKit did not provide complete historical data for our list of videos.
Table 1 gives some summary statistics for maximum and cumulative views across the videos in our set. “Peaks” refers to the distribution of the maximum daily views for each video. “Daily means” refers to the distribution of daily means for each video, which will help later in interpreting the model coefficients. Finally, “cumulative views” refers to the total views of our videos over time.

Table 1. Descriptive Statistics of the Viral Videos Views

<table>
<thead>
<tr>
<th></th>
<th>Minimum</th>
<th>Q1</th>
<th>Median</th>
<th>Mean</th>
<th>Q3</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peaks</td>
<td>5,412</td>
<td>110,200</td>
<td>232,700</td>
<td>292,700</td>
<td>358,600</td>
<td>1,621,000</td>
</tr>
<tr>
<td>Daily means</td>
<td>66</td>
<td>1,346</td>
<td>4,133</td>
<td>6,907</td>
<td>9,562</td>
<td>36,930</td>
</tr>
<tr>
<td>Cumulative views</td>
<td>17,830</td>
<td>394,500</td>
<td>1,373,000</td>
<td>2,255,000</td>
<td>2,577,000</td>
<td>17,870,000</td>
</tr>
</tbody>
</table>

Blogs Linking to Viral Videos

Our next goal was to identify blogs linking to the viral videos. For that purpose, we identified the most popular YouTube URL (i.e., unique identifier) for each of the 120 viral videos identified in the previous stage. Technorati and Google Blog Search were examined to determine which tool would result in the most comprehensive sample of blog posts linking to the videos in our dataset. Google Blog Search was selected because it returned the largest sample of blog posts and included the results returned from Technorati. We created scripts to automatically harvest all of the blog posts linking to these viral videos on a given day for a
given video from Google Blog Search. The data was cleaned to remove duplicates and to identify the blogs in which the posts appeared (see Appendix A for an elaboration). This resulted in a dataset of over 13,173 blog posts from 9,765 unique blogs that linked to our 120 viral videos between March 2007 and June 2009.

**Traffic Data for the Blogs**

We gathered monthly unique-visitors traffic data from the data service Compete. The reasons for using Compete as opposed to other available services (e.g., Hitwise, Technorati, or Alexa) were that Compete provided historical data, going back more than two years, and had the highest data coverage for our blog set. Note that we use this data for two purposes: (1) to identify top-general blogs and (2) as a control variable in our model, which we describe in detail below. The resulting list contained 3,101 blogs (Appendix B explains our use of this data).

**Identifying Four Types of Blogs**

Previous studies have focused mainly on top-blogs or political blogs when examining behavior in the blogosphere. Our study takes a broader perspective in that we examine the whole spectrum of blogs. This means that while we examine top- and political-blog behavior, as others have done, we also develop additional blog categories and include them in our analysis. This gives us a holistic picture of the blogosphere, which we believe is an important contribution to the literature.

A key focus of this article is examining the behavior of viral political videos. Thus, like the previous literature has done successfully, we identify a key group of blogs which are top-political blogs. Within that group of political blogs, two rise far above the others in terms of traffic and video linking activity. As we will justify below, these two blogs form their own elite group. However, the U.S. Presidential Election is not just a political event, and we therefore expect that a number of important, but non-political, blogs have also been important actors in driving viral video views. We refer to these as top-general blogs. Finally, the last group is called tail-blogs—the vast majority of blogs that receive little traffic. Figure 2 illustrates the distribution of these four types of blogs sorted by their

---

2 We harvested any blog post, not just from top-blogs or political blogs, as done in previous studies.

3 Appendix D provides the number and lists of blogs included for each category.
traffic, highest to lowest. The plot clearly shows a power-law distribution, which we discuss in the next section. We also associated with each blog the number of times it linked to a video in our set. This is represented by the size of the bubble for each blog (see Figure 2). We next describe how we collected the data for each one of the above four types and justify these decisions.

