2012-3

# Social Networks and the Diffusion of User-Generated Content: Evidence from YouTube

Anjana Susarla, *et al.*

*Information Systems Research*
Vol. 23, No. 1, March 2012, pp. 23–41
ISSN 1047-7047 (print) | ISSN 1526-5536 (online)

http://dx.doi.org/10.1287/isre.1100.0339


# Social Networks and the Diffusion of User-Generated Content: Evidence from YouTube

Anjana Susarla
Tepper School of Business, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, anjanas@andrew.cmu.edu

Jeong-Ha Oh
School of Management, University of Texas at Dallas, Richardson, Texas 52242, jhoh@utdallas.edu

Yong Tan
Michael G. Foster School of Business, University of Washington, Seattle, Washington 98195, ytan@uw.edu

This paper is motivated by the success of YouTube, which is attractive to content creators as well as corporations for its potential to rapidly disseminate digital content. The networked structure of interactions on YouTube and the tremendous variation in the success of videos posted online lends itself to an inquiry of the role of social influence. Using a unique data set of video information and user information collected from YouTube, we find that social interactions are influential not only in determining which videos become successful but also on the magnitude of that impact. We also find evidence for a number of mechanisms by which social influence is transmitted, such as (i) a preference for conformity and homophily and (ii) the role of social networks in guiding opinion formation and directing product search and discovery. Econometrically, the problem in identifying social influence is that individuals' choices depend in great part upon the choices of other individuals, referred to as the *reflection problem*. Another problem in identification is to distinguish between social contagion and user heterogeneity in the diffusion process. Our results are in sharp contrast to earlier models of diffusion, such as the Bass model, that do not distinguish between different social processes that are responsible for the process of diffusion. Our results are robust to potential self-selection according to user tastes, temporal heterogeneity and the reflection problem. Implications for researchers and managers are discussed.

*Key words*: diffusion; user-generated content; YouTube; social networks; reflection problem
*History*: Sanjeev Dewan, Senior Editor; Indranil Bardhan, Associate Editor. This paper was received July 21, 2008, and was with the authors for $10\frac{1}{4}$ months for 3 revisions. Published online in *Articles in Advance* April 8, 2011.

## 1. Introduction

The past decade has witnessed a tremendous growth in social computing and user-generated content (Peck et al. 2008), shifting the role of technology from information processing to actionable social intelligence embedded in computing platforms (Wang et al. 2007). This research is motivated by the tremendous growth of a social computing platform, YouTube. The usability and functionality of YouTube makes it easy for users to create their own channel and to post content that can be shared almost instantaneously to a wide audience across the world, making this an attractive platform to content creators and media companies alike. YouTube also enables a variety of social interactions whereby users can choose to friend or subscribe to other channels, comment on or choose favorite videos, and even post response videos to other channels. This dual nature of user participation, in content creation as well as opinion formation, is in contrast to earlier online communities that did not enable such rich features of social interaction (Parameswaran and Whinston 2007). Ultimately, the explosion of creativity and self-expression unleashed by YouTube promises to transform consumer engagement with popular culture, and has the potential to alter the structure of industries that deal with digital products such as media and entertainment.

Compared to other models of user-generated content, the democratic nature of content creation and the lack of formal monitoring and reputation mechanisms foster a self-regulating dynamic of social interaction on YouTube. The nature of product search, content discovery, and consumer preferences on YouTube thus entail different assumptions about behavior and decision making, as distinct from that of atomistic market agents maximizing individual utility. For instance, economic models of aggregate influence, such as network externalities, assume that others' actions impact one's own behavior by directly affecting one's payoff, rather than changing the information available to market agents through the networked structure of social interactions (e.g., Easley and Kleinberg 2010).

One of the hallmarks of YouTube is the tremendous variation in the success of content, where a handful

of videos acquire Internet superstar status while most languish in obscurity (e.g., Crane and Sornette 2008). The inequality and unpredictability of success in cultural markets has been attributed to social contagion (Salganik et al. 2006). The central question examined in this paper is the impact of contagion through the networked structures of interaction on the diffusion of digital products on YouTube. Social contagion broadly describes a class of phenomenon where preferences and actions of individuals are influenced by interpersonal contact, impacting the aggregate diffusion and spread of behaviors, new products, ideas, or epidemics (Dodds and Watts 2004, 2005). We build upon a rich set of explanations for social contagion, such as the desire for social conformity, homophily, and awareness diffusion. Using a data set of video information and user information collected from YouTube, we find that social interactions play an important role not only in the success of user-generated content but also on the magnitude of that impact.

Identifying social influence underlying aggregate popularity growth poses a number of econometric challenges. First, individual user preferences might be subject to popularity surges that create a serial correlation. Second, it is difficult to infer true social influence, whereby individual's behavior could result either from the prevailing norms or tastes of the social group that she belongs to, from similarity in behavior due to common characteristics or similar environments. The problem in distinguishing true social influence from that of spurious correlation is referred to as the *reflection problem* (Manski 1993). We build upon a considerable body of research across several fields such as marketing, economics and sociology to identify true social influence or *peer effects* from alternate factors (e.g., Bandiera and Rasul 2006, Bramoulle et al. 2009, Granovetter 1973, Manski 1993, Van den Bulte and Lilien 2001, Sacerdote 2001). In particular, the impact of a focal channel in directing user search behavior can be confounded by user self-selection. In other words, we need to examine whether the diffusion process is driven users selecting which channels they want to view, or whether it is the social structure of interactions that confer an authoritative role to a central channel. Our empirical approach distinguishes between endogenous or true social influence from contextual or correlated effects arising from user self-selection by conducting exclusion restrictions that identify group interactions. Third, aggregate popularity of a video might be a result of unobserved attention-gathering efforts by channels that impacts a potential user's propensity to view particular types of content or to visit certain channels. We employ Hausman-Taylor estimation methods and multilevel models to consider the potential heterogeneity in video and channel characteristics on YouTube.

This paper can make the following contributions to literature. A considerable amount of IS literature has examined the impact of diffusion on individuals' adoption (Fichman 2000). By contrast, this paper examines the dynamics of digital content diffusion structured through a network. We quantify the impact of the social network structure as a pathway in (i) providing information aiding in product search and discovery and (ii) spreading social influence that impacts potential experience. Prior models of diffusion, such as the Bass model (Bass 1969), do not identify the mechanism by which the transmission of social influence occurs. Indeed, it has been posited that a limitation of prior literature is that the S-shaped diffusion curves could result from user heterogeneity (Van den Bulte and Stremersch 2004). This paper disentangles the impact of social contagion on diffusion from that of (i) endogenous self-selection of users into groups dictated by tastes and (ii) channel heterogeneity that could impact the awareness of potential viewers. We also examine whether the diffusion process in the initial phases may be subject to different influences from that in the later phases by distinguishing between search and experience characteristics of a video.

The structure of the paper is as follows. We discuss the context of YouTube and describe the taxonomy of social network structures in §2. Section 3 presents the theory and hypotheses. In §4 we discuss the data collection process and operationalization of measures. Section 5 presents the empirical approach, §6 discusses the results, and §7 concludes.

## 2. The Context

Given the ease of creating a personalized page or channel, a user on YouTube can engage in self-expression (Raymond 2001) as well as obtain peer recognition (Resnick et al. 2000) from social interactions with other users. A channel allows users to display content that they uploaded; videos from other members; videos *favorited* by the channel, their friends, and subscribers; as well as channels that they subscribe to. The ease of creating a personalized channel on YouTube therefore blurs the boundaries between creators and consumers of content. A friend relationship on YouTube is initiated by an invitation from one channel to another, requiring confirmation from the other. Because the friend network is the result of mutual agreement, we characterize such networks as an undirected network (e.g., Newman 2003). By contrast, the act of subscription indicates a willingness to visit and watch the videos uploaded by another channel. Because a subscriber relationship is a one-way relationship representative of user tastes, we characterize it as a directional network. When a new video is posted, all the friends and subscribers of

a channel are alerted through email or RSS feeds. Subscribers and friends can rate and comment on videos in addition to adding videos to a list of favorites. Such choices can also serve as signals to other users, driving the popularity of content.

We therefore identify three distinct mechanisms of social influence on YouTube. First, there are networks of friends within the community of interest that we characterize as a local network of friends. Second, we observe friend ties between users from outside the community of interest, which we characterize as nonlocal or long ties (Centola and Macy 2007). Third, we observe that there are networks of subscribers within the community of interest, or social networks based on instrumental ties, i.e., a pattern of affiliation based on shared interests. In the data collection section, we explain how we demarcate the boundaries of these social network structures. Given the different motives in adding friends or subscribing to another channel, we expect differences in the mechanism of social ties characterizing different types of networks. We do not observe substantial overlap in membership between friend and subscriber networks, which bolsters our argument about the difference between these networks.

The qualitative evidence from YouTube on the networked structure of interactions summarized in the online appendix[1] highlights several patterns of interest. First, friend ties within the community of interest are characterized by greater frequency in interaction compared to subscriber ties, consistent with prior theory that group membership is a basis for defining identity and for social interactions (Watts et al. 2002). Friend relationships could be characterized by homophily and affinity while subscriber ties could exist for informational purposes to obtain content based on users' tastes and interests. Second, we find greater interaction between friends within the network boundary, local friends, than from friends outside the network. We therefore build on prior literature by differentiating between cohesive ties or "local" network effects from that of nonlocal or nonredundant ties (long ties), which have different strengths from a structural or from an information transmission perspective (e.g., Burt 1987, Centola and Macy 2007, Sundararajan 2007). Putnam (2000) differentiates between bridging capital that refer to loose connections between individuals who act as a source of useful information or new perspectives, and bonding social capital that exists in closer relationships, such as kinship or friendships. The subscriber networks on YouTube seem to be indicative of the former, and could be classified as "weak ties" (Granovetter 1973, p. 1362).

[1] An electronic companion to this paper is available as part of the online version that can be found at http://isr.journal.informs.org/.

# 3. Theory and Hypotheses

## 3.1. Prior Literature

The Bass model (Bass 1969) has been widely used to study the diffusion process in marketing (e.g., Mahajan et al. 1990) and other disciplines. Social network methods have provided an important framework to study the diffusion of innovations in the sociology area (e.g., Wejnert 2002) starting with a landmark study by Coleman et al. (1966) that analyzes the impact of social contagion on medical innovation. Strang and Tuma (1993) consider heterogeneity in the susceptibility of individual agents to contagion. Another stream of work has also examined the impact of structural properties of social networks on the propagation of epidemics (Watts 2002, Watts et al. 2002).

