This is the authors' version of the paper accepted for publication in Internet Research journal.

Virality overYouTube: an Empirical Analysis

GoharFeroz Khan*, SokhaVong**

*Corresponding author, School of Industrial Management, Korea University of Technology & Education (KoreaTECH), 1600 Chungjol-roByungcheon-myun, Cheonan city, 330-708, South KOREA. Office: 82-41-560-1415, Email: gohar.feroz@kut.ac.kr or gohar.feroz@gmail.com

** Health, Food Security and Nutrition Department of Council for Agricultural and Rural Development, Office of the Council of Ministers, Russian Bulvd, Room E1-4, Cambodia, email: sokhavong2002@yahoo.com

Acknowledgement:This research was supported by the research subsidy for newly-appointed Professor of Korea University of Technology and Education (KoreaTECH) for the year 2012.

Abstract

Purpose: The purpose of this research is to seek reasons for some videos going viral over YouTube (a type of social media platform).

Methodology: Using YouTube APIs (Application Programming Interface) and Webometrics analyst tool, we collected data on about 100 all-time-most-viewed YouTube videos and information about the users associated with the videos. We constructed and tested an empirical model to understand the relationship among users' social and non-social capital—user age, gender, view count, subscriber, join date, total videos posted, video characteristics—postdate, duration, and video category, external network capital—in-links and hit counts, and Virality—likes, dislikes, favorite count, view count, and comment count. Partial Least Square (PLS) and Webometric analysis was used to explore the association among the constructs.

Main findings: Among other findings, our results showed that popularity of the videos was not only the function of YouTube system per se, but that network dynamics (e.g., in-links and hits counts) and offline social capital (e.g., fan base and fame) play crucial roles in the viral phenomenon, particularly view count.

Originality: We for the first time constructed and tested an empirical model to find out the determinants of viral phenomenon over YouTube.

Keywords: Social media, YouTube, viral videos, Virality, Webometrics, structural equation modeling (SEM).

Introduction

In contemporary society, social media ischanging the way people create, share, and consume information(Mangold & Faulds, 2009). For instance, people tend to electronically gather information from many sources, share back to their networks, and interact with friends on the networks. This makes social media an importantnew communication tool. Social media consists of a variety of tools and technologies that includes collaborative wikis, Blogs, content communities (e.g., YouTube), social networking sites, folksonomies or tagging (e.g., delicious), virtual game worlds, virtual social worlds (e.g., Second Life), and all other internet-based platforms that facilitate the creation & exchange of user generated content (UGC)(Khan, 2013; Khan & Swar, 2013).Diffusion and use of social media in contemporary society is noteworthy.For example, as of May 2013, Facebook possesses more than 1.11 billion users, Google+ has more than 500 million users, Twitter has 500 million accounts, and over 6 billion hours of video are watched each month on YouTube and over 100 hours of video are uploaded to YouTube every minute. The exponential growth of social media in contemporary society makes them necessary tools for communication, content creation, sharing, and business growth(Kaplan & Haenlein, 2010).

Among the different types of electronic communication mediaavailable today—such as electronic mail, graphics, and phone—YouTube, a video based communication medium,is one of the most successful to express feeling, communicate with friends, and advertise business messages. However, not all content posted on YouTube getsthe desiredattention and only a fractioncan reach a large audience, particularly the videos posted by social media marketers expecting millions of views. What causessome videos to get millions of hits? In the other words, why do they “go viral”?

A viral video refers to a video that can be quickly shared and become popular throughout social networks (Broxton, Interian, Vaver, & Wattenhofer, 2010). The term ‘viral’ is a metaphoric reference to a contagious virus which spreads quickly from one host to another. The marketing field uses the term ‘Viral

Marketing', first proposed in 1997 by Steve Jurvetson and Tim Draper, to describe Hotmail ad services (Camarero & San José, 2011). Viral marketing is strategic marketing to promote a business brand by embedding it into content aimed at online social network (Mills, 2012). Some researchers consider viral marketing as a word-of-mouth (WOM) strategy encouraging people to communicate to each other about certain products or services (Phelps, Lewis, Mobilio, Perry, & Raman, 2004a). The purpose of this study is to identify the determinants of virality overYouTube. By understanding virality characteristics of videos, the study will help marketers to determine influential factors in marketing campaign(West, 2011)and to enhance brand advocacy and brand awareness (Kirby & Marsden, 2006).

The rest of this paper is organized as follows: The next section provides theoretical background related to the viral marketing, role of YouTube in viral marketing, and social capital theory and constructs our hypothesis; followed by the methodology and the results sections. We conclude with a discussion and summary of our main findings in discussion section.

Theoretical Background& Hypothesis

Viral Marketing Theory and the YouTube

The term "viral marketing" is used to describes the phenomenon by which individuals mutually share and spread marketing messages or information, initially sent out deliberately by marketers to take advantage of WOM behaviors (Lans, Ralf, Eliashberg, & Wierenga, 2010). Viral marketing acts as an electronicWOM concept (Bampo, Ewing, Mather, Stewart, & Wallace, 2008a).WOM refers to the phenomenon of person-to-person conversation between consumers about a product/service (Sen & Lerman, 2007), and the same phenomenon occurring over some electronic media, such as the Internet and social media, is referred to as electronic WOM (eWOM) (Brown, Broderick, & Lee, 2007; Davis & Khazanchi, 2008; T. Hennig-Thurau & Walsh, 2004). eWOMis defined as, "any positive or negative statement made by potential, actual, or former customers about a product or company which is made available to a multitude of the people and institutes via the Internet" (T. Hennig-Thurau & Walsh, 2004) (p.39). The persuasive use of social media offers a fertile ground for eWOM marketing. For example, more and more consumers use social media tools—such as, blogs, social network sites, and content communities, and online discussion forums—to exchange product information (Lee, Park, & Han, 2008).