Figure 2. Power-law Distribution of Blogs Linking to Viral Videos

Elite Blogs. We found that the Huffington Post and Daily Kos were two unique blogs in our dataset. There are four reasons as to why this group needed to be separated into its own group. First, they have a much higher percentage of videos in our set that they link to (70 percent and 58 percent, respectively) than other top blogs (less than 10 percent). This is represented in Figure 2 by the two largest bubbles. Second, many researchers have acknowledged them as the most influential political blogs. For example, David Karpf calls them “the elite of the elite” (Karpf 2008a, 40). Third, they receive high unique-visitor traffic. Finally, when doing statistical analysis, the results show that differentiating this group is significant.

Top-political Blogs. Most scholarship on the blogosphere focuses on this group of important political blogs. Note that in the literature they are often called elite, top, or A-list blogs, and that in our study we have three types of elites: elite, top-political, and top-general blogs. Our initial set of top-political blogs was drawn from David Karpf’s Blogosphere Authority Index (BAI) (Karpf 2008a; 2008b), which is one measure of a blog’s authority (see Appendix C for a justification of our use of this index). Our set contains the top 25 conservative, and top 25 liberal, blogs from the week of August 8, 2008. We only included in this set the blogs that linked to our videos, so our final set of top-political blogs contained 39 blogs.
Also note that since we placed the Huffington Post and Daily Kos in our elite group, they have been removed from our list of top-political blogs.

**Top-general Blogs.** Our set of top-general blogs was created by taking all blogs from our dataset (excluding those listed in the top-political and elite-blog types) that had more than 250,000 unique visitors as listed by Compete (47 total blogs). Figure 2 shows that the point of 250,000 unique visitors is the inflection point of the power-law distribution of blog traffic. To the right of this point, the curve levels off and approaches zero. Any blogs above this line are certainly more influential, if only in terms of their higher traffic.

Blogs above the inflection point also exhibit a tendency to be farther apart in terms of traffic than those below it. In other words, the gap between the top 47 blogs, above the inflection, is around one million visitors per day, whereas the gap between the next 47 blogs, which fall below the inflection, is only 20,000 visitors per day. After removing the elite and top-political blogs which were included in the other sets of blogs, we are left with 32 top-general blogs.

**Tail-blogs.** Every other blog that linked to our viral videos, that is not in the other three types of blogs, is considered a tail-blog. In other words, tail-blogs represent blogs of users without high authority.

**Modeling**

**Econometric Time Series**

In this section we describe the quantitative method employed in testing our research questions. Borrowing from economics, we use a multiple regression model on time series data with both forward and backward time-lagged variables (Gujarati and Porter 1992). The use of lagged variables in econometric modeling allows us to examine how time plays a factor in the virality of videos. More specifically, it enables the testing of relationships between the daily video views in one time period with changes in our control and independent variables in either the previous or next time period.

Thus, at its most simple, our model examines relationships between daily video views and the following independent variable groups, where each independent variable group comprises both lag and non-lag variables:
Since our primary goal is to present a life cycle of blog post timing in the political information diffusion process, each independent variable group contains two variables:

1. A count of links from blogs in that category to a given video. For example, $\text{ELITE}_t$ would represent all the links from the elite blogs to a given viral video on a given day $t$.
2. A one-day forward-lagged version of the link count variable to the views. This variable associates links from day $t+1$ (tomorrow) to view counts of day $t$ (today). For example, $\text{ELITE}_{t+1}$ would represent all the links from the elite blogs on day $t+1$ to view counts for a given viral video on day $t$.

For each of our independent variables, a negative or a positive relationship to daily video views has the following meaning:

1. If $\text{variable}_t$ is positive, it means we find evidence that blogs of this type are posting links to a video on the day of the peak. This is an important concept because the vast majority of posts from blogs containing links to viral videos are on, or soon after, the day when the video receives its maximum number of views—its peak. Thus, we assume that statements made about the relationships between blogs and a video-view count will be most applicable around a video’s peak in the viral process.