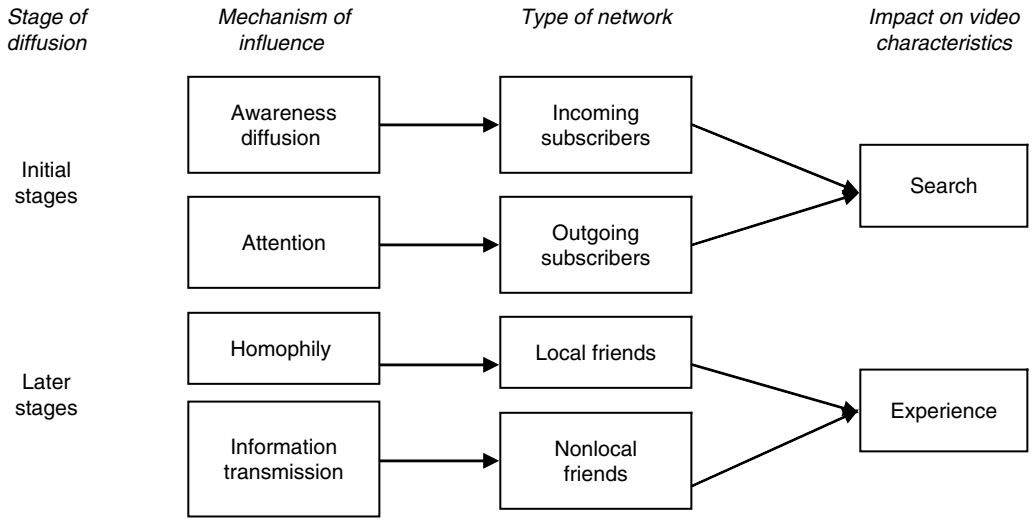
While the question of social influence has been explored in earlier literature in IS (e.g., Armstrong and Sambamurthy 1999, Kraut et al. 1998), there are some crucial distinctions between this study and prior work. First, the literature on product diffusion distinguishes between well-informed early adopters and late adopters, wherein early adopters learn by doing while late adopters learn from observing others.[2] By contrast, we posit that the influence of central channels stems from their structural position in the social network. Second, we consider diffusion in the context of a community where the diffusion process propagates through proximate links in a network. Third, we explicitly address the reflection problem that poses a challenge in inferring social influence. It has been suggested, for instance, that marketing efforts can confound the role of social contagion (Van den Bulte and Lilien 2001). The S-shaped diffusion curves could also result from user heterogeneity in the propensity to adopt (Van den Bulte and Stremersch 2004) rather than the impact of social contagion.

## 3.2. Hypotheses

The two dimensions of user participation on YouTube, in both content creation as well as opinion making, impact two distinct aspects of the diffusion process. Following Kalish (1985), we distinguish between search attributes and experience attributes of a video that impact different stages of diffusion. First, given the bewildering array of content choices and the limitations of keyword search on YouTube (Szabo and Huberman 2009), potential viewers may lack awareness about the range of product choices available. Unlike other types of entertainment products, it is also unlikely that a substantial number of channels engage in promotional efforts or publicity efforts of

[2] Bandiera and Rasul (2006) explicitly distinguish between the two models of learning.

**Figure 1 Conceptual Diagram of Social Influence**



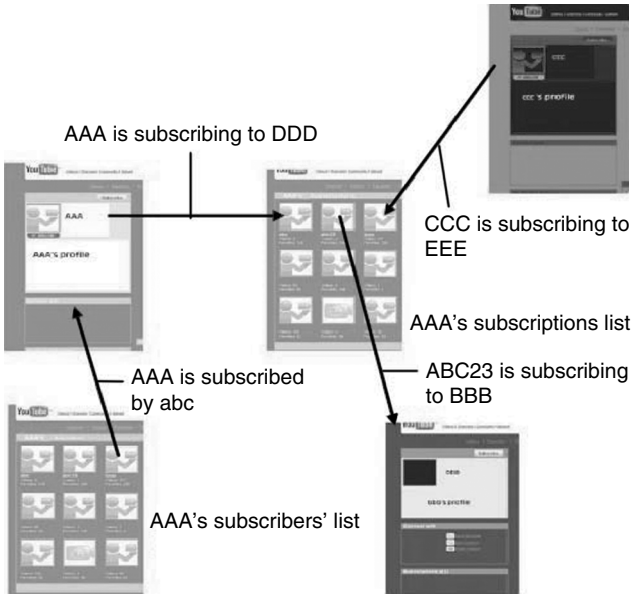

any form, which makes it highly uncertain that a given video can acquire any momentum and reach a wide audience of viewers. The myopic nature of product discovery coupled with the range and depth of offerings and the growth of titles in YouTube substantially increase the uncertainty in searching and locating content. We consider the role of subscriber networks as a conduit to transmit awareness about new videos, which is very important in the early stages of diffusion. Second, a video being an experience good (e.g., Nelson 1970), it is characterized by substantial uncertainty in terms of whether viewers will favorably react to it or not. Because individual preferences are highly influenced by the tastes and opinions of others (e.g., Bikhchandani et al. 1992), social processes such as conformity and peer pressure structured through friend networks could impact a user's perceived experience of a video.

Given the uncertainty in search and experience, this paper examines the impact of the network position of a channel (posting a video) on the process of diffusion of the video over the aggregate YouTube network. The diffusion process begins when a node that is highly influential (measured through degree centrality in our study) generates the infection by posting a video, which then spreads to other nodes through two mechanisms: (i) the ability of the central node to impact awareness of proximate actors by directing search, and (ii) the ability of the central node to influence the potential experience of proximate actors through homophily and cohesion. The type of contagion we consider is that of a central actor increasing the susceptibility of proximate actors by influencing the information available to other actors. In other words, even if proximate actors (alters) do not immediately catch the infection from the focal channel (ego), the latter is still contagious because it is attempting to transmit the infection. The informational role of subscriber networks matters in the early stage in the life of a video due to uncertainty in search characteristics, while the influence of a channel in the friend network matters in the later stage of the life of a video due to the uncertainty in experience faced by a potential viewer. The conceptual diagram of social influence structured through different types of social networks considered in this paper is presented in Figure 1.

**3.2.1. Impact of the Subscriber Network in Directing Search.** We examine the mechanism by which interactions structured through a social network lower the uncertainty in search associated with a new video. Given the limited attention span of viewers and fluctuations in daily viewership of videos (Szabo and Huberman 2009), popularity of user-generated content is relatively ephemeral (e.g., Cha et al. 2007). Potential viewers face an informational bottleneck since they may be aware of only a very small fraction of the available set of videos (e.g., Hendricks and Sorenson 2009). Because that the process of searching and locating content might be local and myopic,[3] a channel with a central position in the subscriber network serves a conduit of information to others, guiding proximate nodes through the *taste neighborhood* (Liu et al. 2006). In the tradition of the epidemiology literature, infection (in our case the number of views of a video) spreads from node to node through interpersonal contact structured through a network (e.g., Newman 2003); Figure 2 depicts a representative subset of nodes

[3] Google has introduced video search optimization tools and enhanced recommendation tools for videos, which can influence the search process, but these features were not present at the time of our data collection.

**Figure 2 Subscriber Relationship (Directed Network)**




in the directed subscriber network, distinguishing between incoming and outgoing connections of a channel.

Channels with several incoming connections might be more popular and might be more likely to disseminate information about a new video among a wider group of actors. When a channel with a greater number of incoming ties from other nodes in a subscriber network posts a video, it lowers the informational bottleneck faced by proximate nodes by increasing exposure and thereby hastening awareness diffusion among the proximate ties (e.g., Kalish 1985). These proximate nodes in turn could be influential in disseminating information and directing search from other nodes, and eventually the video diffuses over the entire network. Thus, the shared interest in sampling videos characterizing social networks of subscriber groups provides an opinion-making role to channels that have a high degree of incoming connections.[4]

Channels with several outgoing connections may be more gregarious, and more likely to be aware about the types of content posted and viewed by nodes that are incident upon the focal channel, i.e., more aware of the preferences of incident ties. The greater connectedness of the channel could provide an informational advantage in directing awareness of incident ties toward itself. The greater the ability of the central channel to seek attention, the more the transmissibility of the contagion, enhancing the likelihood that other viewers will actually watch a video, enhancing the diffusion among the aggregate network. A channel with a large number of outgoing connections is also more susceptible to social contagion, increasing the likelihood of infecting other nodes and thus could be enormously influential in the dynamics of diffusion (Dodds and Watts 2004). Given the importance of both incoming and outgoing ties, we hypothesize as follows.

Hypothesis 1A. *Channels that are central in the subscriber network have a significant impact on the rate of diffusion*.

A channel with a greater number of incoming or outgoing subscriptions has access to a larger the pool of early adopters that are vulnerable to being *infected*. Contagion of new videos can be accelerated when a central node (channel) is connected to a pool of vulnerable nodes that are easily susceptible (Watts 2002). Channels that are more connected have a greater ability to *transmit* information about a video to proximate others. Centrally connected channels occupying a position of greater transmissibility are therefore on the *cutting edge* (Rogers 1995), acting as opinion leaders who influence the spread of information about new videos that have not acquired recognition from the overall YouTube audience (e.g., Bass 1969, Rogers 1995, Mahajan et al. 1990), which is particularly important in the early stages of diffusion. The greater the number of incoming connections of a channel, therefore, we should expect that the greater the influence in the early stages of diffusion, and the more likely that the information about the video is eventually disseminated beyond the proximate actors to the aggregate global network, increasing the likelihood that a new video is viewed.[5] Thus, we hypothesize the following.

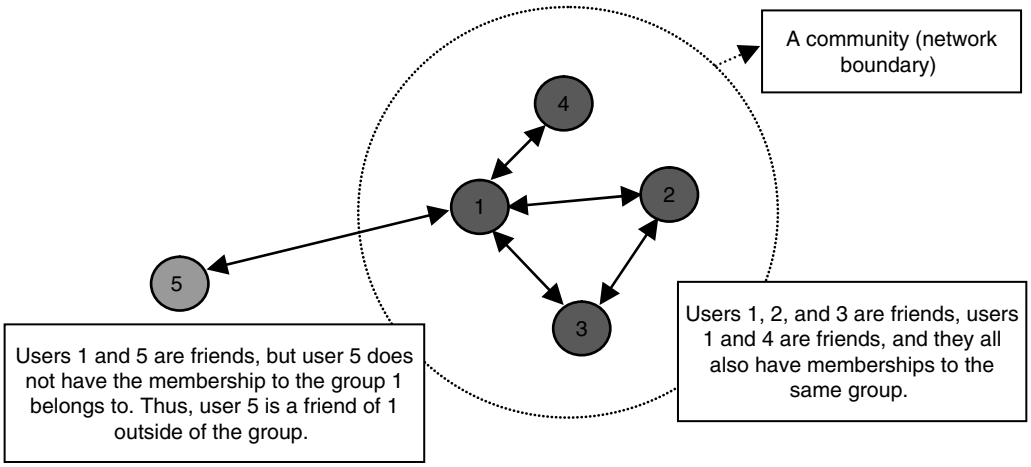
Hypothesis 1B. *A channel's centrality in the incoming subscriber network has a significant positive impact on the rate of diffusion in the initial phase of content diffusion*.

**3.2.2. Impact of the Friend Network Structure in Mitigating Uncertainty in Experience.** We suggest two mechanisms by which social networks impact the perceived value from a video and thus influence diffusion. Friend networks within a community of interest on YouTube might arise from similarity in personal characteristics and consistent interests, which promotes homophily and cohesion (Burt 1987). However such cohesion within the local network, could in turn, also lead to redundancy in the social structure

[4] This explanation only relies on the centrality of the channel and not on the directed path: "(infections) do not have targets and do not prefer to take the shortest paths to any node" (Borgatti 2005, p. 61).

[5] Promotional efforts by channels that are central in a subscriber network can help a recent video attain a "most viewed" or "most discussed" status, which in turn leads to more awareness and drives newer views.

**Figure 3 Friend Relationships Inside and Outside the Group Boundary**




of interaction, creating a *network-redundancy* tradeoff (Reagans and Zuckerman 2008). Another mechanism to activate contagion is that of non-local friendship ties, sometimes characterized as "long ties" (Centola and Macy 2007, p. 704). Figure 3 depicts the friend relationships in the local network and outside the group boundary.

