Rising to Stardom: An Empirical Investigation of the Diffusion of User-generated Content

Yuping Liu-Thompkins [a],[*] & Michelle Rogerson [b],[1]

[a] College of Business and Public Administration, Old Dominion University, 2040 Constant Hall, Norfolk, VA 23529, USA
[b] Reina Communications, P. O. Box 9436, Virginia Beach, VA 23450, USA

Available online 1 February 2012

Abstract

With the explosive growth of online user-generated content and the desire by marketers to better utilize this space, it is beneficial to understand the viral diffusion of such content and to identify messages that are most likely to achieve popularity. In this paper, we combine network analysis and the diffusion literature to study the spreading of user-generated videos online. We identify three groups of factors that affect diffusion outcomes: network structure, content characteristics, and author characteristics. Using a proportional rates model, we analyze the diffusion of a sample of videos on YouTube. Our results show that it is preferable to have many subscribers who each has a few friends than to have a few subscribers with many connections. Furthermore, a curvilinear relationship exists between subscriber network connectivity and diffusion rate such that diffusion is at its highest under moderate connectivity. Examining content characteristics, we show that entertainment and educational values affect diffusion but production quality does not matter. Moreover, we find that quality as manifested by user ratings influences diffusion more than innate content quality. Not surprisingly, an author's past success carries over to the current content, and content from younger authors is more popular.
© 2012 Direct Marketing Educational Foundation, Inc. Published by Elsevier Inc. All rights reserved.

Keywords: User-generated content; Diffusion; Network analysis; Viral marketing; Hazard modeling

Introduction

Along with the Web's transition into a participative environment, consumers have gained influence and credibility as content creators in their own right. They are now contributing to the online universe in a wide variety of ways, including but not limited to blogs, podcasts, videos, online social networks, games, mashups, and user reviews. Such user-generated content (UGC) has important implications for marketers for several reasons. First, UGC pools the ideas of a vast global array of talent, and the cost to facilitate collaboration through UGC is low (Tapscott 2007). This aids UGC's widespread and rapidly accelerating popularity. Furthermore, research shows that creators of UGC are likely to be important brand advocates who share opinions about products and services with others (Luetjens and Stanforth 2007). And because creators of UGC are viewed as "one of us," this makes UGC more influential than traditional

marketing. A December 2008 study by Leo Shapiro and associates showed that consumers find sources such as discussion forums, blogs, online reviews, and social networks three times more influential than TV advertising when making a purchase decision (Leo Shapiro and associates 2008).

A major challenge in utilizing UGC, however, lies in the sheer amount of UGC available and the difficulty in identifying the more valuable pieces that are likely to make a real impact on businesses and other consumers. In reality, every consumer can contribute to the democratic Web. But only some UGC ends up in the popular domain to affect a large number of consumers, while other UGC is left in oblivion, hardly known by anyone else but the original poster and perhaps a small circle of friends.

To businesses looking to "ride the wave" of UGC, therefore, a key task would be to identify a UGC's likelihood of success and to understand what makes a UGC extremely popular while others are not. This understanding will help marketing managers predict

* Corresponding author. Tel.: +1 757 683 6551; fax: +1 646 716 3900.
 E-mail addresses: YXXLiu@odu.edu (Y. Liu-Thompkins), michellerogerson@gmail.com (M. Rogerson).
[1] Tel.: +1 757 589 6404.

the success of UGC components of their own marketing campaigns where consumer participation and contribution are involved (e.g., a video contest). Marketing managers can also learn from the diffusion of UGC to design their own viral marketing campaigns more effectively. In the scenario of a public communication crisis in which a UGC features negative commentaries about a brand, being able to predict the success of the UGC can be critical to formulating a proper response. Ultimately, understanding what drives the success of any UGC will help businesses participate more effectively in the new Web environment.

Addressing this potential, we build on network science and social network analysis to examine the viral diffusion of UGC online. Within the context of online user-generated videos, we trace the pattern of rise to popularity of such content over time and address the factors that affect the diffusion path of new UGC. We accomplish this by modeling aggregate-level diffusion rate on individual network properties and UGC content and author characteristics. In doing so, we provide a mechanism for predicting the success of a UGC based on readily available data at an early stage of the diffusion process. Marketers equipped with such knowledge will then be able to strategically participate in the conversation facilitated through UGC and to better focus their resources on viral content that is likely to affect a larger audience.

The rest of the paper is organized as follows. We first review past research that has incorporated network structure into diffusion modeling, and we identify a few key differences between UGC diffusion and the diffusion of a new product or innovation. We then define a set of network, content, and author-related characteristics that can affect the popularity of a UGC and incorporate these characteristics into a proportional means/rates model. Next, we report the results from our model using data from YouTube, a popular site for consuming and sharing user-generated videos. Finally, we discuss the implications of our research for theory and practice and suggest a set of future research questions.

Background

A large amount of research has been conducted on new product diffusion and its modeling. Rather than going to great length with this literature, we refer interested readers to several excellent past reviews of this topic (Mahajan, Muller, and Bass 1990; Meade and Islam 2006; Muller, Peres, and Mahajan 2009). In this section, we focus more narrowly on the interplay between network properties and diffusion. Given the important role contagion plays in spreading new products and ideas, it is reasonable to argue that network-oriented approaches such as network science and social network analysis are a natural fit to the study of diffusion. While social contagion has long been considered an important force in diffusion, as exemplified by seminal works by Bass (1969) and Rogers (1962), studies that explicitly incorporate concrete network properties into diffusion models are still rare (Valente 1995, 2005). This may be partially due to the cost of collecting social network data in the traditional environment (Van den Bulte and Wuyts 2007). The Internet and the emergence of online social networks, however, have made the task more

feasible and have led to more work in this area in recent years. For instance, using data from a Korean social networking website, Goldenberg et al. (2009) projected individual network property (how many connections an individual has) onto the aggregate diffusion process in a modified Bass model. They found well-connected individuals to have a disproportionate influence on diffusion speed and final adoption level. In another study, Katona, Zubcsek, and Sarvary (2011) considered a large number of network structural properties in the diffusion of a Central European online social network. Using an individual-level hazard model, they confirmed the effects of these network characteristics on individuals' adoption likelihood.

A few other recent studies have also taken network structure into consideration, although structural properties are implicitly assumed in these studies rather than explicitly observed in the data. Toubia, Goldenberg, and Garcia (2009) extended existing aggregate-level diffusion models by considering the number of recommendations consumers sent and received about a new product. They incorporated network structure (network size) as a parameter in the probabilistic recommendation generation process and estimated network size together with the entire model. Van den Bulte and Joshi (2007) devised an asymmetric influence model to take into account the fact that some consumers have the ability to influence others ("influencers") whereas other consumers are "imitators" that do not exert a reciprocal impact on the influencers. In network terms, their model essentially captures the directionality of social ties. But they do not observe actual ties and instead estimate the proportion of consumers that belong to each segment using aggregate-level diffusion data.