2. If $\text{variable}_t$ is negative, it means we find evidence that blogs of this type are posting links to a video during the wind-down from its peak. This can be seen, for example, in Figure 1, where daily views decline after the first peak. If a blog posts during the decline, the link count is increasing while daily views are decreasing, which is a negative relationship.

If $\text{variable}_{t+1}$ is positive, it means we find evidence that blogs of this type are increasingly posting links to a video on a day $t+1$, when the views on day $t$ are increasing. When we take the peak as a reference point, it means that blogs of this type are posting on the day after the peak.

---

4 Error—represents unexplained variation in $Y$ (the dependent variable).
5 $\text{ELITE}_t$, $\text{ELITE}_{t+1}$, $\text{TOP-POLITICAL}_t$, $\text{TOP-POLITICAL}_{t+1}$, $\text{TOP-GENERAL}_t$, $\text{TOP-GENERAL}_{t+1}$, $\text{TAIL}_t$, and $\text{TAIL}_{t+1}$.
Thus, our model is:

\[ \text{VIEWS} = ELITE_t + ELITE_{t1} + TOP-POLITICAL_t + TOP-POLITICAL_{t1} + TOPGENERAL_t + TOP-GENERAL_{t1} + TAIL_t + TAIL_{t1} + \text{CONTROLS} + \epsilon \]

Table 2 summarizes each of the model variables.

Table 2. Description of the Model Components

<table>
<thead>
<tr>
<th>Variable/s</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIEWS</td>
<td>Our dependent variable is the daily-view count for any given video on any given day.</td>
</tr>
<tr>
<td>ELITE_t, ELITE_{t1}</td>
<td>This variable group represents our lagged and non-lagged elite blogs type (Huffington Post and Daily Kos). AQ1: Please confirm the change of “DailyKos” to “Daily Kos” to maintain consistency. It contains the number of links from these blogs to a given video on a given day.</td>
</tr>
<tr>
<td>TOP-POLITICAL_t,</td>
<td>This variable group represents our lagged and non-lagged top-political blog type.</td>
</tr>
<tr>
<td>TOP-POLITICAL_{t1}</td>
<td></td>
</tr>
<tr>
<td>TOP-GENERAL_t,</td>
<td>This variable group represents our lagged and non-lagged top-general blog type.</td>
</tr>
<tr>
<td>TOP-GENERAL_{t1}</td>
<td></td>
</tr>
<tr>
<td>TAIL_t, TAIL_{t1}</td>
<td>This variable group represents our lagged and non-lagged tail-blog type.</td>
</tr>
<tr>
<td>CONTROLS</td>
<td>These variables attempt to hold constant exogenous factors that could influence the virality process in our model.</td>
</tr>
<tr>
<td>VIDEO_ID</td>
<td>This variable controls for the specific characteristics of a particular video. It contains an id number for each of the 65 videos in our dataset.</td>
</tr>
<tr>
<td>VIEWS_{t-1}</td>
<td>This variable tries to control the momentum effect, wherein people who view a video may watch it again. It represents the views of a given video on a day ( t-1 ).</td>
</tr>
<tr>
<td>SUM_UNIQUE_VISITORS</td>
<td>This variable controls for the effect where some proportion of blog visitors will click on a link to video regardless of the blog type, the video, or the blog’s framing of the video. It controls for the fact that, regardless of the blog type, the more traffic a blog gets, the more views a video it links to will receive.</td>
</tr>
</tbody>
</table>
For summary, the model can be represented mathematically as follows:

\[
\text{views}_t = \beta_0 + \beta_1 \text{elite}_t1 + \beta_2 \text{elite}_t1 \text{BC} + \beta_3 \text{top-political}_t1 + \beta_4 \text{top-political}_t1 \text{BC} + \beta_5 \text{general}_t1 + \beta_6 \text{general}_t1 \text{BC} + \beta_7 \text{tall}_t1 + \beta_8 \text{tall}_t1 \text{BC} + \beta_9 \text{views}_t-1 + \beta_{10} \text{sum}_t1 + \text{FACTOR(video}_t1) + \epsilon_t
\]

We note that our model does not account directly for cases where people pass links to videos through email, social networking sites (e.g., Facebook), or micro blogging (e.g., Twitter). However, we believe that a great deal of the variation in our dependent variable, related to these kinds of links, will be captured in our control variable \( \text{VIEWS}_t - 1 \). This is because as \( \text{VIEW}_t - 1 \) is a negative lag variable, it should capture cases where yesterday's views influenced today's views. For example, when people post or mail links to their friends, there is generally some time lag between when the link was posted or mailed and when the receiver followed the link.