**Conformity Through Local Ties.** We consider the degree centrality of a channel in the (undirected) friend network, which denotes that a channel has a greater number of ties or neighbors. Because an actor could influence proximate actors' opinions or preferences, individuals who occupy a key position in a network marked by affinity have more power of influencing perception and impacting potential experience for two reasons. First, individuals care deeply about the opinions of others they interact with, value conformity with the choices of others (Bernheim 1994) and face dissonance when they do not adopt the choices of individuals whose approval they seek. Central actors can then influence the perceptions of others (Ibarra and Andrews 1993). The stronger sense of identification and stronger patterns of interaction within local friend networks, i.e., networks of friends connected by an interest group, promotes cohesion that enhances localized conformity (e.g., Bikhchandani et al. 1992). An actor or channel that is central in a cohesive local network has greater influence on proximate actors' decisions through localized conformity, increasing the likelihood that other nodes connected to the focal channel view a video that the channel has posted. Second, even if we do not ascribe a higher status or greater influence to a central node, in a cohesive local network, ties are characterized by homophily that fosters trust; thus, we should expect that a channel with a greater number of proximate connections is more influential in the local network.

The influence of a central node matters not only in the local network but also matters to the overall network. Conformity preferences within a group are enhanced through cohesive social network structures that foster a sense of social identity. A channel with a greater degree centrality is in a strong position to influence proximate actors' willingness to experience content posted by the channel. At the aggregate level, the individual influence from a central channel is magnified when proximate others are induced to view a video;[6] the resulting viewing patterns could diffuse to other agents and to the overall global network. The result is a social multiplier effect that strengthens the influence of central actors, enhancing the effect of conformity preferences. Thus, we have the following hypothesis.

Hypothesis 2A. *Channels that are central in the local friend network significantly affect the rate of diffusion.*

We expect that the centrality of the channel and the prestige (Bonacich 1987) of a channel would have different impacts during different stages of the diffusion process. Due to localized conformity, a channel with more prestige in the friend network could have significant ability to persuade proximate actors (Rogers and Kincaid 1981) who value their judgment. Thus, we should expect that the prestige of a channel in the local network is more important early in the life of a video. On the other hand, while centrality of a channel within a local friend network might confer influence in persuading others, the patterns of interactions in a cohesive network structure could constrain contagion due to redundancy in information transmission (Burt 1987). A central channel's influence could be limited early in the life of a video when it needs to compete for attention with videos preferred by

[6] When a proximate node watches or rates a video, other nodes incident to the proximate node get updated about this behavior, which provides a channel through which social influence can be transmitted.

other actors connected to those proximate to the focal channel. However, once a video acquires a critical mass of views, the impact of a central channel in the overall global network could be greater due to the lower uncertainty associated with search in the later stage of diffusion. The larger number of proximate ties linked to the central channel could be influential in persuading others to view a video, hastening diffusion through the overall network. We therefore hypothesize the following.

Hypothesis 2B. *A channel's centrality in the local friend network has a positive impact on the rate of diffusion in the later stages of the process of diffusion.*

Hypothesis 2C (H2C). *A channel's prestige in the local friend network has a positive impact on the rate of diffusion in the early stage of the process of diffusion.*

**Information Transmission Through Long Ties.** We consider the impact of nonlocal friend ties, or *long ties* (Centola and Macy 2007) in transmitting contagion beyond the boundaries of the local network structure.[7] Burt (1992) classifies a network contact is characterized as nonredundant if an individual does not share ties with other contacts in a person's immediate social network. Within a community or a local network, greater identification as a result of homophily and the consequent localized conformity and cohesion might also increase the redundancy of information (e.g., Burt 1992), which limits the spread of contagion. We therefore consider the impact of a node with a greater degree of connections outside the local network, or long ties, in facilitating contagion. While the tie strength might be structurally weaker, such nodes can still influence long ties through affinity. Given the cognitive overload involved in choosing between different videos, the time required to sample a variety of videos and the uncertainty in the experience from a video, a channel with a greater number of non-local ties can promote awareness about content, and activate (infect) a greater number of non-local neighbors (e.g., Watts 2002), enhancing the global spread of contagion. Thus, we hypothesize the following.

Hypothesis 3A. *A channels that has a greater number of friend connections outside the local network of friends has a significant impact on the rate of diffusion.*

A channel can activate contagion beyond the cohesive but redundant pathways in the local network structure (e.g., Dodds and Watts 2004) in the following two ways. First, a channel with greater number of connections outside the local network can shape perceptions under conditions of uncertainty or ambiguity when the nonlocal nodes are unsure about the actual experience offered by a video. Because long ties are weaker from a relational perspective (Centola and Macy 2007), what matters is not the channel's ability to persuade nonlocal friends but to inform them. The connection to nonlocal friends becomes more important once the adoption threshold has already been reached. Second, the greater number of ties to individuals outside the local network increases the ability of a channel to transmit a diverse, novel, and rich set of informational cues (Burt 1992) to the global network. This ability to activate contagion to a nonlocal tie is particularly important in the later phases of content diffusion, when the influence of the central node can be extended beyond the local network. Thus, connections to nonlocal ties matter more in the later stage of diffusion both from a relational as well as an information transmission perspective. Thus, we hypothesize as follows.

Hypothesis 3B. *Channels that have a greater number of friend connections outside the local network have a significant impact on the rate of diffusion in the later phases of content diffusion.*

### 3.3. Peer Influence and Heterogeneity in User Tastes

Prior literature suggests that diffusion curves could also result from differences in user propensities to adopt a new product, rather than social contagion (Bemmaor and Lee 2002), which poses an identification challenge to the empirical estimation. In the terminology of the literature on diffusion, heterogeneity in user tastes reflects the unobserved social preferences (e.g., Van den Bulte and Stremersch 2004). Such heterogeneity in tastes might create a self-selection of users clustered into different channels or interest groups. While we do not hypothesize about the impact of user tastes on group formation, we control for this possibility in the empirical estimation.

## 4. The Data

### 4.1. The Data Collection Approach

We employ a panel of data consisting of video information and user information collected from YouTube.com over a period of two months. Our sample focuses on the videos uploaded within the group of our interest, and the members of that specific group for a total of 4,106 videos posted by 913 users. The data were collected for 11 observation points in time, each 5 days apart. At each observation point, the information on each video and each user within the group was collected by taking screen shots. For each

[7] Both explanations (local or nonlocal ties) rely on the ability of a node to influence others nodes or to transmit an infection to other susceptible nodes.

user, we collected the complete list of friends, subscribers, and subscriptions, which are tracked repeatedly over time as multiple events. Because this data collection was repeated, we can get snapshots of the network structure over time.

We define the cumulative demand, $v_{ijt}$, for video $i$, posted by user $j$, at a certain time $t$ as the cumulative number of clicks[8] of video $i$ at time $t$, including the total number of views from the aggregate YouTube network. The age of video $i$, *VAge*, is the number of days since a video has first been posted online. The average age of a video—the number of days a video has been online since it was posted—is 212 days, and the average number of times a video is watched is 14,180 with a standard deviation 247,455. Because there is a high dispersion of the popularity of video clips, to control the skewness we used a log-transformed number of views. We validated that the log-transformed measure obeys a normal distribution by conducting the Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling tests. While viewers may repeatedly watch a video, we can assume that there is a reasonable limit to the number of times any given individual watches a video. Because we take logs, any bias caused by repeated viewings will be only a slight downward revision of the estimates. We also gather data on the number of most important external links leading to a video clip. Because a number of links to YouTube from prominent sites could lead to greater number of views (clicks), the number of outer links provides partial control over the traffic from outside YouTube. Table 1 summarizes the video characteristics in our sample.

**Table 1 Video Characteristics**

| Variables | Mean | St. dev. | Min | Max |
|---|---|---|---|---|
| Age of video (days) | 212 | 144.26 | 0 | 731 |
| Number of times a video is watched | 14,180.80 | 247,455 | 4 | 13,449,210 |
| Log number of times a video is watched | 6.77 | 1.77 | 1.39 | 16.41 |
| Number of external links to a posted video | 3.66 | 1.93 | 0 | 5 |
| Video rating | 3.56 | 1.59 | 0 | 5 |
| Age of channel (days) | 379.96 | 171.29 | 2 | 846 |

**Table 2 Group and Channel Characteristics**

| | |
|---|---|
| Age of the group | 2 years |
| Number of group members | 1,558 |
| Average growth of the number of videos per day | 15 |
| Average growth of the number of members per day | 5 |
| Number of videos | 4,106 |
| Average length of a video (in seconds) | 223.06 |
| Number of users (channels that posted videos) | 913 |
| Data collection duration | 2 months |
| Number of observation points in time | 11 |

### 4.2. Social Network Structures on YouTube

To identify the network structures on YouTube, we define the network boundary by focusing on a community (interest group) on YouTube. In the appendix, we discuss the reasons for this data collection approach. A *community* in YouTube is defined as a group with specific video categories (there are a total of thirteen categories[9] in YouTube). For example, a community listed under the "music" category is a group of people who share the same interest, and upload videos categorized as "music." Any person interested in "music" and who wants to share their videos with other members in the group can join the community. Therefore the interest group itself forms the network boundary. Our sample community (or group) is drawn from the "music" and "people and blog" categories, which not only are representative of the characteristics of YouTube as a medium of sharing video clips but also are representative of the social interactions within YouTube.[10] Table 2 provides the descriptive statistics of the group and channel characteristics.

As described in Table 2, the group is 2 years old and has 1,558 group members and 4,106 videos. There are 15 new videos, on average, uploaded every day in this group, and on average five new members join the group every day. We calculate social network measures using UCINET 6 (Borgatti et al. 2002). An individual user represents a "node" in a social network. As described earlier, a friendship network is an undirected network, while a subscriber network is a directed network. Because a user's social network reaches can reach beyond the boundary of the interest group, a user's friends or subscribers outside the group also get notified when a new video is posted; therefore, we also consider the number of friends and subscribers outside the group boundary as factors influencing the diffusion of a video. The network boundary allows us to distinguish between within-group or local social influence from outside-group or global influence (e.g., Centola and Macy 2007, Sundararajan 2007). Because individuals could interact more if they share common interests, the nature of interpersonal influence might be different when individuals exhibit affinity.

[8] This is consistent with reports from industry observers and independent firms such as TubeMogul (http://www.tubemogul.com).

[9] The 13 categories are Autos and Vehicles, Comedy, Education, Entertainment, Film and Animation, Howto and Style, Music, News and Politics, People and Blogs, Pets and Animals, Science and Technology, Sports, and Travel and Events.

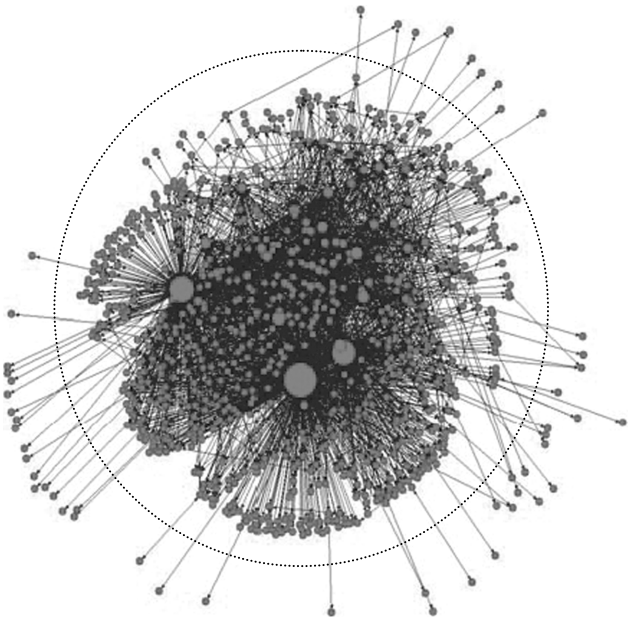
[10] On YouTube, the terms group and community are synonyms.