One central theme in these studies is that incorporating network structure (either explicitly or implicitly) into diffusion models improves the performance of such models. As interpersonal influence among consumers accelerates and becomes more widespread with the help of online social networks, there is much to gain from understanding how consumers' network ties may play into the diffusion process (Van den Bulte and Wuyts 2007). Following this emerging line of research, the current study aims to incorporate social network structure into explaining and predicting the diffusion of UGC online. Within the framework of a recurrent event model, we use UGC originator and subscriber network properties as well as content quality cues to model the diffusion speed of UGC. Different from previous studies, we estimate simultaneously the diffusion of a set of UGC and thereby explain the relative success of one UGC vs. another. Our research also enriches the network science literature by providing empirical evidence of the diffusion path of information in online social media. Previous network research has often utilized computer simulations to address similar questions (e.g., Watts and Dodds 2007). Using real-world data from online social networks, we supplement simulated studies with empirical evidence and offer clues to the true nature of network influence within a social media environment.

Network Effect in UGC Diffusion

Before we move on to describe network and other factors that we believe to affect the diffusion of a UGC, we would like to

briefly discuss the unique nature of UGC that sets it apart from new product diffusion and that may accentuate network effects in the diffusion process. First, UGC typically does not engage in intentional promotion to the mass audience, and mass media exposure usually does not occur until a UGC has already become popular. Therefore, UGC diffusion relies heavily on voluntary sharing among consumers, accentuating the importance of individual social networks. Moreover, the adoption of UGC typically involves very low personal risk. The only visible cost for consuming most UGC is the time and effort involved in the consumption activity itself. This low risk and low cost nature implies that the adoption threshold for UGC is likely to be low, making it easier to spread through the grapevine. It also means that the need to mitigate risk through interpersonal communication is relatively trivial and that awareness through wide reach is likely more important than disproportionate influence through opinion leaders (Watts and Dodds 2007).

At the same time, the democratic nature of UGC means any user can contribute to the universe. As a result, the quality of UGC can vary widely, and a large number of UGCs will probably be of very low quality. This reduces the value of UGC, acting as a counterforce to the aforementioned low risk effect and potentially deterring the diffusion of UGC. Relatedly, unlike content from well-known information sources such as major newspapers, the authoritativeness of UGC remains largely unknown to most individuals. This lack of information increases the externalities of UGC consumption and makes individuals more subject to the opinions of others (Asch 1953), increasing the reliance on network influences.

Factors Affecting UGC Diffusion

Network Sizes

We define a UGC author i's network of directly connected individuals as its first-level network. With direct ties to the UGC creator, these individuals can be expected to be the ones most likely to know the existence of a UGC, consume the UGC, and function as its evangelists in the diffusion process. The size of this first-level network, referred to as "degree" in network analysis, is critical to the initial consumption of the UGC (Van den Bulte and Wuyts 2007). In this research, we are interested in one specific type of first-level network as defined by subscriptions. UGC communities such as YouTube and Twitter allow individuals to subscribe to or follow other users' activities. These types of ties define a directional connection to a UGC author from his/her base of subscribers, and they facilitate the direct flow of information and influence from the author to these subscribers. In network science, this is called the "indegree" of an individual node, and it is considered a measure of one's prestige and popularity and hence one's ability to influence others (Wasserman and Faust 1994). We use N_i to denote the indegree of UGC author i.

Beyond initial adoption by the first-level network, these individuals can start a contagion process that spreads the UGC beyond the small network of the original author and his/her direct subscribers. One factor that can affect this contagion process is

the network size of each of the first-level nodes, denoted as S_{ij} for the jth individual in UGC author i's first-level network, and $\bar{S}_i$ for the average network size (or mean degree in network terms) across all first-level nodes. While this network size defines the population that a first-level node can directly impact, network science suggests a less-than-straightforward relationship between it and contagion outcomes. On one hand, the more individuals who are connected to a first-level node j, the more people j can potentially influence. On the other hand, individuals with a large network size have been found to have on average weaker connections, resulting in less influence on individuals who are connected to them (Katona, Zubcsek, and Sarvary 2011). These two aspects imply potentially opposing effects between network size and contagion outcomes, which may partially explain the lack of relationship found in past research between network size and influence outcomes (Trusov, Bodapati, and Bucklin 2010). In the context of UGC, as we discussed earlier, the low-risk nature of UGC consumption may favor the wider reach created by a large network size despite weak influences. In the meantime, the experience of low value provided by a large amount of low-quality UGC may favor contagion through greater influence. In other words, one is more likely to view UGC that is passed on by someone he or she knows well. Due to the presence of these countering forces, we leave open the question of how the network size of first-level nodes will exactly impact UGC diffusion.

Network Connectivity

Another network property we consider is how well the first-level network is connected. In network theory, connectivity (or sometimes called density) of a network is defined as the total number of present links as a proportion of the maximum number of links possible (Scott 1991). In a highly connected network, everyone is connected to almost everyone else, and information residing at any node can be transmitted to other parts of the network through many routes. In the meantime, such a network also tends to be highly limited to within the group (Watts 2003). While information may transmit quickly and reliably *within* the network, it is often unable to travel very far beyond the small network. In a weakly connected network, in contrast, information transmission depends on a small number of links, which creates instability and affects the success of diffusion. However, weakly connected networks can be critical bridges between networks (Granovetter 1973) and can help information travel further. Combining the respective strengths and weaknesses of densely vs. weakly connected networks, for widespread diffusion to occur for a UGC, a proper balance between high connectivity and low connectivity needs to be present (Watts 2003). What this implies is an inverted-U shaped relationship between network connectivity and UGC success. At the peak of the inverted-U shape is what is defined as a "cascading window" (Watts and Dodds 2007), where widespread diffusion is most likely to occur and contagion effect will be the strongest.

One difficulty in defining connectivity in a UGC environment is that friendship is usually not explicitly observed as in social networks such as Facebook. Even when friendship ties are overt, oftentimes these observed ties have a lot of noise (e.g.,

adding a lot of friends as a popularity contest or accepting friendship out of politeness), and they contain limited information on consumers' true influence on each other (Trusov, Bodapati, and Bucklin 2010). An alternative would be to use subscription to each other to define connectivity. While this is sufficient for defining connections to a UGC author, many average users (subscribers) merely consume rather than contribute UGC of their own, and subsequently the incoming subscriptions for these users will be minimal or non-existent. Simply relying on subscription to each other to define connectivity among these users misses out the full picture.

To solve this issue, we draw from the idea of affiliation network from social network analysis to define network connectivity (Wasserman and Faust 1994). The basic idea of affiliation network is that in the absence of explicit tie information, one can infer the tie between two nodes based on the contexts they belong to. The more contexts two individuals mutually belong to, the more likely the individuals will have crossed path at some point and therefore will have a connection. Inferring social ties using this affiliation network approach has been implemented in online social network studies when explicit social ties are unknown (e.g., Provost et al. 2009). In the current setting, we use the full set of subscriptions by all first-level nodes instead of simply subscriptions to each other, and we define connectivity based on subscriptions in an affiliation network fashion. Specifically, we identify the connection among users based on the common subscriptions that they share, assuming that users who share a large set of common subscriptions are more likely to know each other, through mechanisms such as group-based discussions or interaction on their common channels. This operationalization is based on past findings from social network analysis that individuals with common tastes and values are more likely to be socially connected (McPherson, Smith-Lovin, and Cook 2001). Mathematically, let a_{jg} be a binary variable indicating individual j's subscription status to node g. We can then define a valued rather than binary tie between individuals j and k as:

$$tie_{jk} = \sum_{g=1}^{G} a_{jg} a_{kg} \qquad (1)$$

where G represents the full set of subscriptions by all first-level network nodes. Based on this valued definition of tie, one can then calculate the connectivity of a UGC author *i*'s network as in Eq. (2) (Wasserman and Faust 1994), and this is the formulation we will use to model the effect of network connectivity on UGC diffusion.