**Verifying Regression Assumptions.** We ran Variance Inflation Factor (VIF) tests to ensure that multicollinearity does not negatively impact our results. Each variable’s VIF was under 4, indicating that multicollinearity was not an issue in our model (Kahane 2001). We also ran a Durbin–Watson test, which tests for autocorrelation in the model. The resulting test statistics was 2.49, indicating a lack of autocorrelation (Ott and Longnecker 1993). Also, our Durbin–Watson test statistic is higher than our \( R^2 \)-squared, which is generally accepted as indicating that our time series model is not suffering from spurious regression (Gujarati and Porter 1992). Finally, note that because the dependent variable and the control variables (\( \text{VIEWS}_t1 \) and \( \text{SUM_UNIQUE_VISITORS} \)) follow a power-law distribution, we transform these variables using a natural log, so they are closer to a normal distribution.

Rather than the actual coefficient values, we are interested in interpreting the significance of the relationships between our independent and dependent variables. Interpreting the coefficients is meaningless for two reasons: (1) the distribution of daily video views follows a power law, not a normal distribution; (2) most of our links cluster around the peaks, or the day of maximum daily views. If we chose to interpret the coefficient, it would have provided us with the expected (average) change of video views for a unit change in \( X \), which fails to represent both a power-law distribution and clustering around peaks.
Hypotheses

H1A: We expect ELITE to have a positive relationship with VIEWS because we expect these blogs to be posting links as the daily-view count is rising. Since daily views rises sharpest at the peak (generally at the beginning of the viral process), a positive relationship here implies that our elite blogs are posting links at the very beginning of that process.

H2A: We expect not to find a relationship between ELITE_t1 and VIEWS. Since we expect the elite blogs to post on the day of the peak, we do not expect them to post on any other day.

H3A: We expect to find a negative relationship between TOP-POLITICAL (H3A1), TOP-GENERAL (H3A2) or TAIL (H3A3) and VIEWS, which reflects the idea that if they are follower, and not leader, blogs they will be unlikely to post as the view count is rising, and more likely to post while the view count is falling.

H4A: We expect TOP-POLITICAL_t1 (H4A1), TOP-GENERAL_t1 (H4A2), or TAIL_t1 (H4A3) to be positive with VIEWS, which means that a higher number of views of a video on a given day would cause blogs in these respective types to link to it on the day after the peak.

Results

The results (see Table 3) show that the model explains 96.5 percent of the variance of the daily-view count, and fits the data well. Moreover, it appears to be statistically significant (see F-statistics), indicating that our selection of variables explains the majority of the variation in daily viral video views. Table 3 also shows our variables and their associated coefficients.
Next, we explain the relationships found, which relate our hypotheses to our findings.

**Elite Blogs (H_{1A} and H_{2A}).** We found that there is a positive relationship between the number of links from elite blogs and the number of video views (H_{1A}). We also found no evidence that elite blogs are responding to video-view counts, meaning that H_{2A} was not significant, as expected. This means that elite blogs are posting as the video is going viral; however we don’t find any evidence that they post during any other time in the life cycle.

**Top-general Blogs (H_{3A2} and H_{4A2}).** An important finding suggests a recursive relationship wherein top-general blogs both drive, and were driven by, video views. This finding is supported by a positive relationship between top-general blogs and views both on day $t$ and day $t+1$ (H_{3A2} and H_{4A2}). This means that they are posting right around the peak, but in contrast to tail-blogs, they don’t post to a video that is beyond its prime time. This also suggests that top-general blogs will generally refrain from posting news or content that is old.