Because the prestige and status of an actor can be increased as well as reduced, depending on the structure of the network by connections to powerful others, we calculate two measures of social capital of the friend network. Because we are interested in the impact of localized conformity and homophily, we calculate two measures to denote the social influence of an actor. The degree centrality $Frn\_NrmDeg_{jt}$ of user $j$ in the friend network at time $t$ indicates the importance of a node (channel) in the social network (Wasserman and Faust 1994), which measures the size of the proximate network of the node. The number of times a node is visited increases with the degree centrality, and thus indicates the aggregate level of connectedness of a node and the ability and the opportunity of a node[11] to diffuse information about videos. The centrality measures a node's involvement in the cohesiveness of the network, promoting localized conformity (Borgatti and Everett 2006). We present the graph theoretic interpretation of this measure in the appendix. Figure 4 shows a subset of the friend network.

An actor with a high degree centrality denotes where "the action is" in the network (Wasserman and Faust 1994, p. 179). The more central the network position of the actor, the more the actor is a channel of relational information to others (Wasserman and Faust 1994) and occupies a position of social influence (Burt 1987). We also calculate the Bonacich power (Bonacich 1987), which is based on the insight that, while an individual's status is a function of the status of other actors he is connected to, i.e., connections to many prominent others reduce the power of a node. If channel A and channel B have the same number of connections, but B's connections are well connected with others in the network while A' connections are more isolated, B might have less influence on how proximate connections access information and thereby lower aggregate influence. The descriptive statistics of the friend network for the final period of data gathering are presented in Table 3. The minimum degree centrality is 0 while the maximum degree centrality is 521, indicating that some users are highly connected while others are not. $\log NumFrn\_outside_{jt}$, the log-transformed number of friends of user $j$ at time $t$ outside the local network is the number of proximate ties of user $j$ outside the group boundary. Because our focus is on the spread of information in the network, we examine the structural properties of the various networks that indicate the connectivity and consequently the dynamics of diffusion. The network or global-level *density* is the proportion of ties in a network relative to the total number possible. The

[11] A user is an "actor" or "node," while the relationship between users is a "link" or a "tie."

**Figure 4** **Friend Network, Group Boundary, and Nodes Outside the Network**



*diameter of a network* is the largest geodesic distance in the (connected) network.

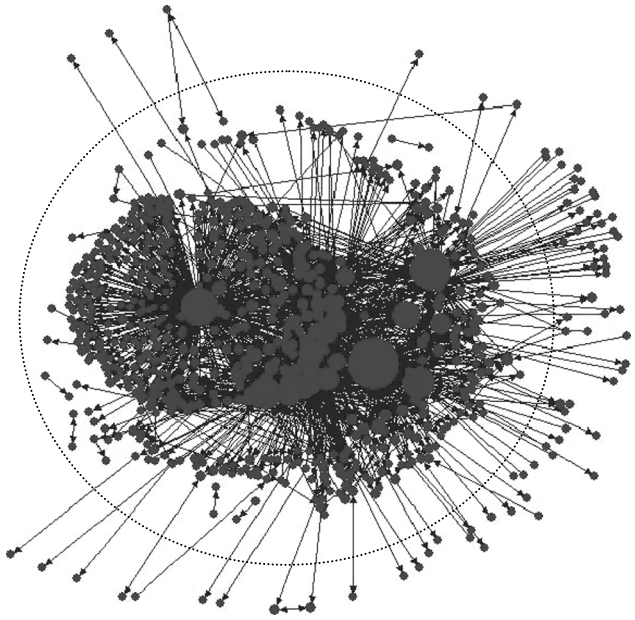
Figures 5 and 6 show a subset of the subscriber network and the group boundary for the last phase of data collection. Because the subscriber relationship is directional, the connectedness of the focal channel in a subscriber network is measured in terms of the in-degree and out-degree degree centrality, $Subs\_NrmInDeg_{jt}$ and $Subs\_NrmOutDeg_{jt}$ of user $j$ in the subscriber network at time $t$, which denotes the extent to which individuals are extensively involved with or adjacent to many other actors in the social network (Freeman 1979, Wasserman and Faust 1994). The interpretation is similar to the centrality measures for undirected graphs except that we distinguish between the directionality of the tie. In-degree centrality indicates the number of ties directed toward a node, while out-degree is the number of ties that a node directs towards others. The network centralization for the out-degree network is greater than that of the in-degree network. While both in-degree and

**Table 3** **Friend Network Description**

| | Degree centrality | Norm. degree centrality | Bonacich power | Norm. BP |
|---|---|---|---|---|
| Mean | 3.852 | 0.261 | −4.83 | −17.15 |
| St. dev. | 20.557 | 1.395 | 9.678 | 34.377 |
| Sum | 5,682.000 | 385.482 | −7,119.7 | −25,291.1 |
| Min | 0.000 | 0.000 | −143.321 | −509.113 |
| Max | 521.000 | 35.346 | 53.135 | 188.749 |

*Notes.* Network centralization = 35.13%; network diameter = 5; network centralization = 0.26%.

**Figure 5** **Subscriber Network (Node Size Based on In-Degree Centrality)**



out-degree indicate that a node is more connected, a channel with a high number of incoming subscriptions could act as a gatekeeper or arbiter, while nodes with a high degree of outgoing subscriptions have greater awareness of what is going on in the adjacent nodes. $\log NumOutSubs\_outside_{jt}$ is the log transformed number of subscriptions (outward directional ties) of user $j$ from outside of the group at time $t$. $\log NumInSubs\_outside_{jt}$ is the log transformed number of incoming subscribers (incoming ties) of user $j$ at time $t$ from outside the group. The descriptive statistics for the subscriber networks for the last phase of data collection are shown in Table 4.

**Figure 6** **Subscriber Network (Node Size Based on Out-Degree Centrality)**

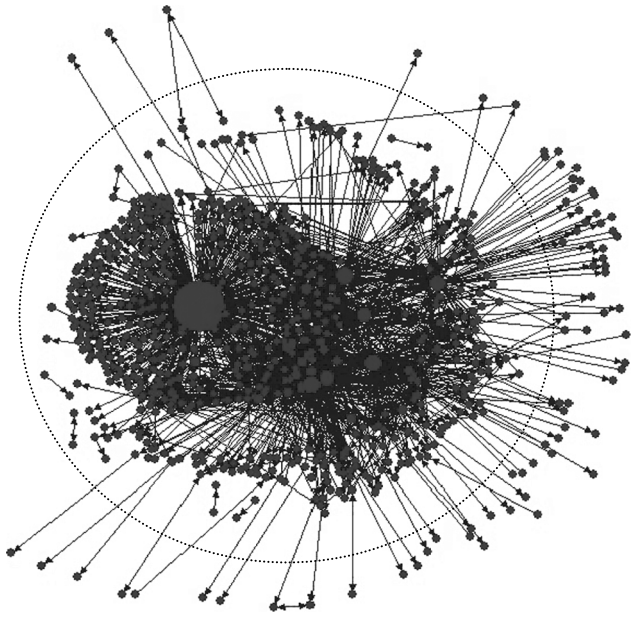


**Table 4** **Subscriber Network Description**

| | Out-degree centrality | In-degree centrality | Nrm. out-deg. cent. | Nrm. in-deg. cent. |
|---|---|---|---|---|
| Mean | 1.352 | 1.352 | 0.092 | 0.092 |
| St. dev. | 12.415 | 3.960 | 0.842 | 0.269 |
| Sum | 1,994.000 | 1,994.000 | 135.278 | 135.278 |
| Min | 0.000 | 0.000 | 0.000 | 0.000 |
| Max | 429.000 | 71.000 | 29.104 | 4.817 |

*Notes.* Network centralization (out-degree) = 29.032%; Network diameter (out-degree) = 4; Network density (out-degree) = 0.09%; Network centralization (in-degree) = 4.728%; Network diameter (in-degree) = 5; Network density (in-degree) = 0.09%.

### 4.3. Control Variables

Because the characteristics of a video affect its popularity, we control for the number of external links, *NumOfLinks*, which enable accesses to the videos from blogs, MySpace, Facebook, or other online communities and forums outside YouTube. This variable provides information of the traffic coming from external outlets other than accesses from within YouTube. We also control for another factor that may influence a video's popularity, the average rating (e.g., Chevalier and Mayzlin 2006) of each video, *Rating*, which are posted by registered YouTube users. Table 5 summarizes the variables. The correlations are presented in the online appendix.

## 5. Empirical Approach

### 5.1. Baseline Model and Variable Descriptions

The baseline model is the standard Bass model that takes into account the social network structure. The Bass model estimates the growth of aggregate demand or diffusion rate as a function of the aggregate demand of the prior time period as well as the time elapsed form the initial launch of a new product. A viewer watching a video on YouTube is equivalent to a consumer adopting a new product and the choice facing a viewer is whether to view a video or not. The dependent variable is the diffusion rate of each video, measured by the growth in views (e.g., Bass 1969), which is the difference of the number of clicks to the video between time $t$ and time $(t-1)$, $\Delta v_{ijt} \equiv v_{ijt} - v_{ijt-1}$, or the popularity growth of video $i$ posted by user $j$ from time $(t-1)$ to time $t$. The measure of aggregate demand is the popularity of each video, $v_{ijt}$, the total number of times a video has been watched (the total number of times a video has been clicked), consistent with the Bass model (Bass 1969) and the prior literature on new product diffusion (e.g., Talukdar et al. 2002). The subscripts $i, j, t$

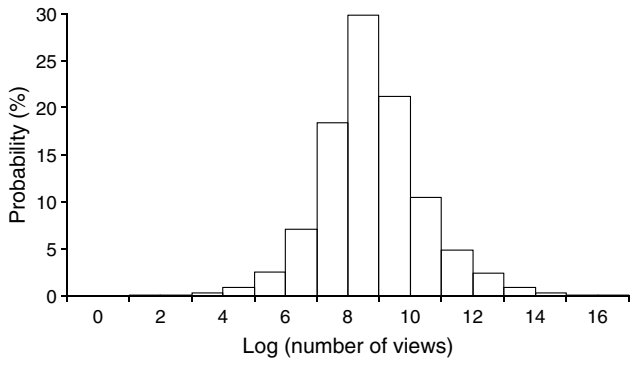
**Table 5 Description of Variables**

| Variable name | Description of variables |
|---|---|
| Video characteristics | |
| log *NumOfViews* | Log number of times a video is watched ($v_{ijt}$) |
| (log *NumOfViews*)$^2$ | The square term of the log number of views ($v_{ijt}^2$). |
| log *VAge* | (log-transformed) Time elapsed since a video has been posted (log($VAge_{ijt}$)) |
| Rating | Video rating (0 to 5) ($Rating_{ijt}$) |
| log *NumOfLinks* | Number of links, which lead to the video, placed outside YouTube |
| Within group social network measures | |
| *Frn_NrmDegree* | Degree centrality of a user in the friend network within the group |
| *Frn_NrmBP* | Normalized Bonacich' power in the friend network within the group |
| *Subs_NrmOut-Degree* | Out-degree centrality of a user in the subscriber network within the group. Connection initiated by the user. |
| *Subs_NrmIn-Degree* | In-degree centrality of a user in the subscriber network within the group. Connection initiated by others. |
| Outside group social network measures | |
| log *NumFrn_Outside* | Number of friends outside the group |
| log *NumOutSubs_Outside* | Number of outgoing subscriptions outside the group |
| log *NumInSubs_Outside* | Number of incoming subscribers from the outside of the group |

denote video $i$, user $j$, and observation time $t$. Figure 7 presents the distribution of the log-transformed views by the percentage of viewers in the group.