$$Conn_i = \frac{\sum_{j=1}^{S(i)} \sum_{k=1}^{S(i)} tie_{jk}}{G*(G-1)}. \qquad (2)$$

Content Quality — Innate vs. Manifested

It is hard to dispute that the quality of a UGC will affect its eventual success. A high-quality UGC increases likelihood of consumption and the possibility that a user will want to share the content with others. We argue that it is important to distinguish between two types of UGC quality: innate vs. manifested. Innate quality refers to the actual quality of a UGC as judged by the viewer. As experience goods, this quality cannot be determined until a UGC has been consumed (Nelson 1970). Given the community features available on many UGC websites, however, there is additional quality information available, in the form of what we call manifested quality. This is quality information as judged by other users who have already consumed a UGC and is available to the current user before actually consuming the UGC, usually through some public rating or commenting system. Utilizing this manifested quality information resembles vicarious learning (Bandura 1977), where an individual observes and learns from another individual's experience.

The need to distinguish between innate vs. manifested content quality is driven by three reasons. First, while a UGC's innate quality is a constant to an individual, the amount and level of manifested quality varies with time. Depending on when a visit occurs, this quality information can be unavailable or the ratings can change over time. Second, unlike commercial product reviews, consumers are likely to feel more lenient when it comes to judging the quality of content created by other users. Consequently, quality ratings can be higher than actual quality, or negative ratings may not always be reported. Both of these can create an upward bias in manifested quality and a mismatch between innate and manifested quality. Lastly, the need to distinguish between innate vs. manifested quality also rests on the differential impact they may have on the diffusion of UGC. Manifested quality as ratings can affect both a consumer's own decision to consume a UGC and a consumer's decision to pass on the UGC to friends. Therefore, it exerts an influence on both first consumptions as well as subsequent contagion likelihood. Innate quality, on the other hand, is only available after consumption and therefore only enters into a consumer's decision to pass on a UGC. For the above reasons, we will include these two types of quality information separately in our model.

Other Covariates — UGC Author Characteristics, Traffic Source, and Content Category

Even before a UGC's quality information is available, another way of judging the potential value of a UGC may be to look at the UGC author's past success. If an author is already known for generating popular content, it is more likely that the author's subsequent UGC will be welcomed. In other words, a UGC author's past success can increase the adoption likelihood of his/her future content. Although this is not the primary focus of our research, we would like to take into account the effect of such factors. Specifically, we consider two aspects of a UGC author's experience: the volume of past contribution and the popularity of past contribution. We also include the author's gender and age, thus allowing gender and age differences in content diffusion.

In addition, because a video can be discovered from sources other than the author's direct subscribers (e.g., through recommendation of related videos or by embedding on other websites), we control for this influence by considering other sources that had led to significant traffic to the video. YouTube lists such

information as "significant discovery events," which includes the date on which the first referral from each significant source occurred and the number of views from each source. We count the total number of views from such events during the 60-day period (*OtherViews*) and include it in our model. Another factor that we control for is the category that a UGC is posted under. As some topics are inherently more appealing than others (e.g., entertainment content may inherently appeal to more people than engineering-related content), a UGC's category can affect its diffusion potential.

The Data

Sample

To examine empirically the diffusion of UGC, we use data from YouTube, a leading online community for posting, watching, and sharing user-generated videos. Recent data show that this online community received about 97 million viewers and 5.1 billion video views each month (an average of more than 52 videos per viewer) (comScore 2009). As a hub for UGC, more than 10 h of video is uploaded onto YouTube each minute (YouTube 2009), making this an ideal environment for studying UGC diffusion. We drew our initial pool of 140 videos in early October 2009. This was done over the course of seven days in order to avoid systematic bias that may be associated with a particular day of the week. Each day, we drew a random sample of 20 user-generated videos from the list of new videos that were added to YouTube on that day. For each video sampled, we collected information about the video poster's network structure, past experience, and demographics.

We then tracked each video on a daily basis over the course of two months between October and December 2009. Every day, we recorded the number of cumulative views for each video and the ratings for each video. Over the course of two months, some videos were removed by the user, and some were removed by YouTube due to rights violations. These removals affected 32 videos in our content pool, leaving a final sample size of 108 videos. These videos were dispersed among 13 content categories as classified by YouTube. The top three most popular categories were music, entertainment, and people and blogs, and the least popular category was science and technology. Table 1 shows descriptive statistics for the sample videos and their authors.

Content Quality

As we discussed earlier, UGC content quality can be of two types: innate quality and manifested quality as indicated by user ratings. We collected manifested quality information (i.e., average rating for each video) with diffusion information each day. To measure the innate content quality of each video, we recruited a convenience sample of 108 individuals to rate the videos. These individuals' age ranged from 18 to 65, with the median age being 33. Males accounted for 34.6% of the sample, and females 65.4%. When asked their frequency of visit to YouTube, 45.1% of the raters reported visiting YouTube several times a week, 32.9% a few times a month, 11% less than once a month, and

6.1% everyday. Another 4.9% of the raters do not visit YouTube on a regular basis. Following past research showing that both utilitarian and hedonic values of a viral message increase consumers' intention to forward the message (Chiu et al. 2007), we asked the respondents to assess each video on its entertainment value and educational value. We also asked them to rate the videos on production quality to account for the possibility that it may influence consumers' perception of the video poster's expertise and credibility, which may subsequently affect voluntary spreading of the video (Dobele, Toleman, and Beverland 2005). All three measures were taken on 10-point semantic differential scales, with 1 being the lowest and 10 being the highest. To avoid fatigue, each individual was asked to watch and rate a random sample of five videos, and each video was rated by five individuals. The average correlation among raters was .87, suggesting a fairly high level of inter-rater agreement (Neuendorf 2002). The ratings were averaged across the five raters to create three innate quality scores for each video.

The Model

As the number of views for some user-generated videos very likely contains multiple views from the same individual, traditional diffusion models with the assumption of single purchases do not work. While diffusion models taking into account repeat purchase have been developed in the marketing literature (Ratchford, Balasubramanian, and Kamakura 2000), they often assume an underlying product lifetime or product failure rate that drives consumers to repurchase. This systematic pattern of purchase and replacement may not apply to UGC diffusion, as the casual nature of UGC consumption can make repeat viewing/consumption of content quite sporadic. Therefore, here we use the proportional rates/means model developed in the biometrics field (Lawless and Nadeau 1995; Lin et al. 2000; Pepe and Cai 1993). Rooted in counting process theory, the proportional rates/means model is a semiparametric model for studying event recurrence. Using a multiplicative formulation similar to proportional hazard modeling, it offers an efficient and parsimonious way of capturing the effects of covariates and at the same time provides a mechanism for inferring the mean function of the view counting process. Compared with other recurrent event models, the proportional rates/means model has the advantage of allowing arbitrary and complex dependence structures among recurrences, therefore in our case permitting future views of a video to be dependent on past events in many ways.