**Top-political Blogs (H_{3A1} and H_{4A1}).** We found that top-political blogs respond to video views. However, we found no evidence that they drive views. The relationship between top-political blogs posting the day after is positive with video views (H_{4A1}). Moreover, we do not see any correlation between posts from these blogs on day $t$ with the views on day $t$ (H_{3A1}). Consequently, we find that
top-political blogs are followers, posting only after the elites and the top-general blogs.

**Tail-blogs (H\textsubscript{3A3} and H\textsubscript{4A3}).** Here we found strong support that tail-blogs respond to views (H\textsubscript{4A3}), and that the relationship of tail-blogs posting links to viral videos is negative with the views (H\textsubscript{3A3}). This strongly suggests that they post on the day after the peak and when the viral cycle is winding down from its peak. In other words, one might infer that this group of blogs is following all the different top-blogs (elite, top-political, and top-general blogs).

Figure 3 represents our findings in an illustrative way, showing the life cycle of a viral process of a video during the 2008 U.S. Presidential Election. In this life cycle, one can see when different types of blogs are driving the view count for videos. Clearly, elites and top-generals are first, followed by top-political, and finally tail-blogs. Note that top-general blogs are the only group that drive, and are driven by, views in the initial viral process.

Figure 3. Life Cycle of a Viral Video in the Blogosphere
Discussion

We would like to discuss here two phenomena that are implicit in our results. First, the blogosphere is not a monolithic sphere, and therefore researchers should not study only elite blogs and expect them to represent the entire blogosphere. Also, scholars should acknowledge the important role of blogs in general, and top-general and tail-blogs in particular, in this ecological blogosphere. Second, while the power-law distribution exists when information goes viral, powerful actors cannot exist as elites without the masses. Nahon (2011) calls them “transient elites.” We discuss these phenomena, while acknowledging that a Presidential Election might have specific characteristics that influence virality.

First Phenomenon: Blogs Are Not Monolithic

Elite Blogs Do Not Represent Blogs in General. Most blogosphere research focuses on elites, either as explicit representatives of the blogosphere (Hargittai, Gallo, and Kane 2008; Adamic et al. 2000; Karpf 2008b), implicit representatives of other blogs (Benkler and Shaw 2010), or as boundary spanners and intermediaries between blogs and other political and media actors (Wallsten 2007). Regardless of how studies have examined blogs in the literature, they mostly overlook the heterogeneity of this sphere. Some researchers acknowledge the un-monolithic nature of the blogosphere (Benkler and Shaw 2010; Munger 2008), but our study is one of the first that is able to show empirically how these types of blogs differ in their behavior and are significant factors in the process of information flow in networks. Moreover, contrary to the literature that focuses on top-political blogs as the unit of analysis for understanding political information diffusion, we found that their significance might be misperceived since top-political blogs, as represented by our typology, are followers of the top-general and elite blogs. Consequently, in many cases, the top-political blogs do not socially construct frames, nor do they set the political agenda. Instead, they replicate the agenda according to the frames created by the elite (the “elite of the elite”) and top-general blogs. This is not to claim that top-political blogs are not significant, but rather that there are other important blog types which have been overlooked in the literature, and which should come forward in future scholarship.

Political Information Is Spread via General Channels. The role of top-general blogs in diffusing political information needs elaboration. Bennett suggests that new media forms, like blogs, transform public communication by removing intermediaries and emphasizing direct contact with individuals (Bennett and Manheim 2006). According to this, blogs impact decision makers and influence
agenda setting directly. On the other hand, Drezner and Farrell (2008) suggest that journalists play the role of boundary spanners between political elites and the blogosphere by focusing the attention of political actors on important information that exists in blogs. For them, journalists are the intermediaries in this process. Our article shows, by looking at the life cycle of virality, that the chain of political information diffusion during elections is much more complicated. It is a top-down process in a bottom-up environment. While blogs may potentially reach out directly to the general public, in practice, virality happens through several intermediaries and not directly. There are several intermediaries in the viral process before information comes to the attention of the general public. More importantly, virality is ignited mainly by the top-general blogs who are the first movers. This finding is of high importance since “bloggers possess first-mover advantages in formulating opinions” (Drezner and Farrell 2008, 17). Now we know that elite and top-general blogs mainly possess that advantage. But, in order for political information to reach the general public and decision makers, boundary spanners (the top-general blogs) need to diffuse this information to an audience that is more general, and less focused on politics. Otherwise, the chances for that information to go viral are slim. This is also why we see the top-political blogs and tails as followers.