The baseline model is presented below. We assume that the average number of views by a single viewer is independent of the exogenous factors in the estimation model. $Y$ is a set of covariates that represent the video characteristics; such as the age of video $i$ at time $t$ uploaded by channel $j$ *VAge*, the number of outer links *NumOfLinks*, and the rating of the video

**Figure 7 Distribution of Log-Number of Views**




$Rating_{ijt}$. The ratings are posted by registered YouTube users and could change over time. The number of external links is the log-transformed number of links outside YouTube, which enable accesses to the videos from blogs, MySpace, Facebook, or other kinds of online forums and communities. Social ties and user preferences outside of the network are also user specific at time $t$. To avoid identification challenges we consider the impact of social network measures up to the period $(t-1)$. The normalized degree centrality of channel $j$ *within the* friend network, and the In- and Out-degree centrality of channel $j$ within the subscriber network are all user specific. Our independent variables are the network measures of friend networks ($Frn\_NrmDeg_{jt-1}$, $Frn\_NrmBP_{jt-1}$, $\log NumFrn\_outside_{jt-1}$) and subscriber network ($Subs\_NrmOutDeg_{jt-1}$, $Subs\_NrmInDeg_{jt-1}$). For both friend networks and subscriber networks, we distinguish between the local network effects ($Frn\_NrmDeg_{jt-1}$, $Subs\_NrmOutDeg_{jt-1}$, $Subs\_NrmInDeg_{jt-1}$) from that of nonredundant ties outside of the group ($Subs\_NrmOutDeg_{jt-1}$, $Subs\_NrmInDeg_{jt-1} \log NumFrn\_outside_{jt-1}$). The diffusion equation incorporating temporal heterogeneity (e.g., Strang and Tuma 1993) is

$$\begin{aligned}\Delta v_{ijt} &= \beta_0+\beta_1 v_{ijt-1}+\beta_2 v_{ijt-1}^2+\beta_3 \log(VAge_{ijt}) \\ &\quad + Y_{ijt-1}\cdot\alpha + X_{jt-1}\cdot\gamma+\varepsilon_{ijt}, \qquad (1)\\ X_{jt-1} &= [Frn\_NrmDeg_{jt-1} Frn\_NrmBP_{jt-1} \\ &\quad \cdot \log NumFrn\_outside_{jt-1} Subs\_NrmOutDeg_{jt-1} \\ &\quad \cdot Subs\_NrmInDeg_{jt-1}], \\ Y_{ijt-1} &= [Rating_{ijt-1} \log NumOfLinks_{ijt-1}].\end{aligned}$$

There are several assumptions underlying the standard Bass model. The population is assumed to be homogeneous, the size of the total population is fixed and known, and the parameters of external and internal influence are assumed to be unchanged over time. Our approach is similar to Strang and Tuma (1993) in that we decompose the diffusion process into two components: (i) the number of individuals who are already infected (and can infect others) and (ii) the hazard rate of contagiousness (as a function of social network position), thereby incorporating heterogeneity in the infectiousness of a channel, depending on its network position, and temporal heterogeneity, which refers to the time-varying influence of factors that drive diffusion. That is, rather than assuming that each node is equally effective in activating contagion, we approximate the individual level of hazard rate by considering the diffusion probability to depend on the degree centrality of the channel posting each individual video. Being a friend or a subscriber to a central node increases the susceptibility of other nodes due to proximity to the central node,

either because the viewer is aware about the video or because the viewer would like to experience the video due to social influence. Because the age variable is skewed, we conducted an alternate specification using a dummy variable for Age instead of the log-transformed measure of age, and found that the coefficients in the baseline model were consistent.

### 5.2. Identification Challenges

Identifying social contagion effects poses a number of statistical and econometric challenges that are not taken into account in the baseline Bass model. First, the baseline model assumes that the difference in the number of views between time periods is a function of the total popularity of a video. When exogenous and unobservable events could trigger a surge of popularity that drives diffusion, $\Delta v_{ijt}$ is influenced by $\Delta v_{ijt-1}$. In other words, there could be a potential bias due to contemporaneous shocks that affect tastes, which creates a serial correlation in estimating the popularity of videos due to unobservable social factors. Such serial correlation needs to be explicitly taken into account in the econometric model.

Second, identifying social influence is complicated by the fact that an individual's choices reflect the choices of the group to which an individual belongs. The difficulty in estimating social contagion effects is that individual behavior is not fixed but varies with the prevailing norms or tastes of the social group. Econometrically, this is referred to as the reflection problem in the literature on peer effects in economics. We address the reflection problem by assuming that the effect of social contagion is based on network composition up to the previous period. This approach can address potential simultaneity between the mean characteristics of a group and the mean outcome. We also observe that users (channels) communicate not only within the group but also with individuals outside the group. When a user (channel) interacts with others outside her subscriber network as well as individuals within the subscriber network, individual decisions are indicative not only of the social influence of the group that the user belongs to (i.e., the subscriber network) but also of the social influence of others from outside the subscriber network.

However, an additional challenge is that we still need to consider whether the measured impact of a user's connectedness in the subscriber networks reflects a systematic pattern of self-selection driven by user preferences rather than the result of informational role of channels that occupy a central role in subscriber networks in directing search and product discovery. Such self-selection could likely affect viewing volume. In that case, users with a high degree of incoming subscriptions might be the ones with *a finger on the pulse*, i.e., arbiters of what is likely to become popular, while those users with a greater degree of outgoing subscriptions have better awareness of others' choices. Given the multiplicity of factors that could confound social influence, we distinguish membership in a social network from other types of social influence by conducting exclusion restrictions on the group composition. Thus, our method of identification is similar to recent approaches that measure social influence by identifying the nature of interactions structured through a network (e.g., Bramoulle et al. 2009).

A third problem is that of unobserved heterogeneity in user tastes, which might result from an unobserved demographic characteristic such as age that reflects the unobserved preferences of viewers. For instance, some users (channels) are likely to be experimenters while others wait to sample content only after it becomes popular. The patterns of diffusion on diffusion and the popularity of content could then result from heterogeneity rather than contagion due to social influence (e.g., Bemmaor and Lee 2002). The structure of network formation in YouTube allows us to address this problem because we can distinguish between a user's membership in a group of social influence (the friend network) from membership in groups dictated by user tastes.

### 5.3. Structural Model Specification

We now consider the interaction between variables that may impact the diffusion process (e.g., Wejnert 2002) and the nature of group formation due to user preferences. The structural estimation proceeds in three stages. First, following the techniques used by Boulding and Christen (2003), we use $\rho$-differencing to remove serial correlation. Second, we conduct exclusion restrictions on the group composition characterizing subscriber networks to address potential self-selection. We conduct two different estimations: (i) Hausman-Taylor estimation (Wooldridge 2002, pp. 225–228) with instrumental variables to take into account unobserved self-selection relationships between users and videos and (ii) a multilevel (hierarchical) estimation to address unobserved channel heterogeneity.

**5.3.1. Rho-Differencing to Remove Serial Correlation.** The Bass model with a serial correlation $\rho$ in the error term, where, $\varepsilon_{ijt} = \eta_{ijt} + \rho\varepsilon_{ijt-1}$ is

$$\Delta v_{ijt} = \beta_0 + \beta_1 v_{ijt-1} + \beta_2 v_{ijt-1}^2 + \beta_3 \log(VAge_{ijt}) + \varepsilon_{ijt}. \quad (2)$$

The autocorrelation effect $\varepsilon_{ijt-1}$ is removed through a serial correlation adjustment, leaving us with a contemporaneous shock term.

$$\begin{aligned}\Delta v_{ijt} = {} & \rho\Delta v_{ijt-1} + \beta_0(1-\rho) + \beta_1 v_{ijt-1} - \beta_1\rho v_{ijt-2} \\ & + \beta_2 v_{ijt-1}^2 - \beta_2\rho v_{ijt-2}^2 + \beta_3 \log(VAge_{ijt}) \\ & - \beta_3\rho \log(VAge_{ijt-1}) + \eta_{ijt}.\end{aligned} \quad (3)$$

After estimating $\rho$ from Equation (3) and taking the first-order difference, we have the variables corrected for serial correlation $\Delta\nu'_{ijt} = \Delta\nu_{ijt} - \hat{\rho}\Delta\nu_{ijt-1}$, $\nu'_{ijt-1} = \nu_{ijt-1} - \hat{\rho}\nu_{ijt-2}$, $\nu'^2_{ijt-1} = \nu^2_{ijt-1} - \hat{\rho}\nu^2_{ijt-2}$, and $\log(VAge'_{ijt}) = \log(VAge_{ijt}) - \hat{\rho}\log(VAge_{ijt-1})$ and the random shock $\eta_{ijt}$ as the error term. Through $p$ values, we establish that the parameter estimate of the auto-regression coefficient is significant.

**5.3.2. Hausman-Taylor Estimation Correcting for Self-Selection.** To identify social effects, Manski (1993) recommends that we understand the factors that dictate the composition of social groups. To consider whether users self-select into different subscriber networks based on their taste preferences, we conduct exclusion restrictions using $NumOutSubs\_outside_{jt}$ and $NumInSubs\_outside_{jt}$ as instruments for out-degree and in-degree centralities of the subscriber network, which can remove the endogeneity in systematic matching across users and social networks depending on unobserved taste preferences. As a robustness check we performed a Wald $F$-test (Angrist and Krueger 1991) for the joint significance of the parameters by including instrumental variables along with the other independent variables in the diffusion estimation, and verified that the test rejected the joint significance of the variables. We estimate the following second-stage models:

$$Subs\_NrmOutDeg_{jt} = f(\log NumOutSubs\_outside_{jt}), \quad (4)$$

$$Subs\_NrmInDeg_{jt} = f(\log NumInSubs\_outside_{jt}). \quad (5)$$

The Hausman-Taylor estimation (Hausman and Taylor 1981) addresses unobserved relationships between popularity growth and video characteristics. In contrast with either fixed effects or random effects estimation, the Hausman-Taylor (hereafter, HT) approach assumes that some and not all of the regressors are correlated with individual effects. Given that a video in our data set can be uploaded by only one channel, we can address a user's time invariant propensity to enjoy a particular video by conducting an instrumental variable (IV) HT estimation (procedure detailed in Wooldridge 2002, pp. 325–328). From the exclusion restrictions described in (4) and (5), the exogenous time-varying variables, the log-transformed number of subscribers from outside the network, are employed as instruments for the endogenous time-varying variables, the out-degree and in-degree centrality of the channel in the subscriber network. That is, because there could be self-selection of users depending on tastes, a channel's position in the subscription network as endogenous and time varying. We then estimate a random effects model using the means of two exogenous time varying variables, video rank, and number of external links as instruments for the endogenous time-fixed variable, the length of the video.[12] This approach allows us to capture the time-invariant characteristics of a video that could impact the diffusion process. Fixed effects estimation, by contrast, removes all sources of time-invariant variation in the explanatory variables. The endogeneity assumptions were validated through a Hausman-Taylor specification test. We also conducted sensitivity tests to establish the appropriateness of the instruments used, consistent with Boulding and Christen (2003) and Wooldridge (2002). We use $X$, $Y$, and $Z$ to represent