Although there is some controversy over the similarity of viral marketing with regard to the WOM phenomenon (Pastore, 2000), there is still a strong support for the similarity between them (Camarero & San José, 2011; Phelps, et al., 2004a). The situation that makes viral marketing and the WOMconcept similar is the network effect and role of influence that causes the message to rapidly reach wider recipients and get positive responses (Vilpponen, Winter, & Sundqvist, 2006). However, viral marketing takes competitive advantages over the traditional WOM in term of lower cost, high credibility, faster diffusion, and better targeting (Bampo, et al., 2008a).

The existing literature points to several factors crucial for viral marketing success including, 1) content (Porter & Golan, 2006) , 2) social network structure(Bampo, et al., 2008a), 3) characteristics of the recipients (Camarero & José, 2011; Thorsten Hennig-Thurau, Gwinner, Walsh, & Gremler, 2004; Phelps, Lewis, Mobilio, Perry, & Raman, 2004b; Wojnicki & Godes, 2008), 4) seeding strategies, and 5) role of influential users(Iyengar, Bulte, & Valente, 2011; Marcus & Perez, 2007). The content hypothesis points to attractiveness (Gladwell, 2002; Porter & Golan, 2006), usefulness(Pousttchi & Wiedemann, 2007), and emotional aspect of content (Berger & Milkman, 2009) in viral phenomenon.The social network hypothesis of viral marketing studies the structure of the social network where the content is posted (Bampo, et al., 2008a), for example, it looks into the viral phenomenon from network characteristicssuch as, the source, structure, and size of network and how the message is propagated through a network (Bampo, et al., 2008a; Ko, Yin, & Kuo, 2008; Liu, Liu, & Li, 2012). Characteristics of the recipients hypothesis looks into role the social, behavioral, and motivation characteristics of the recipients of the content in viral phenomenon (Bampo, et al., 2008a; Camarero & San José, 2011; Cheung, Lee, & Rabjohn, 2008; Thorsten Hennig-Thurau, et al., 2004; Jalilvand & Samiei, 2012; Wojnicki & Godes, 2008). The seeding strategy literature investigate the size of the initial network users required to trigger the viral phenomenon (Bampo, Ewing, Mather, Stewart, & Wallace, 2008b; Kalish, Mahajan, & Muller, 1995; Libai, Muller, & Peres, 2005), and finally, the tastemakers (Allocca, 2011)or influential users hypothesis(Iyengar, et al., 2011; Marcus & Perez, 2007; Subramani & Rajagopalan, 2003)assert the role influential people in the viral phenomenonwho have strong network ties or presence (Iyengar, et al., 2011).These factors are emphasized in the framework of viral marketing to understand the nature of influence in the viral process(Marcus & Perez, 2007; Subramani & Rajagopalan, 2003).

Role of YouTube

In contemporary society, YouTube is becoming an influential medium for social interaction among people.YouTube is one of the leading video-on-demand (VoD) platforms for user generated content (UGC). UGC on YouTube permits users to creatively produce and share content on the platform to empower new ideas and business opportunities (Cha, Kwak, Rodriguez, Ahn, & Moon, 2007)including branding and marketing strategy (Mills, 2012). The contentproductionthrough social media (e.g., YouTube) allow users to fulfill their information, entertainment, and mood management needs,while its generation (or sharing) allows for self-expression and self-actualization(Shao, 2009). On the other hand, the viral video concept plays a crucial role in business marketing for reaching tremendous target audiences within short periods of time. YouTube platform has several embedded features to boost social interaction(Benevenuto et al., 2008), such as, the users' ability to comments on a video, liking/disliking a video, or share a video to other social network platform such as, Facebook or Twitter. These factors may contribute positively to the virality a video(Benevenuto, et al., 2008; Kaplan & Haenlein, 2010).

Particularly, the video response feature of the YouTube plays a crucial role in social network relationship and users interaction (Benevenuto, et al., 2008). On YouTube, social network features permits users to share different categories of videos with different groups of people (Lange, 2007). Thus, we believe that different types or categories of video (e.g., music, comedy, drama, and animation) may affect viral phenomenon differently. An appealing study on user generated content (video content) illustrated the result of difference video popularity and length between user generated content and non-user generated content (professional)(Cha, et al., 2007). YouTube is a content community, containing a social integration concept which enables video rating and commenting activities to boost up popularity of videos (Kaplan & Haenlein, 2010). In addition, the age of video, and duration after the published date contribute to popularity of the video (Cha, et al., 2007). Thus, in light of the content hypothesis (Berger & Milkman, 2009; Gladwell, 2002; Porter & Golan, 2006; Pousttchi & Wiedemann, 2007)which points to the role of content in viral phenomenon, we suggest that video characteristics i.e., Published Date, Video Length, and Video Category will positively affect the viral phenomenon.

H1: Video's characteristics (i.e., Published Date, Video Length, and Video Category) will positively affect viral phenomenon.

Social Capital Theory

If broadly defined,social capital "is the goodwill available to individuals or groups. Its source lies in the structure and content of the actor's social relations. Its effects flow from the information, influence, and solidarity it makes available to the actor."(Adler & Kwon, 2002) (p.23).This working definition illustrates the important dimensions of social capital—social structure and social relations—for the sources of social capital embedded therein. In addition, social capital incorporates three dimensions: structural, relational, and cognitive (Chow & Chan, 2008). Structural social capital contains two imperative properties of individuals' position within the network such as connectedness and integration. Connectedness property refers to the level of associating with others, while integration property focuses on level of communication with others within the network (Camarero & San José, 2011; Kleijnen, Lievens, de Ruyter, & Wetzels, 2009). For instance, if an individual has many friends on their YouTube network and maintains good interaction within the network, it will contribute to thespreadabilityand growth of videos viewed in the network. In other words, accessibility and social interaction within the network stipulate the rapid spread of videos. According to Broxton et al. (2010), social characteristics are defined by connoting external links from other social networks by clicking on the links or copying URL (Uniform Resource Locator) to paste on browser(Broxton, et al., 2010). Furthermore, the structural social capital is stressed in the SPIN—Spreadability, Propagativity, Integration, and Nexus—framework addressing on viral concept of social media (Mills, 2012). Besides structural social capital, there is a relational social capital which is determined by the closeness and strength of ties within the network. The strength of ties has elevated influence on WOM, encourages a reciprocal behavior, and trust for individual relationship (Camarero & San José, 2011). According to Allocca (2011), YouTube's Trend Manager, the source of viral video can be derived by three key characteristics on YouTube: tastemaker, participation in the community, and unexpectedness(Allocca, 2011). Tastemakers are influential people who have strong network ties or presence, for example, in terms of YouTube they are the users with large numbers of followers. As a result, the user characteristics, which are determined in the structural and relational social capital, support videos going viral more rapidly within the network.