The Chicken and the Egg Question. As previously mentioned, Drezner and Farrell suggest that top-blogs scan the blogosphere and highlight important information for decision makers. Their claim implies that tail-blogs create content, and that top-blogs are mainly responsible for replicating and highlighting original content. When it comes to viral information we see otherwise. While we don’t get into a discussion of who creates the original content (such as the viral video itself), we see that elite and top-general blogs are mainly responsible for creating new information in the form of blog posts. These posts act both as a means of propagation of the original content, and add value through additional analysis and spin. According to our picture, tail- and top-political blogs serve as followers: they are far less influential than previously thought. Even when taking into consideration the vast number of tail-blogs, they are not powerful enough to create or sustain the viral process alone, as Anderson (2006) claims in “The Long Tail”. Rather, they are responsible for reducing the rate of decay of viral information. They keep the information in the public awareness longer than it might otherwise stay there.

Thus, like Wallsten argues (2007), we see that tail-blogs are dependent on the elite and top-general blogs to ignite the viral process and reach the attention of wider audiences. Nevertheless, since attention here is measured not only in passive terms of viewing, but also in active terms of posting links, creating and sharing content, this may imply that the tail-blogs are also dependent on their tail-
peers. In contrast, top-general blogs show a recursive relationship, implying that they respond to and generate their own information eco-sphere. So basically, top-general blogs would post information that is similar to their peer competitors. For example, if TechCrunch posts information about Obama’s contributions, The New York Observer would quickly pick up and report on the same story. This was supported also in a study done by Barzilai-Nahon and Hemsley, which found a bandwagon effect among top-blogs (2011).

Second Phenomenon: Transient Elites Are Constituted by the Masses

The phenomenon of a power-law distribution in the blogosphere is documented in many empirical studies (Adamic et al. 2000; Barabasi 2003; Drezner and Farrell 2008). Drezner and Farrell point out that this skewed distribution has important consequences for the salience of different blogs, with respect to their position on the distribution (2008). They also argue that a high degree of disparity in visibility of blogs leads some blogs to become elite blogs, attracting attention from both other bloggers and political elites. In other words, they imply that the role of elite blogs is a transient and dynamic one. Similarly, Karpf suggests that YouTube creates “a dissolution of elite control and the creation of more porous elite networks” (2010). Our empirical investigation shows that their role in the viral life cycle determines the salience of a blog within the blogosphere. Therefore, it supports the idea that elites are transient (Nahon 2011), and that the identification of what constitutes an elite on the Internet changes dynamically over time. Castells asserts that power relationships between elites and non-elites are mainly framed by domination (2009). Moreover, he asserts that “Conflicts never end. They simply pause through temporary agreements and unstable contracts that are transformed into institutions of domination” (Castells 2009, 14). However, if the power relationships of elites vs. non-elites are so dynamic and interchangeable, then the meaning of domination by elites needs to be revisited. In our case, the elites are not only determined by their ability to attract views; they are also constituted by the masses (all other blogs and users) that actively link to them, and that actively crown them as elite. In the blogosphere, these transient elites cannot rule without active participation and support by the masses.

Conclusion

Our article identifies the dynamics of virality of information on the Internet (specifically videos) at the time of the 2008 U.S. Presidential Election by examining the relationship between blogs producing information (posting links)
and users following these links and watching the videos. We present a new methodology in information science and communication, borrowed from economics, that allows us to perform a time series analysis through multiple regression. In this article we examine the linking practices of blogs to the most significant viral videos of the election.