$$X_{jt-1} = [frn\_NrmDeg_{jt-1}\ frn\_NrmBP_{jt-1} \cdot \log NumFrn\_outside_{jt-1}],$$

$$Y_{ijt-1} = [Rating_{ijt-1}\ \log NumOfLinks_{ijt-1}],$$

$$\hat{Z} = [Subs\_\hat{N}rmOutDeg_{jt},\ Subs\_\hat{N}rmInDeg_{jt}],$$

$$\Delta\nu'_{ijt} = \beta_0 + \beta_1\nu'_{ijt-1} + \beta_2\nu'^2_{ijt-1} + \beta_3\log(VAge'_{ijt}) + \alpha Y_{ijt-1} + \gamma X_{jt-1} + \delta\hat{Z}_{jt-1} + \eta_{ijt}, \quad (6)$$

$$\Delta\nu'_{ijt} = \beta_0 + \beta_1\nu'_{ijt-1} + \beta_2\nu'^2_{ijt-1} + \beta_3\log(VAge'_{ijt}) + \alpha Y_{ijt-1} + \gamma_1 X_{jt-1} + \gamma_2(\log(VAge_{ijt}) \cdot X_{jt-1}) + \delta_1\hat{Z}_{jt-1} + \delta_2(\log(VAge_{ijt}) \cdot \hat{Z}_{jt-1}) + \eta_{ijt}. \quad (7)$$

**5.3.3. Hierarchical Estimation to Address Channel Heterogeneity.** An alternate causal explanation we need to consider is that the result of growth in views could be the impact of external linkages and publicity efforts by each channel that results in a video receiving more attention, leading to greater awareness diffusion. That is, it could be the effort made by channels to attract subscribers or greater attention-seeking behavior that is responsible for growth in views rather than the social influence. In the econometric estimation, this leads to unobserved heterogeneity resulting from the efforts of channels at targeting potential subscribers.

We categorize channels into three groups, depending on their connectedness in the overall network structure. The first group consists of channels isolated from the overall network structure within the group. The second group of channels are connected to subscribers and friends outside the group (network boundary) but isolated from interactions within the network boundary. The third group comprises channels that are well connected both within and outside the network boundary. We follow Talukdar et al. (2002) in estimating a hierarchical model with an

[12] The Hausman-Taylor estimation was conducted using Stata. The specification examines the rank of the variance-covariance matrix and detects whether the constraints have been met.

**Table 6 Baseline Model Estimates**

| | Parameter estimates |
|---|---|
| Variable | Baseline |
| Intercept | 0.287600 (0.003544)*** |
| Log *NumOfViews* | 0.001317 (0.000378)*** |
| (Log *NumOfViews*)$^2$ | 0.000625 (0.000094)*** |
| Log (*vAge*) | −0.05003 (0.000596)*** |
| *Rating* | 0.001981 (0.000295)*** |
| log *NumOfLinks* | 0.002463 (0.000861)*** |
| *Frn_NrmDegree* | −0.00080 (0.000203)*** |
| *Frn_NrmBP* | −0.00003 (9.959$E-6$)** |
| *Subs_NrmOut − Degree* | −0.00308 (0.000788)** |
| *Subs_NrmIn − Degree* | 0.000182 (0.001749) |
| log *NumFrn_Outside* | −0.00063 (0.000261)** |
| Adjusted *R* Squared | 0.17 |

*Note.* Standard errors in parentheses; $^{**}p = 0.05$; $^{***}p = 0.01$.

idiosyncratic channel specific component of the error term to capture the heterogeneity.

The estimation proceeds as follows. We re-estimate the models in (6) and (7) but add a channel-specific idiosyncratic term $\zeta_k \sim N(0, \omega_k)$ where we define three levels of hierarchy $k \in \{1, 2, 3\}$. The intercept now becomes $\beta_0 + \zeta_k$. There are two variance components—one represents the variance among videos within the group $\sigma$, and the other represents the variation between group means $\omega_k$ (depending on the connectivity of the different groups of channels).[13] Ignoring this systematic variation in global connectedness of a channel, which impacts the ability of a channel to attract attention, could lead to misleading causal inference (e.g., Gelman and Hill 2007). We assume that observations within each level of hierarchy share certain common characteristics resulting from the connectedness of the channel. The error structure then becomes

$$\begin{pmatrix} \sigma^2 + \omega_1^2 & \cdots & \\ \vdots & \ddots & \vdots \\ & \cdots & \sigma^2 + \omega_k^2 \end{pmatrix}.$$

## 6. Results

### 6.1. Discussion of Estimates

The estimates for the baseline model are presented in Table 6. The video rating and the external links are highly significant, indicating that the perceived video quality and strategic linkages with other online channels of influence positively influence diffusion.

The results from the structural model are shown in Table 7. The *R*-squared values are greater in the baseline model due to serial correlation that has been corrected in the structural models.

The in-degree centrality of users in subscriber networks is highly significant, indicating support for Hypothesis 1A. A channel that is central in a network of subscribers could be more influential in directing search and discovery toward itself, with the result that content posted by the channel is more likely to be disseminated into the overall network. Friend networks have a significant impact on diffusion, validating Hypothesis 2A. Homophily and localized conformity could provide an influential role to channels that are central in a local network of friends, and such social influence eventually diffuses to the aggregate network. Channels that are more connected to friends from outside the group play a significant role in the diffusion of content, validating Hypothesis 3A. Central actors that can enrich the set of experiences and ideas available to the aggregate network could provide the pathways to transmit access to new and nonredundant content in the overall network.

The interaction term between in-degree centrality and video age is negative, indicating support for Hypothesis 1B. However, the coefficient of the interaction term between out-degree centrality and video age is positive. The contrasting impacts of outgoing and incoming subscriptions in the earlier stages of diffusion suggest that a channel's ability to direct search and discovery through its incoming ties might be more important than a channel broadcasting its choices and seeking attention through outgoing ties. The interaction term between degree centrality of the friend network with time is positive, indicating support for Hypothesis 2B, i.e., friend networks play a very important role in the later stages of diffusion when the experience attributes of a product are more important. The Bonacich power, which measures the status of the channel, is more important in the early stages than the later stages, supporting H2C. The relative prestige of a channel within the local network could be very important in the earlier stage when the video is fairly unknown. However, the centrality of the channel could be more important in the later stages due to the ability to quickly transmit contagion to a host of other actors. The interaction term of the number of nonlocal friends with video age has a significant impact on diffusion, validating Hypothesis 3B. Nonlocal friends provide an important pathway for a channel to activate contagion beyond the local network neighborhood. The magnitude of the impact of nonlocal ties is stronger than that of the degree centrality of the actor in a friend network, suggesting that the ability of a node to facilitate contagion to the aggregate network might be more important than the relationships marked by homophily and conformity within a local network.

[13] For robustness we estimated a model with channel-specific heterogeneity specified at the level of each individual channel and verified that the grouped approach fits the data better.

**Table 7 Structural Model Estimates**

| | Parameter estimates | | | |
|---|---|---|---|---|
| Variables | Hausman-Taylor | Hausman-Taylor with interaction terms | Hierarchical model | Hierarchical model with interaction terms |
| Intercept | 0.2317 (0.00354)*** | 0.2525 (0.00628)*** | 0.22800 (0.004360)*** | 0.24650 (0.00684)*** |
| Log *NumOfViews* | 0.00196 (0.00034)*** | 0.00207 (0.00035)*** | 0.00232 (0.000343)*** | 0.00207 (0.00035)*** |
| (Log *NumOfViews*)$^2$ | −0.00014 (0.00009) | −0.00015 (0.00009)* | −0.00017 (0.000085)** | −0.00019 (0.00008)** |
| Log (I) | −0.03810 (0.00059)*** | −0.04196 (0.00117)*** | −0.03726 (0.000603)*** | −0.04071 (0.00120)*** |
| *Rating* | 0.00077 (0.00027)*** | 0.00077 (0.00027)*** | 0.00094 (0.000272)*** | 0.00074 (0.00027)** |
| log *NumOfLinks* | 0.00464 (0.00079)*** | 0.00476 (0.00079)*** | 0.00404 (0.000792)*** | 0.00457 (0.00079)*** |
| *Frn_NrmDegree* | −0.00069 (0.00017)*** | −0.00440 (0.00165)*** | −0.00079 (0.000187)** | −0.00449 (0.00165)*** |
| *Frn_NrmBP* | −0.00001 (9.107$E-6$) | 0.00017 (0.00007)** | −0.00002 (9.123$E-6$)*** | 0.00017 (0.00007)** |
| *Subs_NrmOut-Degree* | −0.00209 (0.00180) | 0.09068 (0.01258)*** | −0.00339 (0.000728)*** | 0.09135 (0.01258)*** |
| *Subs_NrmIn-Degree* | 0.01514 (0.00267)*** | 0.04580 (0.01971)** | 0.00532 (0.001700)*** | 0.05403 (0.01979)*** |
| log *NumFrn_Outside* | −0.00158 (0.00029)*** | −0.01317 (0.00217)*** | −0.00067 (0.000253)*** | −0.01218 (0.00219)*** |
| log *VAge* ∗ *Frn_NrmDeg* | | 0.00070 (0.00031)** | | 0.00071 (0.00031)** |
| log *VAge* ∗ *Frn_NrmBP* | | −0.00004 (0.00001)** | | −0.00004 (0.00001)*** |
| log *VAge* ∗ log *Subs_NrmOut* − *Degree* | | 0.00213 (0.00040)*** | | 0.00199 (0.00040)*** |
| log *VAge* ∗ log *Subs_NrmIn* − *Degree* | | −0.01751 (0.00238)*** | | −0.01763 (0.00238)*** |
| log *VAge* ∗ log *NumFrn_Out* | | −0.00532 (0.00372)** | | −0.00653 (0.00373)** |
| Video length | −0.0010 (0.00010)*** | −0.0010 (0.00010)*** | | |
| Adjusted $R$-squared | 0.1323 | 0.1348 | 0.1287 | 0.1290 |

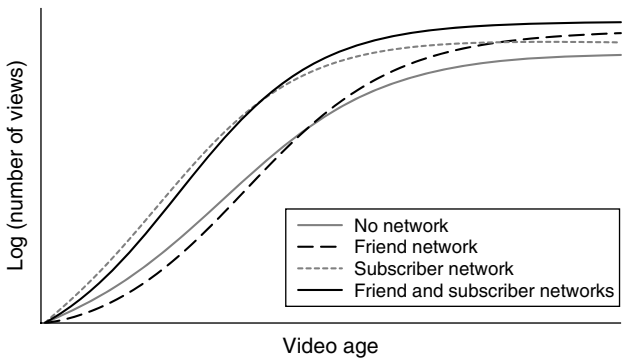
*Note.* Standard errors in parentheses; *$p = 0.10$; **$p = 0.05$; ***$p = 0.01$.