Camarero and San José (2011) looked at determinants of viral dynamics in order to identify the viral process. They adopt social capital theory and social network theory to explain how the viral dynamic process happens by utilizing social capital resources: structural and relational social capital (Camarero & San José, 2011). Also, studies suggest that in the social media context, person-to-person exchanges are associated with an increase in bridging social capital (Burke et al. 2011). Similarly, Broxton, et al. (2010), suggested looking into social and non-social aspects of videos to invistigate the viral phenomenon.Social network structure, for example, the structure and size of network play a vital role in viral marketing (Bampo, et al., 2008a; Ko, et al., 2008; Liu, et al., 2012). Thus, taking into account the role ofnon-social capital (Broxton, et al., 2010) and influential users(Marcus & Perez, 2007; Subramani & Rajagopalan, 2003)i.e., the user with strong social capital (Camarero & San José, 2011) in viral phenomenon, we posit that a video posted by an influential user i.e., having a strong network and non-social capital, may go viral faster as compared to a video posted by an ordinary user. Based on the above reasoning, we proposed the following two hypotheses.

H2: User’s social capital (i.e.,User’s View Count and Subscriber Count) will positively affect viral phenomenon.

H3: User’s non-social (i.e.,Total Videos Posted and Join Date) will positively affect viral phenomenon.

Research Model

To understand the determinants of viral videos, in light of the theoretical background, we constructed a research model (Figure 1) based on YouTube system’s variables (discussed later in the method section). More specifically, we investigated the virality of videos based on two important determinants: YouTube user’s characteristics—i.e., the social, non-social capital of the users who posted the videos and YouTube video’s characteristics—i.e., VideoPublish Date, Duration, and Category. Social capital of a YouTube user is defined as the number of followers or subscribers and number of view counts received by a user. Similarly, non-social capital of a user is the characteristics other than social, such as Total Videos Posted by a user and Join Date.In this study, viral phenomenon is measured with factors relating to Video’s FavouriteCount, Video’s View Count, Video’sComment Count, Video’s Likes, and Dislikes received by a video as shown in the Figure 1. Video’s View Counts indicate the number of time a video is played by users. Video’s Comment Count indicates the number of comments received by a video. Video’s Likes and Dislikes indicate the number of people liked or disliked a video after watching. These properties indicate the popularity of a video. Also two control variables (i.e., User Age and Gender) were included in the model.

<Figure 1 about here>

Data & Method

Data related to the 100 all-time-most-viewed videos was collected from YouTube using standard YouTube API tools in combination with Webometrics Analyst(Thelwall, 2005) on 22 May, 2012.The variables included information related to the YouTube videos i.e.,Video’s Post Date (the date a video was first posted); Video’s Duration (duration of video in minutes); Video’s Category; Video’s Likes (numbers of likes received by a video); Video’s Dislikes (numbers of dislikes received by a video); Video’s Favorite Count (number of times a video is listed as favorite); Video’s View Count (number of times a video is viewed); and Video’s Comment Count (numbers comments received by a video). We also collected information about the users who posted the videos (i.e., User Age, Gender, User’s View Count (number of times a user’s profile is viewed), User’s Subscriber Count (the number of people subscribed to a user), User’s Join Date (i.e., account age), and total Videos Posted by a user).

Post Date (or age of the video) was calculated in terms of months, for example, age of a video was determined by subtracting the data collection date (i.e., 22 May, 2012) from the date the video was first posted on YouTube. In a similar fashion user’s account age was calculated. The Video’s Category is the YouTube classification of a video based on its contents.A poster can classify their video into 15 YouTube standard categories; however, our sample had videos only in six categories: Music, Comedy, Animation, Drama, health, and Ads. And, majority of the videos (75%) were in the Music category; therefore, the Video’s Category was coded as Music=1, and other=0. Gender wascoded as male equal to 1 and female equals to 2. Similarly, Age of the respondent was coded as 1= 18 to 29, 2= 30 to 49, 3= 50 to 69, and 4=70 to 89.

Furthermore, the data was also standardized (i.e., making its mean equal to 0 and variance equal to 1). Due to the multi-item nature of the dependent variable (i.e. Virality), the data was analyzed using structural equation modeling (SEM) technique. Particularly, Smart PLS software package (Ringle, Wende, & Will, 2005) was used for data analysis. PLS (partial least square) is a structured equation modeling technique that can analyze research models involving multiple-item constructs. PLS analysis was performed in two steps: (1) a test of the measurement model, an estimation of internal consistency (composite reliability), and determination of the convergent and discriminant validity of the instrument items; and (2) assessment of the structural model. One advantage of PLS is that is it less demanding on sample size.