Our methodology addresses some gaps present in the blogosphere literature. These gaps are exemplified by (1) focusing mainly on static maps of blogs linking to blogs, surveys, or static design analysis of blogs instead of dynamic behavior of blogs linking to content; and (2) focusing on elite political blogs as representatives of the blogosphere, neglecting the role of other types of blogs. Among other things, our findings illuminate the importance of different types of blogs: elite, top-political, top-general, and tail-blogs. We also find that while elite and top-general blogs create political information, they are responsible for driving the viral process, whereas top-political and tail-blogs act as followers in the process, who reduce the rate of decay of viral information.

While this research explicated the life cycle of information virality of blogs linking to content during the election, further research needs to be undertaken to complement these quantitative results with qualitative research (for example, content analysis). Finally, this study calls for further investigation of the roles of other types of blogs in other contexts.

Appendix A. Blog Data Cleaning

The list of blogs linking to viral videos was cleaned to remove duplicates and to identify the blogs in which the posts appeared. The following criteria were used:

1. Blogs that only differed in URL syntax (i.e., huffingtonpost.com and www.huffingtonpost.com) were merged into the same blog.
2. URLs that automatically redirected to another URL in our dataset were merged into the same blog.
3. Blogs with a clear community and under the same domain with equal or lesser popularity or strength (e.g., the multiple authors at dailykos.com) were merged into the same blog.
4. Blogs with a clear community and under the same domain with different strength were not merged, but a link to the main website was maintained in our database (e.g., Andrew Sullivan, who blogs at theatlantic.com, is more popular than his host).
5. Blogs without a clear community or relation (e.g., the many blogs hosted at wordpress.com) and under the same domain were not merged. Examples of blogs in this rule are exemplified by a site hosting individual blogs. 

Maintaining a link to the main website for blogs in cases 4 and 5 above allowed us to have the option of treating prolific blog authors as separate blogs or collapsing them into their parent websites.

**Appendix B. Using the Compete Service for Daily-Traffic Data**

Compete does not provide data in two cases. First, Compete does not provide data for blogs with daily traffic below 1,000 visitors. Since these blogs will tend to fall into the category of tail-blogs, this has no effect in defining the top-general blogs. Second, Compete does not provide data for sub-site-level blogs. It tracks visitor data at the site level (i.e., site.domain.com and domain.com), so blogs in folders (site.domain.com/myblog) and blogs without a full domain match were excluded. For the vast majority of important blogs in our dataset, this proved not to be a problem since the number of excluded blogs, near the cut-off point (250,000 daily-traffic), that could be mis-categorized is very small. We expect that the error term in our model captures this problem. The resulting list contained 3,101 blogs.

**Appendix C. Using the Blogosphere Authority Index to Assess Top-Political Blogs**

Our set of top-political blogs was drawn from the Blogosphere Authority Index (BAI) (Karpf 2008a; 2008b), which is one measure of a blog’s authority. We used this measure instead of others for two important reasons. First, many other ranking systems rely on distinct measures of influence, such as network centrality, in/out linking measures, site views, daily readership, and contributors. The BAI has an advantage in that it synthesizes these different measures into one authority index. Second, since our video data referred to historical data (we collected it after the fact, instead of in real time), we needed an index that had historical data coverage of influential blogs during the time of the 2008 U.S. Presidential Election. The BAI recalculates the index weekly and archives its historical data, and so obtaining measures of influence for political
blogs during the election was a straightforward process. We selected our BAI rankings from the week that we archived our viral videos. Certainly those rankings changed over the duration of the election, but an informal inspection showed that for our purposes the changes were not significant.