Specifically, let $View_i(t)$ be the cumulative view of video i up to time t, and let $dView_i(t)$ be the increment in views over a small time interval $[t, t+dt]$. The rate function of the counting process is defined as the expectation of $dView_i(t)$ given the observable history of the video and a set of covariates that may affect recurrence. This is shown in Eq. (3):

$$dR_i(t) = E[dView_i(t)|H_i(\tau), 0 \leq \tau \leq t; X_i(t)] \qquad (3)$$

where $H_i(\tau)$ represents the observed history of video i up to time τ, and $X_i(t)$ is a vector of covariates that can be time-independent and/or time-varying. Similar to the proportional hazard

Table 1
Descriptive statistics of the videos and authors.

Variables	Mean	STD	Min	Max	2	3	4	5	6	7	8
1. Final cumulative views	150	178.28	24	41,015	.46	−.25	−.04	.20	.14	.27	.01
2. Number of subscribers	51	108.93	2	27,263		−.04	−.21	.30	−.003	.03	.18
3. Average network size of subscribers	246	504.63	0	3338			.20	−.07	.04	.01	.22
4. Subscriber network connectivity	.22	.34	0	1				−.40	−.09	.01	.003
5. Past videos posted	20	7.96	2	26					.02	.06	.23
6. Average views of past video	276.75	795.39	.68	7171						.05	.06
7. Views from other sources	78	107.70	9	11,984							.04
8. Author age	27	15.43	13	91							
9. Author gender	23.6% females, 77.4% males										

formulation, the proportional rates model presents the recurrence rate in a multiplicative form as:

$$dR_i(t) = \exp(\beta' X_i(t)) * dR_0(t) \tag{4}$$

where $R_0(t)$ is an unspecified continuous function representing the baseline rate. It follows from Eq. (4) that the expected cumulative view for video i at time t is:

$$View_i(t) = \int_0^t \exp\left(\beta' X_i(u)\right) * dR_0(u). \tag{5}$$

Based on our earlier discussion, we include four groups of covariates (X) into the model: (a) network properties, which capture the network structure (size and connectivity) of the UGC author and his/her immediate network of subscribers; (b) content characteristics, which include the innate and manifested content quality and the content topic category; (c) author characteristics, which include past experience and demographics; and (d) traffic impact from non-network sources such as site search. A detailed description of the covariates is provided in Table 2.

Model estimation

The proportional rates/means model can be estimated using a partial likelihood approach. The data used for estimation contain $MV_i + 1$ observations for a given video i, where MV_i is the maximum number of views observed for the video. The one extra observation at $MV_i + 1$ is based on the assumption that there is a possibility for additional recurrence (i.e., additional views), but it simply has not happened yet by the end of the observation period. Therefore this observation is considered to be censored. Let $Y_i(t)$ indicate whether video i is still under observation at time t. Assuming that the videos' diffusion processes are independent from each other, the partial likelihood function can be defined as:

$$L(\beta) = \prod_{i=1}^{N} \int_0^T \left[\frac{Y_i(t)\exp(\beta' X_{it})}{\sum_{j=1}^{N} Y_i(t)\exp(\beta' X_{jt})} \right]^{\Delta View_{it}} \tag{6}$$

Table 2
Covariates included in the proportional rates model.

Covariates	Description
Network properties	
Sub_i	The size of author i's first-level network as measured by the total number of subscribers.
$\bar{S}_i$	The network size of each of author i's subscribers averaged across all subscribers.
$Conn_i$	The connectivity of author i's subscriber network as defined by Eqs. (1) and (2). A quadratic term of this is also included to reflect its expected curvilinear effect.
Content characteristics	
QI_{ij}, $j=1, 2, 3$	The three innate content quality measures from the survey, namely production quality, entertainment value, and educational value.
$QM_i(t)$	The manifested quality of video i at time t, which is the average public rating of the video on YouTube at that time. To accommodate videos that do not have any ratings, we operationalize this as a relative sentiment measure that equals the actual average rating minus three (the neutral point). For videos and time periods that do not have rating information available, we set the variable as 0 representing a neutral impact on consumers.
$Category_{ik}$, $k=1$ to 12	Twelve dummy variables for the thirteen content categories represented by the sample. The most popular category, "Music", was set as the benchmark.
UGC author characteristics	
Vol_i	The total number of videos posted by author i prior to posting the sample video.
$AvgView_i$	The average views across all of author i's past videos.
Age_i	Author i's age.
$Gender_i$	Author i's gender.
Traffic impact from other sources	
$OtherViews_i$	The total number of views that came from non-network related sources such as site search and related videos.

Note: As there is a one-to-one correspondence between videos and authors, we use the subscript i to index both a video and its corresponding author.

where N is the number of videos in the sample, T is the maximum observation period, and $\Delta\text{View}_{it} = 1$ if $\text{View}_{it} - \text{View}_{it-1} > 0$ and 0 otherwise. Essentially, the partial likelihood function is defined over videos under observation and that have experienced a change in views (i.e., recurrence) during time t.

While the partial likelihood function in Eq. (6) is defined the same way as when the observations for each video are independent, Lin et al. (2000) show that the estimates derived from this partial likelihood function are efficient and consistent as long as Eq. (4) holds. However, as the recurrent views for the same video are allowed to follow any arbitrary dependence structure, the covariance matrix for the parameter estimators needs to be computed differently using a robust sandwich approach (Lin et al. 2000). Let $\hat{\Omega}\left(\hat{\beta}\right)$ be the Hessian matrix of the log-likelihood (i.e., observed information matrix) evaluated at the maximum likelihood estimator $\hat{\beta}$, the robust covariance matrix $\hat{\Sigma}\left(\hat{\beta}\right)$ is given by:

$$\hat{\Sigma}\left(\hat{\beta}\right) = \hat{\Omega}\left(\hat{\beta}\right)^{-1}\left[N^{-1}\sum_{i=1}^{N}\hat{W}_i\left(\hat{\beta}\right)\hat{W}'_i\left(\hat{\beta}\right)\right]\hat{\Omega}\left(\hat{\beta}\right)^{-1} \quad (7)$$

where

$$\hat{W}_i\left(\hat{\beta}\right) = \int_0^T\left\{X_i(t) - \bar{X}\left(\hat{\beta}, t\right)\right\}dM_i(t) \quad (8)$$

$$\bar{X}\left(\hat{\beta}, t\right) = \frac{\sum_{i=1}^{N}Y_i(t)X_i(t)\exp\left[\hat{\beta}'X_i(t)\right]}{\sum_{i=1}^{N}Y_i(t)\exp\left[\hat{\beta}'X_i(t)\right]} \quad (9)$$

and

$$M_i(t) = \text{View}_i(t) - \int_0^t Y_i(t)\exp\left[\hat{\beta}'X_i(t)\right]dR_0(t). \quad (10)$$

As the robust estimator needs the baseline rate as the input, it is estimated using the Aalen–Breslow estimator as:

$$R_0(t) = \int_0^t\frac{1}{N\sum_{i=1}^{N}Y_i(t)\exp\left[\hat{\beta}'X_i(t)\right]}d\sum_{i=1}^{N}\text{View}_i(t). \quad (11)$$

While the original proportional rates/means model is based on a continuous time horizon, we used the discrete version of the model with a daily time interval. In order to test the predictive validity of the model, we estimated the model using only 98 of the sample videos and kept the other ten as a holdout sample.

Results

In Table 3, we report the estimates for the model parameters. For space-saving purposes, we omit the estimates for the category dummies and only discuss them in the text instead. Recall that the covariates were entered into the model as exponentials. Therefore, the interpretation of their effect is similar to those in proportional hazard models, where $100 \times [\exp(\beta_m) - 1]$ indicates the percent change in diffusion rate due to one unit change in the mth covariate (Helsen and Schmittlein 1993).

Effect of Network Structure

We intended to examine three aspects of the network structure: the size of a UGC author's immediate network of subscribers, the average network size of each of these subscribers, and the connectivity among these subscribers. Fig. 1 shows the distribution of the size of the subscriber base. Consistent with previous research on scale-free networks (Barabási and Albert 1999), the subscriber base displayed a power law-like distribution, with a majority of authors having a small number of subscribers and a few with a large number of subscribers. As expected, the number of subscribers had a significant positive effect on a video's popularity ($\beta = .0002$; $p < .001$), with each 100 additional subscribers contributing to about 2% increase in the diffusion rate.

The average network size of each subscriber also had a significant effect, but the effect was negative ($\beta = -.0001$; $p < .001$). As we discussed earlier, a larger network size indicates a wider reach by a consumer but a potentially smaller influence over his/her friends, which could dampen the diffusion of UGC. Given that we did not directly observe connection strength on YouTube, we used a crude proxy of network strength to explore whether a larger network size was indeed associated with weaker connections. On YouTube, connections occur in two ways: as friends or as subscribers. Because a friend connection is established only with the approval of both parties, it can be considered a stronger type of tie. We calculated the proportion of network connections that are friends and correlated this measure with network size. Results showed a significant negative correlation ($r = -.21$, $p = .03$), providing preliminary support that connection strength is likely to be weaker for subscribers with a large number of connections. Taken together, our results point to the possibility that influence may still be more important than pure reach in the UGC area. This may be due to the large amount of UGC shared among consumers everyday and the low value generally expected from such UGC. As a result, consumers may have been trained to

Table 3
Parameter estimates from the proportionate rates model.

Covariates	β	Standard error	p
Sub_i	.0002	.0001	<.001
$\bar{S}_i$	−.0001	.00002	<.001
$Conn_i$	.56	.12	<.001
$Conn_i^2$	−.64	.12	<.001
QI_{i1} (production quality)	−.001	.009	n.s.
QI_{i2} (entertainment value)	.03	.009	<.001
QI_{i3} (educational value)	.04	.006	<.001
$QM_i(t)$	.12	.008	<.001
Vol_i	.007	.001	<.001
$AvgView_i$	.0001	.00002	<.001
Age_i	−.001	.0007	.05
$Gender_i$	.001	.02	n.s.
$OtherViews_i$	.003	.0005	<.001

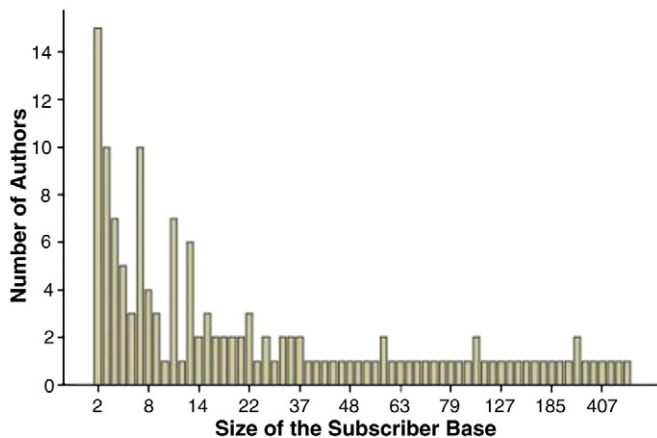Note: See Table 2 for the definitions of these variables.

Fig. 1. Distribution of the size of subscriber base.

ignore most of such sharing unless it comes from someone that they are close to or someone whose opinions they respect.

For network connectivity, the significant positive coefficient for the first-degree term ($\beta=.56$; $p<.001$) combined with the negative coefficient for the quadratic term ($\beta=-.64$; $p<.001$) suggests the presence of a curvilinear effect. As network connectivity rises, the extent of contagion effect first increases and then decreases after it passes a certain threshold. This is consistent with past findings that a highly clustered network (i.e., a social clique) may not allow information to travel far beyond the network, whereas a too disconnected network may not have enough connections for information to spread (Watts 2003). In the YouTube context, although being connected does not necessarily entail direct contact between two individuals, the website does inform users, either on site or through e-mail, of activities from their immediate social circle, such as when a friend uploads or favorites a video. One can also share a video with his or her network on YouTube through its message center. As a result, a similar contagion process can be at work on YouTube as in traditional social networks.

Using our estimates, we can calculate the optimal network connectivity level to be 43.75%. As network connectivity by definition lies between 0 and 1, this means that just under half of all possible ties should be present in a network to achieve maximum contagion for a UGC. We do note, however, that our network connectivity definition is based on an affiliation network structure rather than overtly observed friendship. As a result, the optimal level found here may not apply to directly observed network connections.[2]

---

[2] Due to our unique operationalization of connectivity, we checked the robustness of our findings using an alternative measure of connectivity derived from observed friendship connections among subscribers. This alternative measure was significantly correlated with our subscription-based measure ($r=.27$, $p<.006$). Including this measure in the model yielded the same substantial finding of an inverted-U shaped relationship. The optimal connectivity level was 61.11%. Detailed model results using this measure are available upon request. We note one technical limitation of this friendship-based connectivity measure: the YouTube API limits the number of friend names one can extract for a user to just over 1000. In other words, we could not see the entire picture of friend connections in the data for anyone with more than 1000 connections, which could result in a biased connectivity measure. For this reason, we favored the affiliation network based measure of connectivity and reported our findings based on that measure.

Content Quality and Content Category

As we suspected, the innate quality of the videos varied widely, with production quality ranging from 1 to 7 out of 10, and both entertainment value and educational value ranging from 1 to 8. These innate quality scores were fairly low for most videos too, having a mean of only 3.74 for production quality, 2.94 for entertainment value, and 2.31 for educational value. For manifested quality, many videos (47 videos or 43.5% of the sample) did not have any public rating information on YouTube. For those that were rated by YouTube users, the ratings also varied widely from 1 to 5 stars (5 stars being the best possible rating), with an average rating of 4.57. In comparison, this rating is much higher than the innate quality reported by our sample of consumers. The correlations between the innate quality scores and public ratings at the end of the data collection period were quite low ($r=.18$ for production quality, .30 for entertainment value, and .01 for educational value). These discrepancies between innate and manifest quality support the need to treat the two quality types separately.

As shown in Table 3, out of the three innate quality dimensions, both entertainment value ($\beta=.03$; $p<.001$) and educational value ($\beta=.04$; $p<.001$) had an equally positive impact on video popularity. A one-point increase on our 10-point scale for entertainment value and educational value would increase a video's view growth rate by 3.05% and 4.02% respectively. Production quality did not have a significant impact, possibly due to the fact that consumers realize these are user-contributed videos and therefore do not expect professional production quality. Manifested quality, in the meantime, had a significant positive influence on UGC's rise to popularity ($\beta=.12$; $p<.001$). A one-point increase on the 5-point YouTube rating scale would result in a quite impressive 12.75% gain in view growth rate. A comparison of the standardized coefficients for manifested quality vs. the innate quality dimensions suggests that manifested quality had a significantly larger impact than all innate quality measures ($\chi^2=146.76$, $p<.001$). This is potentially a result of the higher visibility of manifested quality and the fact that it is available prior to consuming a video. It may also reflect a reliance on others' judgments when it comes to consuming UGC.