Based on the estimates from the structural model, we constructed several diffusion curves incorporating social network effects. Figure 8 illustrates the process of diffusion with and without the impact of various social network structures on YouTube.[14]

In the absence of social influence effects, diffusion would have been more rapid in the initial stages but would also have peaked with a lower number of aggregate views. Such diffusion dynamics could occur for two reasons. First, the dynamics whereby the number of adopters reaches a critical mass is likely to be different when we factor in the role of central channels in persuading early adopters. Channels that are highly connected in a subscriber network have the potential to act as *mavens* (Gladwell 2000). A video posted by the central channel first needs to reach a pool of early adopters, which subsequently influences the rate at which the video diffuses through the population. Because the subscriber network is a directed graph, the role of central nodes is to seed content in the initial stages of contagion, influencing awareness diffusion. The initial diffusion rate is then very sensitive to the network structure of incoming and outgoing subscriptions. Second, because we find that the role of conformity preferences and nonlocal ties is greater in the later stages of the diffusion process, diffusion through friend networks occurs only after some proportion of the initial population of users is *infected*, when a certain proportion has already viewed a video, or when a social threshold level of adoption creates a tipping point. One important implication is that the combined effect of friendship networks and subscriber networks reduces the time for the diffusion curve to reach the familiar S-shape. Figure 9 depicts the number of views of a video over time for a representative set of videos.

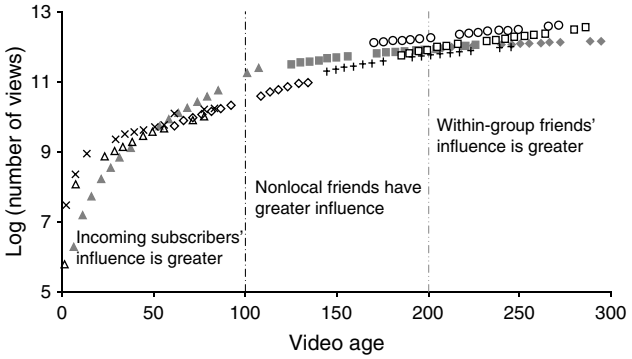
The results offer a contrast between the different types of information transmission in various types of social interactions. Channels occupying positions of influence in the subscribers' network are crucial in directing search and ensuring that a video acquires a critical mass of viewers. While the effects of social influence from friendship networks are not strong in the early stages of the life of a video, it is likely that a small number of highly influential nodes in the subscriber networks enhance the transmissibility of the video by setting off a trajectory of adoption within the subscription networks that cascades to the

**Figure 8 Diffusion With and Without Social Network Effects**




[14] This figure has been modified to scale and is not an exact representation of the estimated coefficients.

**Figure 9 Video Popularity with Video Age**




friend networks, creating a social multiplier effect that magnifies the impact of user tastes through social preferences such as conformity. If the process of information transmission does not depend on user tastes or homophily, we should observe that the impact of the originating channel's position in the friend network and subscriber networks should be similar throughout the diffusion process. We also find that the impact of nonlocal ties is stronger in the process of diffusion than that of within-group ties. Theory suggests that users place different weights on information acquired from other users (e.g., Bala and Goyal 1998). A preference for conformity leading to frequent and sustained connections with others in a local network creates a strong sense of identification; however, such conformity limits the amount of nonredundant information that can be transmitted across the local network. Long ties that are weaker in tie strength can be very influential in extending diffusion beyond a cohesive group. Our results suggest that the informational role of long ties could be more important than the relational role of stronger ties. The contrasting effects of degree centrality and the Bonacich power measures in a friend network also suggest that homophily and social status might have subtle differences in the process of diffusion.

### 6.2. Contributions to Literature and Practice

Prior research has extensively analyzed the role of online dissemination of consumer opinions such as online word-of-mouth (e.g., Chevalier and Mayzlin 2006), consumer-generated media (Dewan and Ramaprasad 2008), and social identity disclosure in user-generated content (Forman et al. 2008). The focus on this paper is on the impact of the network position of the content creator, which not only impacts whether a video achieves success or failure but also affects the magnitude of the impact. This inquiry into the role of network position distinguishes this study from prior studies on user generated content in IS literature. Prior literature on diffusion highlights the importance of promotional efforts such as targeted communication (Manchanda et al. 2008), advertising (Van den Bulte and Lilien 2001) and the role of influentials (Dodds and Watts 2007). In addition to promotional efforts, content creators control the timing of the release of an album as well as promotional effort such as radio play (e.g., Bhattacharjee et al. 2007). Such promotional efforts play a much more limited role on YouTube where the model of content creation is more democratic as evidenced by the blockbuster popularity of amateur content.

The results in this paper demonstrate that social networks impact economic outcomes by structuring on the information available to other actors, which influences others' decisions, perceptions, and behavior. The social capital fostered through networked interactions might also mitigate the potential for information asymmetry, suggesting that research on reputation systems on the Internet (e.g., Resnick et al. 2000) could incorporate social networks based explanations. This paper also highlights that various forms of networked social interactions exert different types of interpersonal influence and the nature of information transmission can be different depending on whether ties are homophilous or instrumental ties. This perspective could enrich the stream of research on the adoption and diffusion of technological innovations by highlighting the process of interpersonal influence. Multiplier effects arising from social contagion within a social network can be instrumental in shaping perceptions of the usefulness of innovations (e.g., Bandiera and Rasul 2006), explaining the trajectory of diffusion of technological innovations.

While technology-mediated communications existed earlier, the new models of social computing are characterized by a spontaneous emergence of communities, with a wealth of opportunity for participatory interaction, self-expression, and collective action. The ease with which users can participate in social computing platforms and seek out digital content such as music and entertainment is of tremendous interest to marketers and content creators. Businesses are attempting to enhance customer engagement through a community-centered approach to product development and brand building (Wells 2009), shifting power to the edge of the network. With the strong rate of growth of YouTube (Peck et al. 2008), and as online video supplants traditional media channels such as television, the question facing companies is whether to sponsor a set of content creators to post their content on YouTube or to fund a set of videos that have already acquired momentum. We find that the number of outgoing subscriptions has a different impact from that of the number of incoming subscriptions of a channel, indicating that who initiates the

connections in a network matters to how the content is diffused. A push model of content creation, where a channel tries to initiate connections to other users, might be less successful than a pull model, where a channel with high in-degree centrality acts as a fashion leader by initiating contagion. Therefore, one implication for practitioners is to consider the role of the prestige of a channel in strategically seeding content in order to maximize impact.

One perspective in marketing contends that a small number of *mavens* or *influentials* play a disproportionate role in spreading new ideas or triggering adoption of a new product (Gladwell 2000). Another perspective questions whether influentials can act as gatekeepers and suggests that a critical mass of easily persuadable individuals drives collective contagion (Dodds and Watts 2007). The results in this paper suggest that central channels in the subscriber networks act as *influentials* in directing search and product discovery. Given the critical importance of incoming subscriptions, channels could try to build an initial user base through purchasing key words to improve their listing in sponsored search, through promotional videos, or by listing in YouTube's featured videos listing.

While subscription networks are important in directing viewer attention and acquiring a critical mass of viewers, the magnitude of the popularity also depends on a channel's position in the friend networks. Content creators could monetize content through subscription fees or micro payments; however, an advantageous position in the subscriber network alone might not be enough for a channel to obtain revenues. Rather, it could be greater user engagement driven by homophily or affiliation that could be important in monetization of content. Thus different types of networks play different roles in ensuring channel loyalty and user engagement, and they consequently impact the monetization of content.

It has been suggested that online video communities are characterized by user stickiness (Peck et al. 2008). Content creators need to comprehend how networks of interpersonal influence promote users' identification with their channel to design future promotional efforts and advertising campaigns. Businesses seeking to monetize social networks, in particular, could obtain insights to maximize the impact of targeted marketing, advertising, and the spread of digital content. Understanding the social structure of interactions is also important in identifying authoritative nodes that influence product discovery and viewing behavior. For instance, the local popularity of a video is considerably different from that of global popularity (Baluja et al. 2008). An understanding of local network effects, and particularly user engagement within the local network, is important to develop greater knowledge of what users search for and what impacts user experience in online social networks. Analyzing such patterns of social interactions could prove valuable in developing recommendation tools and collaborative filtering systems (e.g., Oestricher-Singer and Sundararajan 2008).

### 6.3. Limitations

This study has a few limitations. We also cannot rule out whether attention-seeking efforts by content creators impact diffusion by increasing the saliency of each video, rather than the social interactions posited here. While we distinguish between the impacts of incoming and outgoing subscriptions on the aggregate diffusion process, we do not examine the deeper theoretical differences in the different types of subscriptions. Future work can examine these differences in terms of network structure and conditions that can trigger a large-scale epidemic propagation. We also do not consider whether social influence through offline interactions bolsters the impact of conformity pressures and status in networked interactions online. Another issue is that the impact of channel centrality in directing search and influence potential experience could be a result of the involvement of the channel in the overall YouTube network, increasing the salience of videos posted by a channel among potential viewers rather than the impact of social contagion.

While we control for the differences in individual connectedness of a channel, we do not investigate whether individual heterogeneity results in different agents updating their beliefs differently. It is possible that contagion might be triggered through a complex interaction between network structure and decision making in groups, whereby agents' decisions could depend not only on the incident ties of an agent but also on the characteristics and actions of the agents' neighbors. Bala and Goyal (1998) characterize learning from neighbors as a Bayesian updating process whereby an agent updates her beliefs based on the actions of other agents. Dodds and Watts (2007) model interpersonal influence as a function of not only network characteristics such as proximity but also the influencer's expertise and characteristics of other individuals adjacent to the influencer. It might be necessary to conduct experimental studies to tease out such impacts.

## 7. Conclusions and Future Research

While the proliferation of social computing models is of interest to academics and practitioners alike, an analysis of social computing platforms cannot take place without an understanding of the interpersonal influence underlying individual behavior. For businesses seeking to monetize social search and digital content, in particular, it is increasingly important

to understand how users interact and participate in these settings. This study investigates how contagion impacts the trajectory of diffusion in a social computing platform. We integrate multiple theoretical perspectives to highlight that social interactions structured through a network significantly influence not only the spread of contagion but also the magnitude of the contagion. We find that diffusion proceeds in a two-stage process in which a product's search characteristics matter in the early stage in triggering awareness diffusion creating a role for subscriber networks, while the experience characteristics matter more in the later stage, and thus the role of a central channel in the friend networks in disseminating experience-related information.

Identifying social influence is complicated due to the difficulty in distinguishing between social influence and self-selection of users. We address the reflection problem by systematically separating between factors that affect membership in a social network from other types of social influence. We can thus rule out the possibility that an individual's viewing pattern is either solely influenced by the social group that they belong to. The Hausman-Taylor estimation considers the impact of a viewer's time invariant loyalty to a video. To address the issue of the unobservable attention-seeking efforts, we consider channel specific heterogeneity resulting from the aggregate connectedness of a channel. Given the multiplicity of factors constituting social influence, this study makes a first step in demonstrating not only that social influence matters but also that we can disentangle the different mechanisms by which social influence is structured. Our estimation is robust to unobservable self-selection of users, unobserved heterogeneity in user preferences and robust to contemporaneous shocks that might result in serial correlation.