### Appendix D. The Four Categories of Blogs

<table>
<thead>
<tr>
<th>Blog Category</th>
<th>Number of Blogs</th>
<th>Example Blogs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elite blogs</td>
<td>2</td>
<td>The Huffington Post (huffingtonpost.com), Daily Kos (dailykos.com)</td>
</tr>
<tr>
<td>Top-political blogs</td>
<td>39</td>
<td>MyDD (mydd.com), Talking Points Memo (talkingpointsmemo.com), RedState (redstate.com), Washington Wire (blogs.wsj.com/washwire), Firedoglake (firedoglake.com), The Next Right (thenextright.com), NewsBusters (newsbusters.org), Open Left (openleft.com), Crooks and Liars (crooksandliars.com), Americablog News (americablog.com), Hit &amp; Run (reason.com/blog), Boing Boing (boingboing.net), The Daily Dish (andrewsullivan.theatlantic.com/the_daily_dish), Althouse (althouse.blogspot.com), ReadWriteWeb (readwriteweb.com), Power Line Blog (powerlineblog.com), Pajamas Media (pajamasmedia.com), The Gateway Pundit (gatewaypundit.rightnetwork.com), Stop The ACLU (stoptheaclu.com), Balloon Juice (balloon-juice.com), FiveThirtyEight Blog (fivethirtyeight.blogs.nytimes.com), Mo Jo (motherjones.com/mojo), Feministing (feministing.com), Hullabaloo (digbyblog.blogspot.com), Informed Comment (juancole.com), Shakesville (shakespearessister.blogspot.com), Pam’s House Blend (pamshouseblend.com), RealClearPolitics (realclearpolitics.com), Political Radar (blogs.abcnews.com/politicalradar), Gadling (gadling.com), Official Google Blog (googleblog.blogspot.com), TV Squad (tv squad.com), Teagan Goddard’s Political Wire (politicalwire.com), Ezra Klein (voices.washingtonpost.com/ezra-klein), Yglesias (yglesias.thinkprogress.org), Patterico’s Pontifications (patterico.com), Pandagon (pandagon.net), IMAO (imao.us), QandO (qando.net)</td>
</tr>
<tr>
<td>Top-general blogs</td>
<td>32</td>
<td>my.barackobama.com, NewTeeVee (newteevée.com), MetaFilter (metafilter.com), Urlesque (urlesque.com), moli (moli.com), The New York Observer (observer.com), Google Blogoscoped (blogoscoped.com), neatorama (neatorama.com), Paste Magazine (pastemagazine.com), College OTR: Omega Tau Rho (collegeotr.com), Brave New...</td>
</tr>
</tbody>
</table>
Films (bravenewfilms.org), Portfolio.Com (portfolio.com), Glamour magazine (glamour.com), VentureBeat (venturebeat.com), GigaOM Tech News (gigaom.com), Telegraph Blogs (blogs.telegraph.co.uk), Web Hosting Talk (webhostingtalk.com), POLITICO.com (politicocom), TechCrunch (techcrunch.com), TeamSugar—Celebrity Gossip, Entertainment News & Celebrity News (teamsugar.com), RNC: Republican National Committee (gop.com), PopSugar (popsugar.com), Political Irony > Humor and Hypocrisy from the World of Politics (politicalirony.com), The Inquisitr (inquisitr.com), SideReel (sidereel.com), The LRC Blog (lewrockwell.com), College Confidential (talk.collegeconfidential.com), Plurk is a social journal for your life (plurk.com), Best City Guide (digitalcity.com), green-living—Decorating Ideas & Tips from Shelterpop (greendaily.com), JackBook.Com (jackbook.com), Laughing Squid (laughingsquid.com)

Tail-blogs⁶ 5,566 In Light of the Gospel (www.inlightofthegospel.org), iPandora (http://www.ipandora.net), Democralypse Now (democralypsenow.blogspot.com), Criticism As Inspiration (criticismas.wordpress.com), My View from the Center (webcentrist.wordpress.com)

References


⁶ This is not an extensive list of the tail-blogs of our dataset, but rather a few random examples. The full list can be retrieved from http://retroV.org.

http://www.psocommons.org/policyandinternet/vol3/iss1/art2
DOI: 10.2202/1944-2866.1108


http://www.psocommons.org/policyandinternet/vol3/iss1/art2
DOI: 10.2202/1944-2866.1108