For content category, recall that we used Music as the benchmark category. Compared with the benchmark category, the categories of Entertainment and People and Blogs showed significantly higher diffusion rate, whereas the more specialized categories of Autos & Vehicles, Howtos & Style, Nonprofits & Activism, and Science & Technology had a significantly lower view growth rate. The rest of the categories (Sports, Film & Animation, Comedy, Pets & Animals, Gaming, and Education) were not significantly different from the Music category.

Author and Other Source Impacts

We expected author experience, as manifested by the number of videos uploaded and the average view of past videos, to signal an author's ability and as a result to affect diffusion of the author's new video. Consistent with our expectation, both factors had a significant impact on the success of the author's current

video. Every ten more videos uploaded by an author in the past increase his/her current video view growth rate by 6.79% ($\beta = .007$; $p < .001$), and every 100 additional average views for past videos can increase the current video growth rate by 1.23% ($\beta = .0001$; $p < .001$). Note that we already included subscriber base size in the model, which may absorb some of the effects of author experience. The fact that these factors still had a separate impact suggests that they reflect an author's credibility above and beyond the popularity of the author alone. For author demographics, an author's gender did not have a significant effect on the success of his/her video. Age, in contrast, had a significant negative influence ($\beta = -.001$; $p = .05$), which means that contributions by younger people are more likely to be popular than those by their older counterparts. This result is consistent with YouTube visitor demographics, where 60% of the visitors are in the 34 and under age categories (http://www.youtube.com/t/advertising_targeting).

To control for the effect of non-network sources and mechanisms such as site search bringing significant traffic to a video, we included the number of views from these other sources as a control variable in our model. On average, these sources accounted for 48.58% of the videos' final number of views. Not surprisingly, the significance of these sources had a positive impact on a video's popularity ($\beta = .003$, $p < .001$). Together, these results suggest that there are multiple ways for a UGC to gain popularity. While the subscriber network we studied here plays an important role in diffusion, additional factors such as relationship with other contents (consequently, chance of being recommended as a related video) and publicizing efforts outside of the YouTube community (e.g., announcing on Twitter or embedding on other websites) can also affect the success of a UGC.

Model Fit and Predictive Validity

The semiparametric proportional rates model does not specify a baseline rate, and therefore it is impossible to explicitly calculate the expected number of recurrences. However, it does allow the estimation of mean recurrence within the observation time frame using the Nelson estimator, given that the covariates are time-constant (Lin et al. 2000). The only time-varying variable in our model was manifested quality. We averaged the ratings across the entire observation period for each video to estimate the predicted cumulative views for the videos during each time period. These predicted cumulative views correlated highly with the observed cumulative views, ranging from .85 to .99 for the sample videos. The mean correlation coefficient was .95. We noticed that the predictive accuracy generally deteriorated for videos that had more views. This is likely caused by the fact that a daily interval may have been too crude for videos that have amassed a large number of views. For these videos, too many recurrences were lumped into a single time interval and therefore reduced the accuracy of the model predictions.

One of our research goals was to provide a mechanism for forecasting the likely success of a UGC in the early stage of the diffusion process. To see if our model can accomplish this goal, we applied the model estimates to the ten videos in our holdout sample. Fig. 2 shows the actual vs. predicted cumulative views

for four of the holdout videos. A visual examination of the figures suggests that our model predictions are fairly close to the actual views for these videos. The overall correlation among the observed and the predicted views were .94 across the holdout sample.

Conclusions

The participatory nature of Web 2.0 has fostered the quick growth of the UGC space. While the low cost and highly viral nature of this new content format presents valuable opportunities to marketers, the sheer volume of UGC available makes it difficult to determine the right ones to invest effort in. Addressing this challenge, we merge social network analysis and the diffusion literature to identify the key factors related to network structure, content quality and topic, and author characteristics that may affect the diffusion of UGC. Within the context of YouTube, a popular online video community, we traced the diffusion of a sample of new user-generated videos over the course of two months. To account for the sporadic repeat views of a video, we built our model on the proportional rates model that was developed in the biometrics field to analyze recurrent events. The model showed a good fit to the data and performed well with a holdout sample of videos.

Our results suggest that a UGC author's subscriber base as well as his/her past experience (in terms of total videos posted and average views of past videos) has a positive impact on the success of the new video. We also found that the connectivity among existing subscribers has an inverted-U shaped effect on the diffusion of a new video. Diffusion rate at first increases with network connectivity, and then decreases with connectivity after it passes a certain threshold (43.75% to be exact, based on our model estimates). This supports the premise in network science that shows both advantages and disadvantages associated with either too poorly or too highly connected networks with respect to information transmission (Watts 2003). Translating these findings into practice, a UGC creator should avoid exclusively targeting its content to individuals in a closely knit community (e.g., only members of a motorcycle enthusiast group). Instead, a better strategy is to initially spread the content to individuals from multiple social networks representing different interests. This will result in reach to more moderately connected individuals and will be most advantageous to content diffusion.

We found a significant negative effect of secondary network size. We surmise that this may be due to weaker average connections as a result of a large network. Although we did not test network strength directly, our preliminary findings suggest that influence may be more important than reach in the diffusion of UGC, suggesting the value in targeting highly influential individuals when dealing with UGC. With regard to content quality, we demonstrate that quality as manifested by other users' public ratings differed from innate content quality as judged by our sample of consumers. Both the entertainment value and educational value dimensions of innate quality had a significant impact on the diffusion of a video, whereas video production quality did not matter. Manifested quality had the biggest impact on diffusion, with one point gain on ratings leading to 12.75% increase in diffusion rate. Content topic mattered as well, and broader
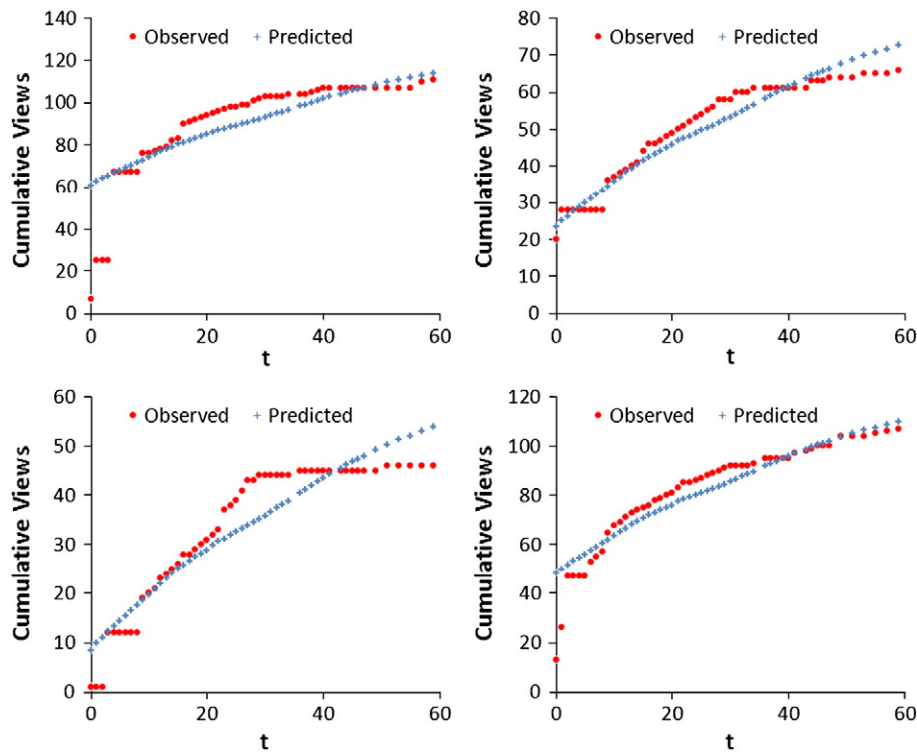
Fig. 2. Observed vs. predicted views for four videos in the holdout sample.

categories such as Entertainment and People & Blogs exhibited higher diffusion rates than more specialized categories such as Autos & Vehicles and Nonprofit & Activism. A UGC author's age also showed a significant impact, with younger authors' content spreading faster.

Limitations and Future Research

In interpreting the results from our study, we note a few limitations of the current research that may be explored in the future. First, we only observed the diffusion of our sample videos for a relatively short period of time. This may affect the generalizability of our findings to a longer time horizon. In reality, it can take a while for some videos to reach a cascading point (i.e., to take off), which may not be captured in our two-month time horizon. Furthermore, in the case of videos that amass views at a very rapid pace, the daily time interval used in our analysis may be too crude, and either the time interval needs to be more refined (e.g., hourly) or the unit of analysis can be increased (e.g., at an increment of 100 views).

Another limitation of this study is that the network characteristics we included in our analysis were only derived from publicly available information on YouTube. While most of our variables were accurately captured through this means, connectivity was inferred from shared subscriptions. This may or may not reflect the true connections among subscribers. Using this operationalization of connectivity also did not allow us to pinpoint the underlying process that may have contributed to the observed relationship. Although the use of an alternative connectivity measure based on overt friendship largely corroborated our findings, future research is needed to verify the inverted-U shaped relationship with a more accurate connectivity measure and to establish the real underlying process through which connectivity works. Relatedly, we operationalized secondary network size as an average across a UGC author's set of subscribers. This mean-based approach may have masked some of the true effects of the variable. For instance, the balance between reach and influence could have resulted in a similarly non-linear relationship between secondary network size and diffusion outcome. This possibility should be explored in future research with an individual-level measure for the variable.

Third, we only captured network structure at the beginning of the diffusion process. Due to the relatively short time span that we covered, it is reasonable to assume that individual networks have not changed dramatically. However, when forecasting over a longer time horizon is necessary, it would be desirable to take network dynamics into consideration. This can be incorporated into the current model by treating network properties as time-varying variables. It would also be interesting to examine the reciprocal effect that UGC diffusion may have on the network structure among users. Given the large universe associated with many UGC communities, identifying network structural properties such as connectivity may require high computational cost, and a user-centered rather than video-centered sampling approach may be more appropriate.

Fourth, we used a convenience sample of consumers to determine innate quality, and only a small number of quality ratings were obtained for each video. As a result, our innate quality measures may not be very accurate. However, the main point we are trying to make here is a need to distinguish between innate and manifested quality of UGC. Their differential impact on diffusion as revealed by our analysis points to the value in making this

distinction. Besides the magnitude of influence studied here, future research is needed to examine *how* these two types of quality may affect the diffusion process differently. For instance, being available both before and after consumption, manifested quality is likely to affect both initial adoption and subsequent contagion. In contrast, innate quality may be mostly responsible for consumers' decision to share a UGC after they have consumed the UGC.

Lastly, while our focus here is on user-generated content, there has been an increasing presence of branded viral messages in the UGC sphere, such as branded channels on YouTube. Many of these branded messages are minimally produced to give the appearance of a UGC. They are often used in conjunction with grassroots campaigns and therefore also rely heavily on the voluntary word-of-mouth sharing among consumers. While the similarity in format of these branded messages justifies the belief that their diffusion may have much in common with that of a UGC, their branded nature can still bring a few key differences, such as consumers' willingness to share without appearing as a commercial agent. It would be interesting to extend the current research to study the diffusion of these branded messages in the UGC space and discern how their diffusion process differs from that of regular UGC.

Implications and Contributions for Research and Practice

It is with academic curiosity that we embarked on this research project to find out why some UGC quickly rises to popularity, while others remain forgotten. Although we do not claim to have exhausted all the factors that can potentially lead UGC to stardom, we believe our research contributes important insight to both marketing research and practice. From a practical standpoint, the explosive growth of the online UGC sphere presents a great opportunity for marketers, especially in this age of cluttered marketplace and declining traditional marketing influence. At the same time, this opportunity has remained elusive, and few companies have been able to capitalize on its potential (Luetjens and Stanforth 2007). Part of the challenge is to sift through the large quantity of UGC available and identify the ones with high potential so that maximum impact can be achieved. In this research, we suggest three types of factors that can affect the diffusion of UGC: network properties, content characteristics, and author characteristics. Our findings suggest that there are indeed systematic variations among UGCs that function as markers of diffusion potential. While the context studied here is online user-generated videos, the basic ideas can be applied to other types of UGC such as consumer blogs and customer reviews. Coupled with more extensive studies in these other contexts, it will be possible to formulate empirical generalizations so that predictions can be made with known values of network properties and other covariates even when limited diffusion information is available (Gatignon, Eliashberg, and Robertson 1989).

Furthermore, as the Internet moves toward and beyond the second-generation Web, the vast social affiliations available through online social networks and the constant sharing among consumers are likely to make network effects even more salient in the marketplace (Van den Bulte and Wuyts 2007). While formal considerations of social network structure are present in the business-to-business literature, similar analysis is far less common in the business-to-consumer and consumer-to-consumer realms. As the current research and a few other recent studies (e.g., Goldenberg et al. 2009; Iyengar, Van den Bulte, and Valente, 2011; Katona, Zubcsek, and Sarvary 2011) demonstrate, valuable insight can be gained from understanding how consumers' connections with each other can affect the way marketing information transmits through the marketplace. For instance, our findings suggest that it is more productive to have a wider network of subscribers that each has a small number of friends than to have a small network of subscribers each with a lot of friends. We also show that a moderate level of connectivity among direct subscribers is optimal for UGC diffusion. Incorporating social network properties such as these can enrich the diffusion literature, which has long implicitly assumed certain network effects but rarely examined the actual processes explicitly (see Valente 1995, 2005). With individual consumers' network connections made more readily available by online social networks, such network analysis can lead to a better understanding of social dynamics in consumption and inform the effective use and optimal design of grassroots marketing campaigns.

References

Asch, Solomon E. (1953), "Effects of Group Pressure upon the Modification and Distortion of Judgments," in *Group Dynamics: Research and Theory,* D. Cartwright and A. Zander, editors. Evanston, IL: Row, Peterson & Co., p. 151–62.

Bandura, Albert (1977), *Social Learning Theory*. Englewood Cliffs, NJ: Prentice Hall.

Barabási, Albert-László and Réka Albert (1999), "Emergence of Scaling in Random Networks," *Science*, 286, 5439, 509–12.

Bass, Frank M. (1969), "A New Product Growth for Model Consumer Durables," *Management Science*, 15, 5, 215–27.

Chiu, Hung-Chang, Yi-Ching Hsieh, Ya-Hui Kao, and Monle Lee (2007), "The Determinants of Email Receivers' Disseminating Behaviors on the Internet," *Journal of Advertising Research*, 47, 4, 524–34.

comScore (2009), "Americans View 34 Percent More Online Videos in November 2008 Compared to a Year Ago," (available at http://www.comscore.com/Press_Events/Press_Releases/2009/1/US_Online_Video_Viewing) (last accessed on March 9, 2011).

Dobele, Angela, David Toleman, and Michael Beverland (2005), "Controlled Infection! Spreading the Brand Message through Viral Marketing," *Business Horizons*, 48, 2, 143–9.

Gatignon, Hubert, Jehoshua Eliashberg, and Thomas S. Robertson (1989), "Modeling Multinational Diffusion Patterns: An Efficient Methodology," *Marketing Science*, 8, 3, 231–47.

Goldenberg, Jacob, Sangman Han, Donald R. Lehmann, and Jae Weon Hong (2009), "The Role of Hubs in the Adoption Process," *Journal of Marketing*, 73, 2, 1–13.

Granovetter, Mark S. (1973), "The Strength of Weak Ties," *The American Journal of Sociology*, 78, 6, 1360–80.

Helsen, Kristiaan and David C. Schmittlein (1993), "Analyzing Duration Times in Marketing: Evidence for the Effectiveness of Hazard Rate Models," *Marketing Science*, 11, 4, 395–414.

Iyengar, Raghuram, Christopher Van den Bulte, and Thomas W. Valente (2011). "Opinion Leadership and Social Contagion in New Product Diffusion," *Marketing Science*, 30, 2, 195–212.

Katona, Zsolt, Peter Pal Zubcsek, and Miklos Sarvary (2011), "Network Effects and Personal Influences: Diffusion of an Online Social Network," *Journal of Marketing Research*, 48, 3, 425–43.

Lawless, Jerald F. and Claude Nadeau (1995), "Some Simple Robust Methods for the Analysis of Recurrent Events," *Technometrics*, 37, 2, 158–68.

Shapiro, Leo J. and Associates (2008), "User Generated Content Three Times More Influential than TV Advertising on Consumer Purchase Decisions," *Marketing Business Weekly*, 34 (December 28, 2008).

Lin, D.Y., L.J. Wei, I. Yang, and Z. Ying (2000), "Semiparametric Regression for the Mean and Rate Functions of Recurrent Events," *Journal of the Royal Statistical Society: Series B Statistical Methodology*, 62, 4, 711–30.

Luetjens, Peter and Simon Stanforth (2007), "UGC's Untapped Potential," *Marketing Week*, 30, 35, 24–5.

Mahajan, Vijay, Eitan Muller, and Frank M. Bass (1990), "New Product Diffusion Models in Marketing: A Review and Directions for Research," *Journal of Marketing*, 54, 1, 1–26.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook (2001), "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology*, 27, 1, 415–44.

Meade, Nigel and Towhidul Islam (2006), "Modelling and Forecasting the Diffusion of Innovation — A 25-year Review," *International Journal of Forecasting*, 22, 3, 519–45.

Muller, Eitan, Renana Peres, and Vijay Mahajan (2009), *Innovation Diffusion and New Product Growth*. Cambridge, MA: Marketing Science Institute.

Nelson, Philip (1970), "Information and Consumer Behavior," *Journal of Political Economics*, 78, 2, 311–29.

Neuendorf, Kimberly A. (2002), *The Content Analysis Guidebook*. Thousand Oaks, CA: Sage Publications.

Pepe, Margaret Sullivan and Tianxi Cai (1993), "Some Graphical Displays and Marginal Regression Analyses for Recurrent Failure Times and Time Dependent Covariates," *Journal of the American Statistical Association*, 88, 423, 811–20.

Provost, Foster, Dalessandro Brian, Hook Rod, Zhang Xiaohan, and Murray Alan (2009), "Audience Selection for On-line Brand Advertising: Privacy-Friendly Social Network Targeting," Proceedings of the Fifteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Paris, France.

Ratchford, Brian T., Siva K. Balasubramanian, and Wagner A. Kamakura (2000), "Diffusion Models with Replacement and Multiple Purchases," in *New Product Diffusion Models,* V. Mahajan, E. Muller, Y. Wind, editors. Boston, MA: Klower Academic Publishers, p. 123–40.

Rogers, Everett M. (1962), *Diffusion of Innovations*. New York: The Free Press.

Scott, John (1991), *Social Network Analysis: A Handbook*. London: Sage Publications.

Tapscott, Don (2007), "Web 2: Some Bubble," *European Business Forum*, 31, 9–11 (December).

Toubia, Olivier, Jacob Goldenberg, and Rosanna Garcia (2009), "Diffusion Forecasts Using Social Interaction Data," MSI Special Report 09-210. Cambridge, MA: Marketing Science Institute.

Trusov, Michael, Anand V. Bodapati, and Randolph E. Bucklin (2010), "Determining Influential Users in Internet Social Networks," *Journal of Marketing Research*, 47, August, 643–58.

Valente, Thomas W. (1995), *Network Models of the Diffusion of Innovations*. Cresskill, NJ: Hampton Press.

——— (2005), "*Models and Methods in Social Network Analysis*," P. J. Carrington, J. Scott, S. Wasserman, editors. Cambridge, UK, Cambridge University Press, p. 98–116.

Van den Bulte, Christophe and Yogesh V. Joshi (2007), "New Product Diffusion with Influentials and Imitators," *Marketing Science*, 26, 3, 400–21.

——— and Stefan Wuyts (2007), *Social Networks and Marketing*. Cambridge, MA: Marketing Science Institute.

Wasserman, Stanley and Katherine Faust (1994), *Social Network Analysis: Methods and Applications*. Cambridge, UK: Cambridge University Press.

Watts, Duncan J. (2003), *Six Degrees: The Science of a Connected Age*. New York: W. W. Norton & Company.

——— and Peter Sheridan Dodds (2007), "Influentials, Networks, and Public Opinion Formation," *Journal of Consumer Research*, 34, December, 441–58.

YouTube (2009), "Youtube Fact Sheet," (available at http://www.youtube.com/t/fact_sheet) (last accessed on March 9, 2011).

Yuping Liu-Thompkins is an Associate Professor of Marketing and E. V. Williams Faculty Fellow in the College of Business and Public Administration at Old Dominion University. She received her MBA and Ph.D. in Management with a concentration in Marketing from Rutgers University in 2002. Dr. Liu-Thompkins' research focuses on the intersection among marketing, technology, and consumer psychology. Her main research areas include Internet marketing, loyalty programs, and customer relationship management. Dr. Liu-Thompkins' publications have appeared in *Journal of Marketing, Journal of Advertising, Journal of Advertising Research*, and *Business Horizons*, among others. More information about Dr. Liu-Thompkins' research can be found on her website at http://www.yupingliu.com.

Michelle Rogerson is an independent marketing and public relations consultant specializing in social media doing business as Reina Communications (www.ReinaCommunications.com). She holds a master's degree in communications management from Syracuse University's S.I. Newhouse School of Public Communications. Rogerson is accredited in public relations (APR) through the Public Relations Society of America.