In addition, for the studies investigating network phenomenon, relying only on conventional statistical tools may not be enough. Conventional statistical tools fill short in understanding a network phenomenon beyond the system boundaries. For example, to understand the network dynamics beyondthe YouTube system per se, such asthe role of network capital—external links pointing to the videos and hit count received outside YouTube's domain—invirality, we need to employ specialized network tools that can provide us a window into the network structures. In fact network capital inform of external links is linked to the popularity of videos (Cha, et al., 2007). Therefore, for a post-hoc analysis,in order to further strengthen our understanding of the viral phenomenon, we used Webometrics analysis technique(described below).

Webometrics Analysis

Webometricsanalysis is a well-established technique and is concerned with the measurement of different aspects of the World Wide Web (WWW), such as, web impact analysis, hyperlinks analysis, and web search engine results(Thelwall, 2005). Using Webometricsanalysis, we calculated network capitals of the YouTube videos and of the users who posted the videos. . Network capital of a user (or video) is defined as the number of external links and internet domains pointing to or mentioning the user (or the video). Therefore, we collected data related to the videos' network capital using Webometrics Analyst(formally known as "LexiURL searcher" is a well-known tool for analyzing the WWW; more details on the tool can be found here: http://lexiurl.wlv.ac.uk/). We collected a number of URLs of the pages pointing to a video and hit counts received by the videos outside YouTube systems i.e., over the internet. For example, the Table 5 lists the sites of pages (top 10) matching the base query: "*www.youtube.com/watch?v=kffacxfA7G4" -site: youtube.com"* (note that it is a video having an ID "kffacxfA7G4" posted by the user "Justin Bieber" on the YouTube). The URLs column lists the number of URLs returned by the query with the given site. The information was collected for all 100 video. Also, this data was used to construct network diagrams(i.e., Figure 2 and 3) for better understanding using UCINET social networking tool.

This may provide a good understanding of a videos influence or visibility outside the YouTube system (results are discussedin the later section).

<Table 5 about here>

Results

Descriptive Statistics

Of the 100 videos that went viral on YouTube, almost half of them were posted by Male users (49%)and 36 percent were posted byFemale users. The remainder (14%) did not disclose their gender. In some cases, gender was not relevant, for example, with videos posted by a company. Similarly, the majority of the 100 most-viewed videos were posted by users belonging to the 30 to 49 age group (67%) followed by users in the 18 to 29 age group (16%). A large majority of the most-viewed videos (78%) were posted by users in the USA, followed distantly by GreatBritain (GB) with 8% of videos. A large majority of the videos (75%) that went viral were in the music category, followed distantly by Comedy (12%), and Animation (8%).

Correlation Analysis

Table 1 shows Pearson Correlation among the variables used in the study. Likes and Dislikes received by video are mildly positively correlated with a correlation coefficient (r) of .268 significant at the level of 0.01. In other words, this means that there is a tendency that when one variable goes up the others go up as well. This can be interpreted as people preference for a particular video i.e., the split among the viewers' choice related to the videos in terms of likes and dislikes. It is interesting to note that a Video's View Count and Video's Dislikes have moderate positive correlation (r=.543, p <0.01). This suggests that when the Video's View Count goes up the rate of dislikes also go up. The same is also true for the correlation among Video's View Count and Likes rate (r=.685, p<0.01); however, the correlation among

Video’s View Count and Likes was slightly stronger than the relationship among Video’s View Count and dislike rate. This suggests that the more a video is viewed there is a tendency of receiving more likes than dislikes. It is also interesting to note that Video’s Comment Count and Dislikes have the strongest correlation among all the variables (r=.936, p<0.01). This suggests that the more comments on a video go up the rate of dislikes also goes up. Also, Video’sView Count has a moderate positive correlation to the User’s (profile’s) View Count (r= .418, p<0.01). It is not surprising to note that a strong positive correlation between Video’s Favourite Count and Likes (r=.825, p<0.01). Surprisingly, the correlation among the Video’s Published Date and Video Likes is negative (r=-.436, p<0.01). In other words, the older videos received less likes or the younger videos (i.e., posted recently) received more likes. The Video’sPublished Date is also negatively correlated with Video’s Favourite Count and Video’s View Count. This might be that the older videos went viral when there were fewer people on social media. Thus the older the video, the smaller its total views.

<Table 1 about here>

Test of Reliability and Validity

Correlation analysis is a great way to understand mutual association among the variables, but they cannot show cause and effect relationships. Therefore, we conducted SEM analysis to test reliability and validity of the measurement items.

For the *Virality*latent variable Cronbachs Alphawas 0.80and composite reliability was 0.9; the measures are robust and well above the recommended level of 0.70 (Nunnally 1978). The average variance extracted (AVE) value was 0.6; well above the accepted level of 0.50. The AVE, composite reliably, and Cronbachs Alpha were 1.0 for all other variables, because they are single item variables.

Table 2 shows inter variable correlations. Most of the correlations are low and pose no problem to the validity of the instrument. AVE indicates the reliability of the construct and allows the evaluation of discriminant validity. To indicate satisfactory discriminant validity, the AVE of the construct should be greater than the variance shared between the construct and other constructs in the model. The discriminant validity of the measurement scales is depicted in Table 2; the square roots of the AVEs were greater than the off-diagonal elements in their corresponding row and column in all cases, supporting the discriminant validity of our scales (the square roots of the AVE is shown only for the *Virality* (0.80) variable, for other variables it was 1.0 and not shown).

<Table 2 about here>

Similarly, Table 3 shows loadings and cross loadings of the items on its respective variables. Convergent validity is demonstrated when items load highly (loading> 0.50) on their associated factors. Convergent validity was fully demonstrated as all items loaded highly (Loading> 0.50) on their associated factors. The important thing to note is the loadings of the Virality variable's items, as the other variables are single item and loads fully on its respective variables.

<Table 3 about here>

Assessment of the Structural Model

Results of structural model carried out through SEM are show in Table 4, where the independent variables were user's characteristics and video's characteristics and the dependant variable was the Virality. Out of seven variables used in the model only three variables had significant effect on Virality (See Figure A). User's Total Videos had a moderate significant negative effect on Virality (β =-0.158, $p < 0.01$). In simple words, it means that as the number of total videos posted by a user increase, the Virality of a video posted by the user decreases proportionally. Video Category had a week significant positive effect on Virality (β =0.101, $p < 0.05$). In other words, the negative effect of video category on Virality shows that music videos are more likely to go viral (because it was coded as music=1 and other=0) as compared to the other YouTube categories.

Video Published Date had a moderate significant negative effect on Virality (β =-0.206, $p < 0.05$). To put it simply, this means that as the video published date (in months) increases (i.e., age of the video increases), the chance of a video to go viral decreases proportionally i.e., the video posted more recently gets more likes vice-versa. Overall, the User’s Characteristics and Video Characteristics accounted for 16% of variance in Virality (see Table 4).

<Table 4 about here>
<Figure A about here>

Webometrics Analysis Results

As noted in the SEM analysis, themajority of theindependent variables could not explainVirality ; therefore, in a post-hoc analysis, we tried to understand more about the viral phenomenon by studying network dynamics beyond the YouTube system using Webometrics analysis (as explained in the method section).

Based on the URLs and hit counts information explained in the method section, we constructed a correlation metrics for the number of URLspointingto a video, hits countsreceived over the Internet, and the virality variables as shown in Table 6. As shown in the Table 6, Hit and URLs count are highly correlated with the viral variables with an exception of Dislikes which is not correlated to any one of the variables. For example, Video’s Likes (VL) is strongly correlated with Hit count ($r=839$, $p<0.01$) and URL count ($r=825$, $p<0.01$); Video’sFavoriteCount (VFC) is also highly correlated with Hit count ($r=831$, $p<0.01$) and URL count ($r=827$, $p<0.01$); and Video’s View Count (VVC) is strongly is correlated with Hit count ($r=630$, $p<0.01$) and URL count ($r=640$, $p<0.01$). These correlations are positive and in most cases are much stronger than the correlation among the variables related to the YouTube system shown earlier inTable 2. This means that with increase/decrease in the number of URLs pointing to a video and hit counts received by a video outside the YouTube domain also increases/decreases viral phenomena (likes, views, comments, favourite count) over the YouTube system. Further, we did a nonparametric test to see if the distribution of Hits and URLs was the same across different video categories. In both cases, the results were highly significant with a level of 0.05 ($p<0.001$). It shows that different categories (music, comedy, drama, etc) received a different number of Hits counts and numbers of URLs. This might be one reason that number of videos in some categories, for example, in the music category,are more likely to go viral (e.g., 75% of the video that went viral were in the music category). These results are encouraging and support our assumption to look beyond the YouTube system to better understand the viral phenomenon.

<Table 5about here>

Network Diagram

In order to further explore the effects of external URLs on the viral phenomenon, we constructed a two mode network of the Internetdomains that were sending links to the YouTube videos and the YouTube users. Figure 2 shows the network of YouTube videos and the domain names pointing to YouTube videos. Figure 3 the network of YouTube Users and the domain names pointing to YouTube Users). In the Figure 2, the squares indicate unique videos IDs (videos IDs are automatically assigned when videos are posted over YouTube) and circle nodes indicate the domain names. Similarly, in the Figure 3, the squares indicate unique users IDs (users IDs are automatically assigned when users join/registerYouTube) and circle nodes indicate the domain names. The links among nodes are URLs (in-link) received by the videos and users from a specific domain (arrow heads pointing from a domain towards a video or user are removed for the sake of clarity). In the case of the videos IDs, the size of a node indicates the number of URLs (in-links) received by the videos (the size of a node is bigger when it received more URLs) and in case of the domain names it means the number of out-link the domains are sending to the videos (the size is bigger when a domain sent more URLs).Width of the links among the nodes indicates the number of URLs sent a by a domain to a video: width is bigger when more URLs are sent by a domain.

It is clear from the Figure 2 (domain name and video network) that most of the videos received URLs from a common set of domains (as shown by the circle nodes in the middle of the figure); however, the number of URLs received are different i.e., some videos received more URLs than others and some domains (e.g., com, net, and org) sent more links compared to others. The important point to note is the diversity domains receivedby some videos (as shown on the right side of the Figure 2). For example, the videos on the right side of the Figure 2 have not only attracted URLs from the common set of domains(e.g., com, net, and org), but also from several other diverse domains (e.g., tv, ca,sk, uk, be, tw, fi, cc). The differenceamong the videos receiving several diverse domain and links (as indicated by the size of the nodes on the right side of the Figure 2) vs., the videosreceiving limited number of links (shown on the left side of the Figure 2) isquite visible.

<Figure 2 about here>

A similar pattern is visible in the domain name and users network (Figure 3). Some users (the one on left side of the Figure 3) received several diverse domains both in numbers and diversity as compared to the other users (the one on the right side of the Figure 3). This analysis shows that apart from the popularity inside YouTube system the most popular videos (and users) had a strong in-links network (links received by videos and users) originating from diverse domains over the internet. In other words, the most popular videos and users on YouTube are quite visible over the Internet and these network dynamics or internet structural capital might be one of the reasons that the videos posted by certain users go viral (also evident in the correlation the Table 9).

<Figure 3 about here>

Discussion

In this research, we tried to understand why some videosposted over the YouTube goes viral. To this end, we collected information about the 100 all-time most-viewed YouTube videos and analyzed it using SEM and Webmoterics analysis. Some interesting findings are discussed below.

Surprisingly, total videos posted by a user and video published date had a moderate negative effect on Virality. In simple words, it means that as the number of total videos posted by a user increase, the chance of a video to go viral posted by the same user decreases proportionally. On the other hand, as the video published date (in months) increases (i.e., age of the video increases), the chance of the video getting viral also decreases proportionally i.e., the chance of a video getting viral posted more recently is higher and vice-versa. These findings are consistent with previous research on the YouTube (Cha, et al., 2007). This suggests that user/marketers interested inviral phenomenon should avoid posting a lot of videos, because this will negatively affect the chance of a video getting viral. This also suggests that YouTube fans/users are very selective in their choice of videos and do not just like content solely posted by famous users. Also, the results suggest thatsocial media users like fresh and new content more than old content (Cha, et al., 2007). Conventional wisdom dictates that the older the content is (i.e., published earlier) the greater it will received attention (i.e., get viral); however, this research shows that the opposite is true: the more recent the content, the greater the chances it has of receiving attention (i.e., getviral). This might be an evidence of decreeing return principle at work when it comes to social media content or the decay principle of famous contents (Cha, et al., 2007). In other words, as the content on social media becomes aged, the value/impact (in this case to get viral) is reduced with aging process. These findings are similar to a conclusion reached by Cha, et al. (2007) in their study of YouTube. These findings indicate that companies/individuals should frequently update their social marketing videos strategically.

However, most of the YouTube system variables in our research model could not explain the reasons behind a viral video posted over YouTube. Therefore, in a post-hoc analysis, we looked into the network dynamics beyond the YouTube systemusing Webometrics analysis. More specifically, we looked for the role of external network capital (such as, external links pointing to the videos/users and hit count received by a video/user outside the YouTube domain) and offline social capital, such as, existing fan base and fame in the offline world. Consistent with the network structure hypothesis (Bampo, et al., 2008a; Cha, et al., 2007; Ko, et al., 2008; Liu, et al., 2012), we found that popularity of the videos was not only the function of YouTube system per se, but network dynamics and offline capital play crucial role in the viral phenomenon. For example, upon a closer inspection, we found that the most popular videos (and user associated with the videos) were quite visible over the Internet attracting several URLs and diverse domains and that the majority of the users were well known celebrities, such as, Justin Bieber, Lady Gaga, and Shakira. Thus, the role of network dynamics and existing offline social capital cannot be ruled out in the viral phenomenon. The role of existing offline/online social capital(such as, fan base and fame)in viral phenomenon is something similar to what Allocca(2011) calls “Tastemakers”: the influentialindividuals whom people listen to (e.g., in terms of Twitter, they are the users with large numbers of followers). And the networks dynamics arethe Internet structural capital inform of in-links and hit counts(Broxton, et al., 2010).

From the data we also found that some video went viral quickly than others.Since no correlation was observed between the videos age and view counts, we suspected this may be due to the external network capital i.e., external links (and domains) pointing to the videos. To this end, wematched the number of domains attracted by video with the age of the videos, and found thatthe videos that attracted more diverse domain (e.g., the video at the right and upper-right side of the Figure 2) tends to go viral faster as compared to the videos with few diverse domains (e.g., the video at the left and bottom-left side of the Figure 2).This finding is interested and will help marketers to determine influential factor in the marketing campaign (West 2011) and to enhance brand advocacy and brand awareness (Kirby & Marsden, 2006). The important implication for viral marketers here is that network dynamics (i.e., internet structural capital inform of in-links and hit counts) and “Tastemakers”should be taken into account while formulating virality strategies. For example, linking the videos/contents posted over the YouTube in several external platforms (e.g., blogs, social network sites, and online discussion communities) and listing support from online influential users (e.g., a Tweet containing link to the video) may increase its chances of getting viral and getting viral faster.

Based on the findings of the study, we suggest that virality may take either positive or negative forms(Sen & Lerman, 2007; Wojnicki & Godes, 2008). For example, we found a strong correlation among the likes and dislikes received by video and the views count (e.g., see Table 1),suggesting that content virality on YouTube may take either a positive (i.e.,a video goes viral mainly because it is liked and appreciated) or a negative form (i.e., a video goes viral mainly because it is disliked and not appreciated). This finding has very important implications for the viral marketers, for example, strategies should be in place to encounter negative virality (the phenomenon that contentover YouTube is viral in a negative sense); failing to do somay hurt brand name. However, discussing such strategiesisbeyondthe scope of this study.Further research is needed to establish these findings beyond the correlation statistics and discussthe effects of negative virality on brand name and how to encounterit(Lee, et al., 2008). Referring tothe methodology employed in the study, we also suggest that while investigating social media related network phenomenon, relying only on conventional statistical tools may not provide a complete understanding of the phenomenon under investigation. As demonstrated in this study, conventional statistical tools if complemented with specialized network tools may provide richer insights into the social network phenomenon.

The results are interesting and in some cases surprising; however, it comes with certain limitations. For example, we only analyze 100 all-time most-viewed videos to study what drives viral behaviour; whereas, hundreds of videos are uploaded to the YouTube every day, for example, over 100 hours of video are uploaded to YouTube every minute(YouTube, 2013). Also, the videos analyzed were examples of successful or viral videos, perhaps a wider spread of sampled videos, including both viral and non-viral videos, could be more revealing and might be an interesting avenue for future research to explore. However, this limitation stems from the fact that YouTube APIs does not allow data for more than 100 videos to be extracted. Even though, the study has some limitations, we believe that we provided a good starting point to the researcher interested in viral phenomena over social media, for example, we proposed and empirically proofed a research model that can be enhanced and used on a more robust data in future research related to viral phenomenon.

Also, we only looked into the different categories of videos and not into the contents of the videos per se; which plays an important role in the viral phenomenon (Phelps, et al., 2004). For example, some of the viral videos were posted by ordinary users; however, all these videos had something in common i.e., either they captured the unexpected and/or had a funny touch. Thus, we believe that content of a video may also play an important role in the Virality. Further research is needed to look into the role of contents of a video in the viral phenomenon. We did not analyze the contents of the comments made by the others users in response to the videos posted over YouTube. Understanding effects of users' comments(specially the negative comments) on attitude toward product/service is very crucial(Lee, et al., 2008), for example, higher proportion of negative online consumer reviews on a product can have negative effect on the attitudes of consumers toward a product (Hennig-Thurau & Walsh, 2004; Lee, et al., 2008).Future research is needed to thoroughlyanalyze and categorize the user's comments and its effects on viral phenomenon over YouTube. This will further enrichour understanding of the role of users'sentimentsin viral phenomenon. Finally, we acknowledge that the data collected through Webometricstechnique (and other transaction log methods) has unique edge (e.g., it is larger and is collected in natural settings) over the conventional methods, such as, survey or laboratory settings, it also has several limitations (Jansen, et al., 2009). For example, the data may suffer from abstraction (difficulty in relating data to higher contexts), selection (difficulty in separating necessary from unnecessary), reduction (difficulty in reduction of size and complexity), context (difficulty in meaningful interpretation), and evolution problems(Jansen, et al., 2009).

Conclusion

The purpose of this research was to seek reasons for some content going viral on social media, specifically, the videos going viral over the YouTube. For this purpose, we constructed and tested an empirical model to understand the relationship among users' social and non-social capital, video characteristics, external network capital (in-links and hit counts), and Virality. Among other findings, our results showed that popularity of the videos was not only the function of YouTube system per se, but that network dynamics (e.g., in-links and hits counts) and offline social capital (e.g., fan base and fame) play crucial roles in the viral phenomenon, particularly view count. The study has important implications. For example, we for the first time constructed and tested an empirical model to find out the determinants of viral phenomenon over YouTube. By understanding virality characteristics of videos, the study will help marketers to determine influential factor in marketing campaign(West, 2011)and to enhance brand advocacy and brand awareness (Kirby & Marsden, 2006).

References

Adler, P. S., & Kwon, S.-W. (2002). Social Capital: Prospects for a New Concept. *The Academy of Management Review, 27*(1), 17-40.

Allocca, K. (2011). Why videos go viral Retrieved 05, 2012 (available at: http://www.ted.com/talks/kevin_allocca_why_videos_go_viral.html)

Bampo, M., Ewing, M. T., Mather, D. R., Stewart, D., & Wallace, M. (2008a). The effects of the social structure of digital networks on viral marketing performance. *Information Systems Research, 19*(3), 273-290.

Bampo, M., Ewing, M. T., Mather, D. R., Stewart, D., & Wallace, M. (2008b). The Effects of the Social Structure of Digital Networks on Viral Marketing Performance,. *Information Systems Research, 19*(3), 273-290.

Benevenuto, F., Duarte, F., Rodrigues, T., Almeida, V. A. F., Almeida, J. M., & Ross, K. W. (2008). *Understanding video interactions in youtube*. Paper presented at the Proceedings of the 16th ACM international conference on Multimedia, Vancouver, British Columbia, Canada.

Berger, J. A., & Milkman, K. L. (2009). What Makes Online Content Viral? *Working paper series, Available at SSRN: http://ssrn.com/abstract=1528077 or http://dx.doi.org/10.2139/ssrn.1528077*.

Brown, J., Broderick, A. J., & Lee, N. (2007). Word of mouth communication within online communities: conceptualizing the online social network. *Journal of Interactive Marketing, 21*(3), 2-20.

Broxton, T., Interian, Y., Vaver, J., & Wattenhofer, M. (2010, 13-13 Dec. 2010). *Catching a Viral Video.* Paper presented at the Data Mining Workshops (ICDMW), 2010 IEEE International Conference on.

Camarero, C., & José, R. S. (2011). Social and attitudinal determinants of viral marketing dynamics. *Computers in Human Behavior, 27*(6), 2292-2300.

Camarero, C., & San José, R. (2011). Social and attitudinal determinants of viral marketing dynamics. *Computers in Human Behavior, 27*(6), 2292-2300. doi: 10.1016/j.chb.2011.07.008

Cha, M., Kwak, H., Rodriguez, P., Ahn, Y.-Y., & Moon, S. (2007). *I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system*. Paper presented at the Proceedings of the 7th ACM SIGCOMM conference on Internet measurement, San Diego, California, USA.

Cheung, C., Lee, M., & Rabjohn, N. (2008). The impact of electronic word-of-mouth: The adoption of online opinions in online customer communities. *Internet Research, 18*(3), 229 - 247.

Chow, W. S., & Chan, L. S. (2008). Social network, social trust and shared goals in organizational knowledge sharing. *Information & Management, 45*(7), 458-465. doi: 10.1016/j.im.2008.06.007

Davis, A., & Khazanchi, D. (2008). An empirical study of online word of mouth as a predictor for multi-product category e-commerce sales. *Electronic Markets, 18*(2), 130-141.

Gladwell, M. (2002). *The Tipping Point: How Little Things Can Make a Big Difference*. Boston: Back Bay Books.

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? *Journal of Interactive Marketing, 18*(1), 38-52. doi: http://dx.doi.org/10.1002/dir.10073

Hennig-Thurau, T., & Walsh, G. (2004). Electronic word-of-mouth: motives for and consequences for reading customer articulations on the Internet. *International Journal of Electronic Commerce, 8*(2), 51-74.

Iyengar, R., Bulte, C. V. d., & Valente, T. W. (2011). Opinion Leadership and Social Contagion in New Product Diffusion. *Marketing Science, 30*(2), 195-212.

Jalilvand, M. R., & Samiei, N. (2012). The impact of electronic word of mouth on a tourism destination choice: Testing the theory of planned behavior (TPB). *Internet Research, 22*(5), 591 - 612.

Jansen, B. J., Taksa, I., & Spink, A. (2009). Research and Methodological Foundations of Transaction Log Analysis. In B. Jansen, A. Spink, & I. Taksa (Eds.)*Handbook of Research on Web Log Analysis* (pp. 1-16). Hershey, PA: Information Science Reference. doi:10.4018/978-1-59904-974-8.ch001

Kalish, S., Mahajan, V., & Muller, E. (1995). Waterfall and Sprinkler New-Product Strategies in Competitive Global Markets. *International Journal of Research in Marketing, 12*(2), 105-119.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons, 53*(1), 59-68. doi: 10.1016/j.bushor.2009.09.003

Khan, G. F. (2013). The government 2.0 utilization model and implementation scenarios. *Information Development,* online first, doi: 10.1177/0266666913502061

Khan, G. F., & Swar, B. (2013, June 4). *Government 2.0: Utilization Model, Implementation Scenarios, and Relationships.* Paper presented at the Pre-ECIS (21st European Conference on Information Systems) workshop: E-Government 2.0: Case studies and experience reports, Utrecht, Netherlands.

Kirby, J., & Marsden, P. (2006). *Connected Marketing: the Viral, Buzz and Word-of-mouth Revolution*. Oxford, UK: Butterworth-Heinemann.

Kleijnen, M., Lievens, A., de Ruyter, K., & Wetzels, M. (2009). Knowledge Creation Through Mobile Social Networks and Its Impact on Intentions to Use Innovative Mobile Services. *Journal of Service Research, 12*(1), 15-35. doi: 10.1177/1094670509333238

Ko, H.-C., Yin, C.-P., & Kuo, F.-Y. (2008). Exploring individual communication power in the blogosphere. *Internet Research, 18*(5), 541 - 561.

Lange, P. G. (2007). Publicly Private and Privately Public: Social Networking on YouTube. *Journal of Computer-Mediated Communication, 13*(1), 361-380. doi: 10.1111/j.1083-6101.2007.00400.x

Lans, V. d., Ralf, G. v. B., Eliashberg, J., & Wierenga, B. (2010). A Viral Branching Model for Predicting the Spread of Electronic Word-of-Mouth. *Marketing Science, 29*(2), 348–365.

Lee, J., Park, D. H., & Han, I. (2008). The effect of negative online consumer reviews on product attitude: An information processing view. *Electronic Commerce Research and Applications, 7*(3), 341.

Libai, B., Muller, E., & Peres, R. (2005). The Role of Seeding in Multi-Market Entry. *International Journal of Research in Marketing, 22*(4), 375-393.

Liu, Z., Liu, L., & Li, H. (2012). Determinants of information retweeting in microblogging. *Internet Research, 22*(4), 443 - 466.

Mangold, W. G., & Faulds, D. J. (2009). Social media: The new hybrid element of the promotion mix. *Business Horizons, 52*(4), 357-365. doi: 10.1016/j.bushor.2009.03.002

Marcus, A., & Perez, A. (2007). m-YouTube Mobile UI: Video Selection Based on Social Influence. In J. Jacko (Ed.), *Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments* (Vol. 4552, pp. 926-932): Springer Berlin Heidelberg.

Mills, A. J. (2012). Virality in social media: the SPIN Framework. *Journal of Public Affairs*, *12*(2), 162-169. doi: 10.1002/pa.1418

Pastore, M. (2000). The Value of Word of Mouth, from http://adres.intemet.com/feature/article/0,1401,8961_395371,00.htm]

Phelps, J. E., Lewis, R., Mobilio, L., Perry, D., & Raman, N. (2004a). Viral Marketing or Electronic Word-of-Mouth Advertising: Examining Consumer Responses and Motivations to Pass Along Email. *Journal of Advertising Research, 44*(04), 333-348. doi: doi:10.1017/S0021849904040371

Phelps, J. E., Lewis, R., Mobilio, L., Perry, D., & Raman, N. (2004b). Viral Marketing or Electronic Word-of-Mouth Advertising: Examining Consumer Responses and Motivations to Pass Along Email. *Journal of Advertising Research, 44*, 333-348.

Porter, L., & Golan, G. J. (2006). From Subservient Chickens to Brawny Men: A Comparison of Viral Advertising to Television Advertising. *Journal of Interactive Advertising, 6*(2), 30-38.

Pousttchi, K., & Wiedemann, D. G. (2007, 9-11 July 2007). *Success Factors in Mobile Viral Marketing: A Multi-Case Study Approach.* Paper presented at the Management of Mobile Business, 2007. ICMB 2007. International Conference on the.

Ringle, C. M., Wende, S., & Will, S. (2005). SmartPLS 2.0 (M3) Beta, Hamburg <http://www.smartpls.de/forum/index.php>.

Sen, S., & Lerman, D. (2007). Why are you telling me this? An examination into negative consumer reviews on the web. *Journal of Interactive Marketing, 21*(4), 76-94.

Shao, G. (2009). Understanding the appeal of user-generated media: a uses and gratification perspective. *Internet Research, 19*(1), 7 - 25.

Subramani, M. R., & Rajagopalan, B. (2003). Knowledge-sharing and influence in online social networks via viral marketing. *Commun. ACM, 46*(12), 300-307. doi: 10.1145/953460.953514

Thelwall, M. (2005). Webometrics. In M. A. Drake (Ed.), *Encyclopedia of Library and Information Science* (Second ed.). New York: Marcel Dekker, Inc.

Vilpponen, A., Winter, S., & Sundqvist, S. (2006). Electronic Word-of-Mouth in Online Environments: Exploring Referral Network Structure and Adoption Behavior. [Article]. *Journal of Interactive Advertising, 6*(2), 71-86.

West, T. (2011). Going Viral: Factors That Lead Videos to Become Internet Phenomena. *The Elon Journal of Undergraduate Research in Communication, 2*(1).

Wojnicki, A. C., & Godes, D. (2008). Word-of-Mouth as Self-Enhancement. *HBS Marketing Research, Paper No. 06-01, Available at SSRN: http://ssrn.com/abstract=908999 or http://dx.doi.org/10.2139/ssrn.908999*.

YouTube. (2013). YouTube Statistics, available at: http://www.youtube.com/yt/press/statistics.html