In an increasingly hypernetworked age, individuals have access to informational content from a wide variety of online sources such as blogs, consumer forums, podcasts, and social media that influence their tastes and preferences. Individuals whose networks bridge a variety of sources have access to a diversity of information and can translate information across groups. Agents who broker across structural holes in a network might have an informational advantage resulting from access to multiple sources of information. Future research can explore how information is transmitted across different networks and whether prominent users function as conduits of information linking networks across different forms of social media.

## Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at http://isr.journal.informs.org/.

## Acknowledgments

The authors thank Sanjeev Dewan and the ISR review team for constructive feedback and suggestions; V. Sambamurthy, Ritu Agarwal, Alok Gupta, Mani Subramani, and seminar participants at the University of Minnesota, Hong Kong University of Science and Technology, University of Washington, Singapore Management University, Towson University, University of Washington Information School, City University of Hong Kong, Workshop in E-Business 2007, International Conference in Information Systems 2008, Informs Conference 2009, and the Symposium on Statistical Challenges in eCommerce Research 2008 for comments on earlier versions of this paper.

## References

Angrist, J., A. Krueger. 1991. Does compulsory school attendance affect schooling and earnings? *Quart. J. Econom.* **106**(4) 979–1014.

Armstrong, C. P., V. Sambamurthy. 1999. Information technology assimilation in firms: The influence of senior leadership and IT infrastructures. *Inform. Systems Res.* **10**(4) 304–327.

Bala, V., S. Goyal. 1998. Learning from neighbors. *Rev. Econom. Stud.* **65**(3) 595–621.

Baluja, S., R. Seth, D. Sivakumar, Y. Jing, J. Yagnik, S. Kumar, D. Ravichandran, M. Aly. 2008. Video suggestion and discovery for YouTube: Taking random walks through the view graph. *Proc. 17th Internat. Conf. World Wide Web, Beijing*, 895–904. http://www.conference.org/www2008/papers/pdf/p895-svakumarA.pdf.

Bandiera, O., I. Rasul. 2006. Social networks and technology adoption in Northern Mozambique. *Econom. J.* **116**(514) 862–902.

Bass, F. 1969. A new product growth for model consumer durables. *Management Sci.* **15**(5) 215–227.

Bemmaor, A. C., J. C. Lee. 2002. The impact of heterogeneity and ill-conditioning on diffusion model parameter estimates. *Marketing Sci.* **21**(2) 209–220.

Bernheim, D. A. 1994. A theory of conformity. *J. Political Econom.* **102**(5) 841–877.

Bhattacharjee, S., R. D. Gopal, K. Lertwachara, J. R. Marsden, R. Telang. 2007. The effect of digital sharing technologies on music markets. *Management Sci.* **53**(9) 1359–1374.

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

Bonacich. 1987. Power and centrality: A family of measures. *Amer. J. Sociol.* **92**(5) 1170–1182.

Borgatti, S. P. 2005. Centrality and network flow. *Soc. Networks* **27**(1) 55–71.

Borgatti, S. P., M. G. Everett. 2006. A graph-theoretic framework for classifying centrality measures. *Soc. Networks* **28**(4) 466–484.

Borgatti, S. P., M. G. Everett, L. C. Freeman. 2002. *Ucinet for Windows: Software for Social Network Analysis*. Analytic Technologies, Harvard, MA.

Boulding, W., M. Christen. 2003. Sustainable pioneering advantage? Profit implications of market entry order. *Marketing Sci.* **22**(3) 371–392.

Bramoulle, Y., H. Djebbari, B. Fortin. 2009. Identification of peer effects through social networks. *J. Econometrics* **150**(1) 41–55.

Burt, R. 1987. Social contagion and innovation: Cohesion versus structural equivalence. *Amer. J. Sociol.* **92**(6) 1287–1335.

Burt, R. 1992. *Structural Holes: The Social Structure of Competition*. Harvard University Press, Cambridge, MA.

Centola, D., M. Macy. 2007. Complex contagions and the weakness of long ties. *Amer. J. Sociol.* **113**(3) 702–734.

Cha, M. H., P. Kwak, Y. A. Rodriguez, S. Moon. 2007. I Tube, YouTube, everybody Tubes: Analyzing the world's largest user generated content video system. *Proc. ACM Special Interest*

*Group Data Comm. (SIGCOMM) Internet Measurement Conf., San Diego.*

Chevalier, J., D. Mayzlin. 2006. The effect of word of mouth online: Online book reviews. *J. Marketing Res.* **43**(3) 345–354.

Coleman, J. S., E. Katz, H. Menzel. 1966. *Medical Innovation: Diffusion of a Medical Drug Among Doctors*. Bobbs-Merrill, Indianapolis.

Crane, R., D. Sornette. 2008. Robust dynamic classes revealed by measuring the response function of a social system. *Proc. Natl. Acad. Sci.* **105**(41) 15649–15653.

Dewan, S., J. Ramaprasad. 2008. Consumer blogging and music sampling. Working paper, University of California at Irvine, Irvine.

Dodds, P. S., D. J. Watts. 2004. Universal behavior in a generalized model of contagion. *Phys. Rev. Lett.* **92**(21). http://prl.aps.org/abstract/PRL/v92/i21/e218701.

Dodds, P. S., D. J. Watts. 2005. A generalized model of social and biological contagion. *J. Theoret. Biol.* **232**(4) 587–604.

Dodds, P. S., D. J. Watts. 2007. Influentials, networks, and public opinion formation. *J. Consumer Res.* **34** 441–458.

Easley, D., J. Kleinberg. 2010. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, Cambridge, UK.

Fichman, R. G. 2000. The diffusion and assimilation of information technology innovations. R. W. Zmud, ed. *Framing the Domains of IT Management: Projecting the Future Through the Past*. Pinnaflex Educational Resources, Inc. Cincinnati, OH.

Forman, C., A. Ghose, B. Wiesenfeld. 2008. Examining the relationship between reviews and sales: The role of reviewer identity information. *Inform. Systems Res.* **19**(3) 291–313.

Freeman, L. C. 1979. Centrality in social networks: A conceptual clarification. *Soc. Networks* **1**(3) 215–239.

Gelman, A., J. Hill. 2007. *Data Analysis Using Regression and Mulitlevel/ Hierarchical Models*. Cambridge University Press, Cambridge, UK.

Gladwell, M. 2000. *The Tipping Point: How Little Things Can Make a Big Difference*. Little Brown, Boston.

Granovetter, M. 1973. The strength of weak ties. *Amer. J. Sociol.* **78**(6) 1360–1380.

Hausman, J. A., W. E. Taylor. 1981. Panel data and unobservable individual effects. *Econometrica* **49**(6) 1377–1398.

Hendricks, K., A. Sorenson. 2009. Information and the skewness of music sales. *J. Political Econom.* **117**(2) 324–369.

Ibarra, H., S. B. Andrews. 1993. Power, social Influence, and sense making: Effects of network centrality and proximity on employee perceptions. *Admin. Sci. Quart.* **38**(2) 277–303.

Kalish, S. 1985. A new product adoption model with price, advertising and uncertainty. *Management Sci.* **31**(12) 1569–1585.

Kraut, R. E., R. E. Rice, C. Cool, R. S. Fish. 1998. Varieties of social influence: The role of utility and norms in the success of a new communication medium. *Organ. Sci.* **9**(4) 437–453.

Liu, H., P. Maes, G. Davenport. 2006. Unraveling the taste fabric of social networks. *Internat. J. Semantic Web Inform. Systems* **2**(1) 42–71.

Mahajan, V., E. Muller, F. M. Bass. 1990. New product diffusion models in marketing: A review and directions for research. *J. Marketing* **54**(1) 1–26.

Manchanda, P., Y. Xie, N. Youn. 2008. The role of targeted communication and contagion in new product adoption. *Marketing Sci.* **27**(6) 950–961.

Manski, C. F. 1993. Identification of endogenous social effects: The reflection problem. *Rev. Econom. Stud.* **60**(3) 531–542.

Nelson, P. 1970. Information and consumer behavior. *J. Political Econom.* **78**(2) 311–329.

Newman, M. E. J. 2003. The structure and function of complex networks. *SIAM Rev.* **45**(2) 167–256.

Oestreicher-Singer, G., A. Sundararajan. 2008. The visible hand of social networks in electronic markets. Working paper, New York University, New York.

Parameswaran, M., A. B. Whinston. 2007. Social computing: An overview. *Comm. Assoc. Inform. Systems* **19** 762–780.

Peck, R. S., L. Y. Zhou, V. B. Anthony, K. Madhukar. 2008. Consumer Internet, Bear Stearns equity research report. Bear Stearns, New York.

Putnam, R. 2000. *Bowling Alone: The Collapse and Revival of American Community*. Simon and Schuster, New York.

Raymond, E. 2001. *The Cathedral and the Bazaar*. O'Reilly, Sebastopol, CA.

Reagans, R. E., E. W. Zuckerman. 2008. Why knowledge does not equal power: The network redundancy trade-off. *Indust. Corporate Change* **17**(5) 903–944.

Resnick, P., R. Zeckhauser, E. Friedman, K. Kuwabara. 2000. Reputation systems. *Comm. ACM* **43**(12) 45–48.

Rogers, E. 1995. *Diffusion of Innovations*, 4th ed. The Free Press, New York.

Rogers, E. M., L. D. Kincaid. 1981. *Communication Networks*. Free Press, New York.

Sacerdote, B. 2001. Peer effects with random assignment: Results for Dartmouth roommates. *Quart. J. Econom.* **116**(2) 681–704.

Salganik, M., P. S. Dodds, D. Watts. 2006. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* **311**(5762) 854–856.

Strang, D., N. B. Tuma. 1993. Spatial and temporal heterogeneity in diffusion. *Amer. J. Sociol.* **99**(3) 614–639.

Sundararajan, A. 2007. Local network effects and complex network structure. *Berkeley Electronic J. Theoret. Econom.* **7**(1) Article 46. http://www.bepress.com/bejte/vol7/iss1/art46

Szabo, G., B. A. Huberman. 2009. Predicting the popularity of online content. *Comm. ACM.* Forthcoming.

Talukdar, D., K. Sudhir, A. Ainslie. 2002. Investigating new product diffusion across products and countries. *Marketing Sci.* **21**(1) 97–114.

Van den Bulte, C., G. Lilien. 2001. Medical innovation revisited: Social contagion versus marketing effort. *Amer. J. Sociol.* **106**(5) 1409–1435.

Van den Bulte, C., S. Stremersch. 2004. Social contagion and income heterogeneity in new product diffusion: A meta-analytic test. *Marketing Sci.* **23**(4) 530–544.

Wang, F.-Y., K. M. Carley., D. Zeng, W. Mao. 2007. Social computing: From social informatics to social intelligence. *IEEE Intelligent Systems* **22**(2) 79–83.

Wasserman, S., K. Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, UK.

Watts, D. J. 2002. A simple model of global cascades on random networks. *Proc. Natl. Acad. Sci.* **99**(9) 5766–5771.

Watts, D. J., P. S. Dodds, M. E. J. Newman. 2002. *Science* **296**(5571) 1302–1305.

Wejnert, B. 2002. Integrating models of diffusion of innovations: A conceptual framework. *Annual Rev. Sociol.* (28) 297–326.

Wells, J. 2009. It's crunch time. *Globe Mail* (February 13). http://www.theglobeandmail.com/report-on-business/article971011.ece.

Wooldridge, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA.